Московский государственный университет имени М.В. Ломоносова
Факультет вычислительной математики и кибернетики
Кафедра исследования операций

Теория эконометрики

Лекция 1

Москва, 2020

Содержание

Введение

Эконометрика как наука расположена где-то между экономикой, статистикой и математикой. Эконометрика - это наука, связанная с эмпирическим выводом экономических законов. То есть эконометристы используют данные или "наблюдения"для того, чтобы получить количественные зависимости для экономических соотношений.

Также эконометрист формулирует **экономические модели**, основываясь на экономической теории или на эмпирических данных, оценивает **неизвестные величины (параметры)** в этих моделях, делает прогнозы (и оценивает их точность) и дает рекомендации по экономической политике.

Типы моделей

Математические модели полезны для более полного понимания сущности происходящих процессов, их анализа. Можно выделить три основных класса моделей, которые применяются для анализа и/или прогноза.

Модели временных рядов

К этому классу относятся модели:

тренда:

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

где $T(t)-$ временной тренд заданного параметрического вида(например, линейный $T(t) = a + bt$), ε_t- случайная (стохастическая) компонента;

сезонности:

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

где $S(t)-$ периодическая (сезонная) компонента, ε_t- случайная (стохастическая) компонента;

тренда и сезонности:

$$y(t) = T(t) + S(t) + \varepsilon_t, \text{ (аддитивная) или}$$

$$y(t) = T(t) \cdot S(t) + \varepsilon_t, \text{ (мультипликативная),}$$

где $T(t)-$ временной тренд заданного параметрического вида, $S(t)-$ периодическая (сезонная) компонента, ε_t- случайная (стохастическая) компонента.

Их общей чертой является то, что они объясняют поведение временного ряда, исходя только из его предыдущих значений. Более сложные модели временных рядов: модели адаптивного прогноза, модели авторегрессии и скользящего среднего (ARINMA) и другие.

Регрессионные модели с одним уравнением

В таких моделях **зависимая (объясняемая)** переменная y предоставляется в виду функции $f(x, \beta) = f(x_1, ..., x_k, \beta_1, ..., \beta_p)$, где $x_1, ..., x_k$ - **независимые (объясняющие)** переменные, а $\beta_1, ..., \beta_p$ - параметры. В зависимости от вида функции $f(x, \beta)$ модели делятся на *линейные и нелинейные.*

Область применения таких моделей, даже линейных, значительно шире, чем моделей временных рядов.

Системы одновременных уравнений

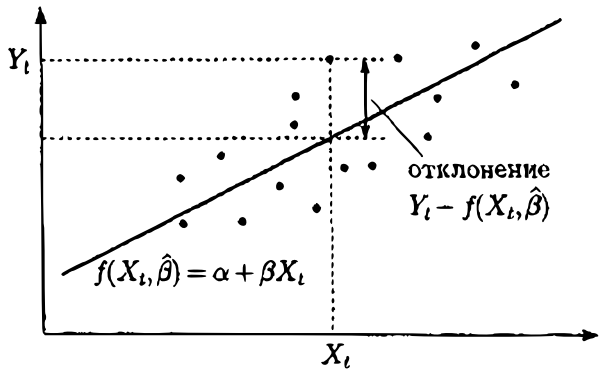
Эти модели описываются системами уравнений. Системы могут состоять из тождеств и регрессионных уравнений, каждое из которых может, кроме объясняющих переменных, включать объясняемые переменные из других уравнений системы.

Типы данных

При моделировании экономических процессов встречаются два типа данных: **пространственные данные** *(cross-sectional data)* и **временные ряды** *(time-series data).*

Модель парной регрессии

Пусть у нас есть набор значений двух переменных X_t, Y_t, $t = 1, ..., n$; можно отобразить пары (X_t, Y_t) точками на плоскости.



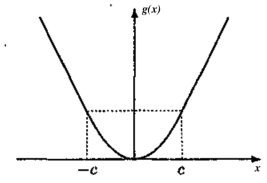
Предположим, что нашей задачей является подобрать ("подогнать") функцию $Y = f(X)$ из параметрического семейства функций $f(X, \beta)$, "наилучшим"способом описывающую зависимость Y от X, то есть выбрать наилучшее значение $\hat{\beta}$.

В качестве меры отклонений функции $f(X, \beta)$ от набора наблюдений можно взять:

1. сумму квадратов отклонений $F = \sum_{t=1}^{n} (Y_t - f(X_t, \hat{\beta}))^2$,
2. сумма модулей отклонений $F = \sum_{t=1}^{n} |Y_t - f(X_t, \hat{\beta})|$,
3. $F = \sum_{t=1}^{n} g(Y_t - f(X_t, \hat{\beta}))^2$, где g – мера, с которой отклонения $Y_t - f(X_t, \hat{\beta})$ входит в функционал F.

Примером такой "меры"может служить **функция Хубера**, которая при малых отклонениях квадратичная, а при больших линейная:

$$g(x) = \begin{cases} x^2, & \text{если } |x| < c, \\ 2cx - c^2, & \text{если } x \geq c, \\ -2cx - c^2, & \text{если } x \leq -c. \end{cases}$$



Рассмотрим достоинства и недостатки рассмотренных функционалов.

Сумма квадратов отклонений

Плюсы метода:

– легкость вычислительной процедуры;

– хорошие статистические свойства, простота математических выводов делает возможным построить развитую теорию, позволяющую провести тщательную проверку различных статистических гипотез;

минусы метода:

– чувствительность к "выбросам"(*outliers*).

Сумма модулей отклонений

Плюсы метода:

– робастность, т. е. нечувствительность к выбросам;

минусы метода:

– сложность вычислительной процедуры;

– возможно, большим отклонениям надо придавать больший вес (лучше два отклонения величиной 1, чем два отклонения величиной 0 и 2);

– неоднозначность, т. е. разным значениям параметра β могут соответствовать одинаковые суммы модулей отклонений.

Метод наименьших квадратов (МНК)

Рассмотрим задачу "наилучшей"аппроксимации набора наблюдений $X_t, Y_t, t = 1, ..., n$, линейной функцией $f(X) = a + bX$ в смысле минимизации функционала

$$F = \sum_{t=1}^{n} (Y_t - (a + bX_t))^2 \to \min_{\forall a,b}.$$

Запишем необходимые условия экстремума *(First Order Condition, FOC)*:

$$\frac{\partial F}{\partial a} = -2 \sum_{t=1}^{n} (Y_t - a - bX_t) = 0, \quad \frac{\partial F}{\partial b} = -2 \sum_{t=1}^{n} X_t (Y_t - a - bX_t) = 0, \text{ или}$$

$$\sum_{t=1}^{n} (Y_t - a - bX_t) = 0, \quad \sum_{t=1}^{n} X_t (Y_t - a - bX_t) = 0.$$

Раскроем скобки и получим *стандартную форму нормальных уравнений* (для краткости опустим индексы суммирования у знака суммы $\sum$):

$$an + b\sum X_t = \sum Y_t, \quad a\sum X_t + b\sum {X_t}^2 = \sum X_t Y_t.$$

Решением этой системы будут $\hat{a}, \hat{b}$:

$$\hat{a} = \frac{1}{n}\sum Y_t - \frac{1}{n}X_t\hat{b} = \bar{Y} - \bar{X}b,$$

$$\hat{b} = \frac{n\sum X_t Y_t - (\sum X_t)(\sum Y_t)}{n\sum {X_t}^2 - (\sum X_t)^2} = \frac{Cov(X,Y)}{VarX}.$$

Замечание 1. Из первого уравнения следует

$$\bar{Y} = \hat{a} + \hat{b}\bar{X},$$

то есть уравнение прямой линии $Y = \hat{a} + \hat{b}X$, полученное в результате минимизации функционала, проходит через точку $(\bar{X}, \bar{Y})$. Здесь через $\bar{X}, \bar{Y}$ обозначены выборочные средние значения переменных X_t и Y_t: $\bar{X} = (1/n)\sum X_t, \bar{Y} = (1/n)\sum Y_t$. А

$$Cov(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}), \quad Var(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Уравнения в отклонениях

Обозначим через $x_t = X_t - \bar{X}$, $y_t = Y_t - \bar{Y}$ отклонения от средних по выборке значений X_t и Y_t, $\bar{X} = (1/n)\sum X_t, \bar{Y} = (1/n)\sum Y_t$. *(Проверьте, что $\bar{x} = \bar{y} = 0$).*

Решим теперь ту же задачу: подобрать линейную функцию $f(x) = a + bx$, минимизирующую функционал

$$F = \sum_{t=1}^{n}\left(y_t - (a + bx_t)\right)^2.$$

Из геометрических соображений ясно, что решением задачи будет та же прямая на плоскости (x, y), что и для исходных данных X_t, Y_t. В самом деле, в силу $\bar{Y} = \hat{a} + \hat{b}\bar{X}$ переход от X, Y к отклонениям x, y означает лишь перенос начала координат в точку $(\bar{X}, \bar{Y})$. Вычисления, которые необходимо проделать для решения задачи, вполне аналогичны предыдущим (с заменой X, Y на x, y). Сделав замену X_t, Y_t на x_t, y_t в равенствах для $\hat{a}, \hat{b}$, учитывая, что

$\bar{x} = \bar{y} = (1/n)\sum x_t = (1/n)\sum y_t = 0$, получим

$$\hat{a} = 0, \hat{b} = \frac{\sum x_t y_t}{\sum x_t^2} = \frac{\sum (X_t - \bar{X})(Y_t - \bar{Y})}{\sum (X_t - \bar{X})^2}.$$

Таким образом, мы получили другое выражение для углового коэффициента прямой $\hat{b}$.

Матричная форма записи

Обозначим через X матрицу размерности $n \times 2$; β - матрица (вектор) коэффициентов размерностью 2×1; i - единичный вектор; $\hat{y} = ai + bx = X\beta$ - вектор, лежащий в двумерной гиперплоскости π, натянутой на векторы i, x; $e = y - \hat{y} = y - X\beta$:

$$X = \begin{bmatrix} 1 & X \\ \dots & \dots \\ 1 & X_n \end{bmatrix}, \quad y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \quad \beta = \begin{bmatrix} a \\ b \end{bmatrix}, \quad i = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \quad e = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix} \quad x = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}.$$

Условие ортогональности вектора e плоскости π записывается как:

$$X'e = 0, \text{ где } X' \text{ - транспонированная матрица } X, \text{ или } X'(y - X\hat{\beta}) = X'y - X'X\hat{\beta} = 0,$$

таким образом мы получили необходимые условия экстремума. Преобразовав последнее равенство получаем:

$$X'X\hat{\beta} = X'y, \text{ или } \hat{\beta} = (X'X)^{-1}X'y,$$

в предположении, что векторы x, i линейно независимы и, следовательно, матрица $X'X$ обратима. Также нетрудно проверить, что:

$$\hat{\beta} = (X'X)^{-1}X'y = \begin{bmatrix} n & \sum X_t \\ \sum X_t & \sum {X_t}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum Y_t \\ \sum X_t Y_t \end{bmatrix}^{-1} = \dots = \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix}.$$

Отметим, что матрица $X'X$ невырождена, т.к. матрица X имеет максимальный ранг 2 $(rank(X'X) = 2 \Rightarrow \exists (X'X)^{-1})$.

Линейная регрессионная модель с двумя переменными

Добавим к постановке задачи некоторые статистические свойства данных.

На самом деле, для одного мы можем наблюдать разные значения Y.

Пример 1. -возраст индивидуума, Y — его зарплата.

Пример 2. X — доход семьи, Y расходы на питание.

Запишем уравнение зависимости Y_t от X_t, в виде

$$Y_t = a + bX_t + \epsilon_t, t = 1, ..., n,$$

где X - неслучайная (детерминированная) величина, а Y_t, ε_t - случайные величины, Y_t называется объясняемой (зависимой) переменной, а X_t - объясняющей (независимой) или *регрессором*. Уравнение, приведенное выше, также называется *регрессионным уравнением*.

Какова природа ошибки ε_t?

Есть две основные возможные причины случайности:

1. Наша модель является упрощением действительности и на самом деле есть еще другие параметры (пропущенные переменные, *omitted variables*), от которых зависит Y. Зарплата, например, может зависеть от уровня образования, стажа работы, пола, типа фирмы (государственная, частная) и т. п.

2. Трудности в измерении данных (присутствуют ошибки измерений). Например, данные по расходам семьи на питание составляются на основании записей участников опросов, которые, как предполагается, тщательно фиксируют свои ежедневные расходы. Разумеется, при этом возможны ошибки.

Таким образом, можно считать, что ε_t — случайная величина с некоторой функцией распределения, которой соответствует функция распределения случайной величины Y_t.

Основные гипотезы:

1. $Y_t = a + bX_t + \varepsilon_t, t = 1, ..., n$, - спецификация модели.

2. X_t - детерминированная величина; вектор $(X_1, ..., X_n)'$ не коллинеарен вектору $i = (1, ..., 1)'$.

3. (a) $E\varepsilon_t = 0$, $E(\varepsilon_t)^2 = V(\varepsilon_t) = \sigma^2$ - не зависит от t ($E\varepsilon = 0$, $V(\varepsilon) = \sigma^2 I_n$)

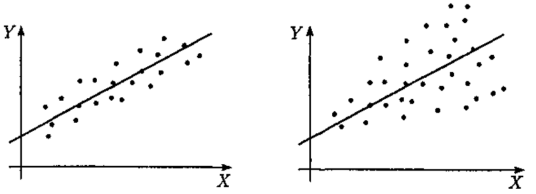
   (b) $E(\varepsilon_t \varepsilon_s) = 0$ при $t \neq s$, некоррелированность ошибок для разных наблюдений. Часто добавляется условие:

   (c) Ошибка $\varepsilon_t, t = 1, ..., n$, имеют совместное нормальное распределение: $\varepsilon_t \sim N(0, \sigma^2)$

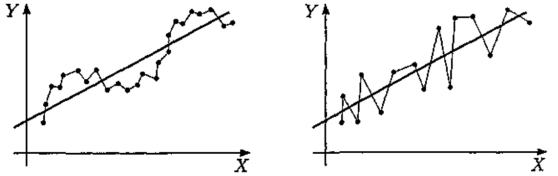
В этом случае модель называется нормальной линейной регрессионной *(Classical Normal Linear Regression model)*.

Условие $E\varepsilon = 0$, означает, что $EY_t = a + bX_t$, то есть при фиксированном X_t среднее ожидаемое значение Y_t равно $a + bX_t$.

Условие независимости дисперсии ошибки от номера наблюдения (от регрессора X_t): $(\varepsilon_t^2) - V(\varepsilon_t) = \sigma^2, t = 1, ..., n$, называется *гомоскедастичностью (homoscedasticity)*; случай, когда условие гомоскедастичности не выполняется, называется *гетероскедастичностью (heteroscedasticity)*. На рисунках ниже приведены пример типичной картинки для случая гомоскедастичности ошибок и пример данных с гетероскедастичными ошибками (возможно, что в этом примере $V(\varepsilon_t) \sim X_t^2$). Условие $(\varepsilon_t \varepsilon_s) = 0, t \neq s$ указывает на некоррелированность ошибок для разных наблюдений. Это условие часто нарушается в случае, когда наши данные являются временными рядами.В случае, когда это условие не выполняется, говорят об *автокорреляции ошибок (serial correlation)*.



Для простейшего случая автокорреляции ошибок, когда $(\varepsilon_t \varepsilon_s) = \rho \neq 0$, типичный вид данных представлен на рисунках для $(\rho > 0)$ и $(\rho < 0)$ соответственно.



Теорема Гаусса–Маркова. Оценка дисперсии ошибок σ^2

Итак, мы имеем набор данных (наблюдений) $(X_t, Y_t), t = 1, \ldots, n$, и модель $1-3ab$. Наша задача — оценить все три параметра модели: a, b, σ^2.

Мы хотим оценить параметры a и b «наилучшим» способом. Что значит «наилучшим»? Например, найти в классе линейных (по Y_t) несмещенных оценок наилучшую в смысле минимальной дисперсии *(Best Linear Unbiased Estimator, BLUE)*.

Заметим, что когда такая оценка найдена, это вовсе не означает, что не существует нелинейной несмещенной оценки с меньшей дисперсией. Кроме того, например, можно отбросить требование несмещенности оценки и минимизировать среднеквадратичное отклонение оценки от истинного значения: $\mathrm{E}(\widehat{b} - b)^2$.

Теорема Гаусса–Маркова. *В предположениях модели* $1-3ab$*:*

1. $Y_t = a + bX_t + \varepsilon_t, t = 1, \ldots, n;$
2. X_t*– детерминированная величина;*
3. *(a)* $\mathrm{E}\varepsilon_t = 0$, $\mathrm{E}(\varepsilon_t^2) = \mathrm{V}(\varepsilon_t) = \sigma^2$*;*
   (b) $\mathrm{E}(\varepsilon_t \varepsilon_s) = 0$*, при* $t \neq s$*;*

оценки $\widehat{a}, \widehat{b}(2.4a), (2.4б)$*, полученные по методу наименьших квадратов (МНК),*

$$\widehat{b} = \frac{n\sum X_t Y_t - (\sum X_t)(\sum Y_t)}{n\sum X_t^2 - (\sum X_t)^2}$$

$$\widehat{a} = \frac{1}{n}\sum Y_t - \frac{1}{n}\sum X_t \widehat{b} = \overline{Y} - \overline{X}\widehat{b}.$$

имеют наименьшую дисперсию в классе всех линейных несмещенных оценок.

Доказательство. 1) несмещенность $\widehat{a}, \widehat{b}$

2) значение дисперсий $\mathrm{V}(\widehat{a})$, $\mathrm{V}(\widehat{b})$ вычсил.

3) проверка, что $\nexists$ оценки с меньшей дисперсией

1. Проверим несмещенность МНК-оценок $\widehat{a}, \widehat{b}$

$$\mathrm{E}\widehat{b} = \mathrm{E}\frac{\sum x_t y_t}{\sum x_t^2} = \frac{\sum x_t \mathrm{E}y_t}{\sum x_t^2} = (*) = \frac{b\sum x_t^2}{\sum x_t^2} = b.$$

(*) $\mathrm{E}y_t = \mathrm{E}(Y_t - \overline{Y}) = \mathrm{E}(a + bX_t + \varepsilon_t) - (a + b\overline{X} + \varepsilon_t) =$
$= \mathrm{E}(b(X_t - \overline{X})) = bx_t.$

$$\mathrm{E}\widehat{a} = \mathrm{E}(\overline{Y} - \widehat{b\overline{X}}) = a + b\overline{X} - \overline{X}\mathrm{E}(\widehat{b}) - b\overline{X} = a$$

2. Вычислим дисперсии оценок $\widehat{a}, \widehat{b}$.

$$\widehat{b} = \frac{\sum x_t y_t}{\sum x_t^2} = \sum w_t y_t, \text{где } w_t = \frac{x_t}{\sum\limits_s x_s^2}$$

w_t удовлетворяет следующим условиям:

1) $\sum w_t = 0$
2) $\sum w_t x_t = \sum w_t X_t = 1$
3) $\sum w_t^2 = \frac{1}{\sum x_t^2}$
4) $\sum w_t y_t = \sum w_t Y_t$

Проверим эти условия.

1)
$$\sum w_t = \frac{\sum x_t}{\sum\limits_s x_s^2} = \frac{n\overline{x}}{\sum\limits_s x_s^2} = 0(\overline{x} = 0) \Rightarrow \sum w_t = 0.$$

2)
$$\sum w_t x_t = \frac{\sum x_t^2}{\sum\limits_s x_s^2} = 1 \Rightarrow \sum w_t x_t = 1$$

$$\sum w_t x_t = \sum w_t (X_t - \overline{X}) = \sum w_t X_t - \overline{X}\sum w_t = \{1)\} = \sum w_t X_t$$

$$\Rightarrow \sum w_t x_t = \sum w_t X_t$$

3)
$$\sum w_t^2 = \sum \left[\frac{x_t}{\sum\limits_s x_s^2}\right]^2 = \frac{\sum x_t^2}{(\sum\limits_s x_s^2)^2} = \frac{1}{\sum\limits_s x_s^2}$$

4)
$$\sum w_t y_t = \sum (Y_t - \overline{Y}) = \sum w_t y_t - \overline{Y}\sum w_t = \{1)\} = \sum w_t Y_t$$

$$\Rightarrow \sum w_t y_t = \sum w_t Y_t.$$

Дисперсия оценки параметра b:

$$\mathrm{V}(\widehat{b}) = \mathrm{V}(\sum w_t y_t) = \mathrm{V}(\sum w_t Y_t) = \sum w_t^2 \sigma^2 = \frac{\sigma^2}{\sum x_t^2}$$

Дисперсия параметра a:

$$\widehat{a} = \overline{Y} - \overline{X}\widehat{b} = \overline{Y} - \overline{X}\sum w_t y_t = \overline{Y} - \overline{X}\sum w_t Y_t =$$

$$= \frac{1}{n}\sum_t Y_t - \overline{X}\sum w_t Y_t = \sum_t (\frac{1}{n} - \overline{X} w_t) Y_t$$

Понадобится еще одно тождество:

$$\sum x_t^2 = \sum_t X_t^2 - n\overline{X}^2 (*)$$

Докажем его:

$$\sum x_t^2 = \sum (X_t - \overline{X})^2 = \sum X_t^2 - 2\overline{X}\sum X_t + \overline{X}^2 n =$$

$$= \sum X_t^2 - 2n\overline{X}^2 + n\overline{X}^2 = \sum_t X_t^2 - n\overline{X}^2.$$

Тогда

$$\mathrm{V}(\widehat{a}) = \mathrm{V}(\sum_t (\frac{1}{n} - \overline{X} w_t) Y_t) = \sigma^2 \sum (\frac{1}{n} - \overline{X} w_t)^2 =$$

$$= \sigma^2 \sum (\frac{1}{n^2} - 2\frac{1}{n}\overline{X} w_t + \overline{X}^2 w_t^2) = \sigma^2 (\frac{1}{n} - 2\frac{1}{n}\overline{X}\sum w_t + \overline{X}^2 \sum w_t^2) =$$

$$= \sigma^2 (\frac{1}{n} + \frac{\overline{X}^2}{\sum x_i^2}) = \{(*)\} = \sigma^2 (\frac{1}{n} + \frac{\sum X_t^2 - \sum x_t^2}{n \sum x_t^2}) = \sigma^2 \frac{\sum X_t^2}{n \sum x_t^2}$$

3. Покажем, что МНК-оценки обладают наименьшей дисперсией в классе всех линейных несмещенных оценок.

   Пусть $\tilde{b} = \sum C_t Y_t$ - любая другая несмещенная оценка.

   Представим c_t в виде:

$$c_t = w_t + d_t.$$

Тогда в силу несмещенности $\widehat{b}$ и $\tilde{b}$:

$$\mathrm{E}(\tilde{b} - \widehat{b}) = 0 = \mathrm{E}((\sum w_t Y_t + \sum d_t Y_t) - \sum w_t y_t) =$$

$$= \mathrm{E}(\sum w_t Y_t + \sum d_t Y_t - \sum w_t y_t) = \mathrm{E}(\sum d_t Y_t) =$$

$$= \sum d_t (a + bX_t) \text{для всех } a \text{ и } b.$$

Тогда

$$\sum d_t = 0, \sum d_t X_t = \sum d_t x_t = 0 \text{ (проверить)}$$

$$\mathrm{V}(\tilde{b}) = \mathrm{V}(\sum C_t Y_t) = \sigma^2 \sum c_t^2 = \sigma^2 \sum (w_t + d_t)^2 =$$

$$= \sigma^2 (\sum w_t^2 + 2\sum w_t d_t + \sum d_t^2) = \sigma^2 (\sum w_t^2 + \sum d_t^2) =$$

$$= \mathrm{V}(\tilde{b}) + \sigma^2 \sum d_t^2.$$

$$\Rightarrow \mathrm{V}(\tilde{b}) \geq \mathrm{V}(\widehat{b})$$

Здесь использовали, что $\sum w_t d_t = 0$ (проверить!).

$$\sum w_t d_t = \frac{\sum x_t d_t}{\sum_s x_s^2} = \frac{0}{\sum_s x_s^2} = 0.$$

Аналогично, можно доказать, что $\mathrm{V}(\tilde{a}) \geq \mathrm{V}(\widehat{a})$.

□

С помощью Теоремы Гаусса-Маркова получили «наилучшие» оценки $\widehat{a}$ и $\widehat{b}$ коэффициентов регрессии a и b.

В регрессионном уравнении помимо коэффициентов регрессии a и b есть еще один параметр – дисперсия ошибок σ^2.

Пусть $\widehat{Y}_t = \widehat{a} + \widehat{b} X_t$ – прогноз значения Y_t в точке X_t. Остатки регрессии e_t определяются из уравнения

$$Y_t = \widehat{Y}_t + e_t = \widehat{a} + \widehat{b} X_t + e_t$$

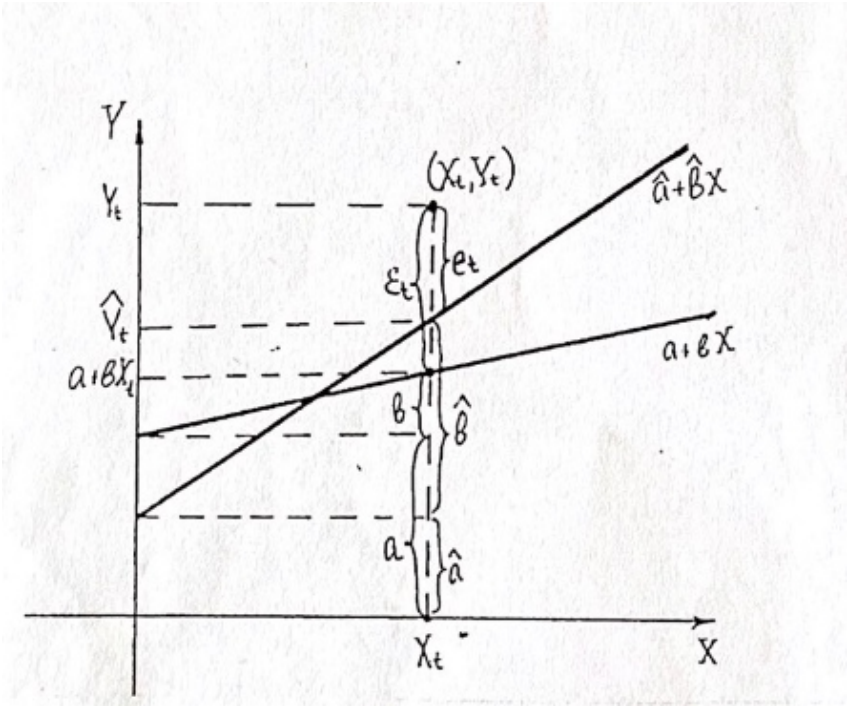
$$e_t = Y_t - \widehat{Y}_t$$

Остатки регрессии e_t – случайные величины, также как и ошибки регрессии ε_t в уравнении

$Y_t = a + bX_t + \varepsilon_t$. Тем не менее не следует их путать:

ε_t получается из уравнения истинной регрессии,

а e_t – из оценки истинной линии регрессии, т.е. e_t наблюдаемы. Гипотеза: между оцен-



кой σ^2 и суммой квадратов остатков регрессии $e_t = Y_t - \widehat{a} - \widehat{b}X_t$ существует связь.

Для её проверки надо рассмотреть $\mathrm{E}\sum_t e_t^2$.

$$\mathrm{E}\sum_t e_t^2 = (n-2)\sigma^2$$

Таким образом,

$$\boxed{s^2 = \widehat{\sigma}^2 = \frac{1}{n-2}\sum e_t^2}$$

является несмещенной оценкой дисперсии ошибок σ^2.

Дисперсии оценок $\widehat{a}, \widehat{b}$ коэффициентов регрессии, полученные при доказательстве Тео-

ремы Гаусса-Маркова могут быть вычислены в случае, если дисперсия ошибок σ^2 известна.

На практике дисперсия ошибок σ^2, как правило, не известна и рассчитывается одновременно с коэффициентами a, b.

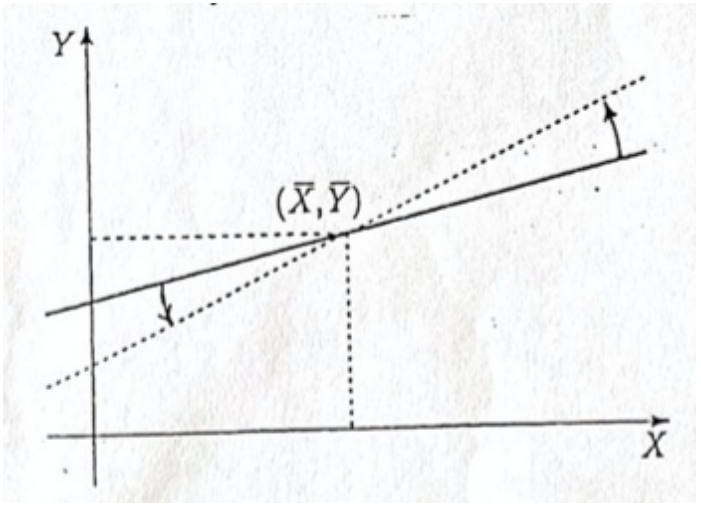
В этом случае могут быть получены только оценки дисперсий $\widehat{a}, \widehat{b}$ (при замене σ^2 на s^2):

$$\boxed{\widehat{V}(\widehat{b}) = s^2\frac{1}{\sum x_t^2} = \frac{s^2}{\sum (X_t - \overline{X})^2}},$$

$$\boxed{\widehat{V}(\widehat{a}) = s^2\frac{\sum X_t^2}{n\sum x_t^2} = \frac{s^2\sum X_t^2}{n\sum (X_t - \overline{X})^2}}$$

В статистических пакетах используются вышеуказанные формулы для нахождения стандартных отклонений оценок коэффициентов регрессии: $s_{\widehat{b}} = \sqrt{\widehat{V}(\widehat{b})}, s_{\widehat{a}} = \sqrt{\widehat{V}(\widehat{a})}$.

Поскольку график уравнения регрессии $Y = \widehat{a} + \widehat{b}X$ проходит через точку $(\overline{X}, \overline{Y})$, то при увеличении $\widehat{b}$ величина $\widehat{a}$ уменьшается.

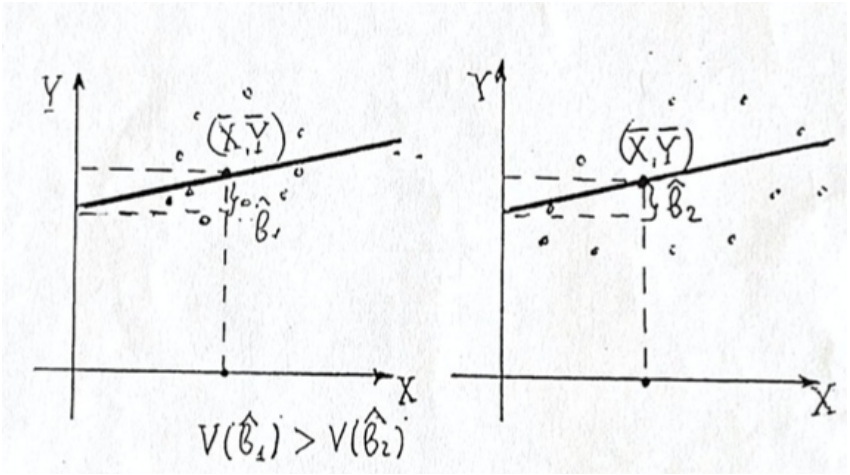


Замечание1. Предположим, что изучается зависимость Y от X и число наблюдений n задано, но можно выбирать набор значений X, т. е. $(X_1, X_2, \ldots, X_n)$.

Как выбрать X_t, чтобы точность оценки коэффициента b была наибольшей?

Из формулы для дисперсии оценки $\widehat{b}$ коэффициента регрессии видим, что чем больше $\sum x_t^2 = \sum (X_t - \overline{X})^2$, тем меньше дисперсия $V(\widehat{b})$.

Следовательно, если выбираем X_t с большим разбросом вокруг среднего значения, то получаем меньшую оценку дисперсии, т. е. большую точность оценки коэффициента регрессии b.



Доказательства фактов из лекции

Доказательство факта $\bar{x} = \bar{y} = 0$

$$x_t = X_t - \bar{X}, \quad y_t = Y_t - \bar{Y}, \quad \bar{x} = \bar{y} = 0$$

Доказательство

$$\bar{x} = \frac{1}{n}\sum_t (X_t - \bar{X}) = \frac{1}{n}\sum_t X_t - \frac{1}{n} n\bar{X} = 0,$$

$$\bar{y} = \frac{1}{n}\sum_t (Y_t - \bar{Y}) = \frac{1}{n}\sum_t Y_t - \frac{1}{n} n\bar{Y} = 0.$$

Для задачи минимизации функционала

$$F = \sum_{t=1}^{n} (y_t - (a + bx_t))^2 \to \min_{\forall a,b}$$

из МНК следует, что

$$\hat{a} = \bar{y} - \bar{x}b,$$

$$\hat{b} = \frac{n\sum x_t y_t - (\sum x_t)(\sum y_t)}{n\sum {x_t}^2 - (\sum x_t)^2} = \frac{n\sum x_t y_t - (n\bar{x})(n\bar{y})}{n\sum {x_t}^2 - (n\bar{x}^2)} = \frac{\sum x_t y_t}{\sum {x_t}^2}.$$

Семинар № 1

№ 2.1(1) $\widehat{\alpha}, \widehat{\beta}-?$

№ 2.9 (а) построить регрессию со свободным членом

Д/з:

1) № 2.1(1) $\widehat{\sigma}^2-?$

2) № 2.9 (б) построить регрессию без свободного члена

3) Доказать, что $\overline{x} = \overline{y} = 0$, $\widehat{a} = 0$, $\widehat{b} = \frac{\sum x_t y_t}{\sum x_t^2}$

4) Док-во теоремы Г-М.

2 лекция

Лекция 2

Статистические свойства
МНК-оценок параметров регрессии.
Проверка гипотезы $b=b_0$.
Доверительные интервалы для
коэффициентов регрессии

Основные гипотезы, лежащие в основе линейной регрессионной модели

1. $Y_t = a + bX_t + \varepsilon_t$, $t = 1,..., n$, — спецификация модели.
2. X_t — детерминированная величина; вектор $(X_1 ..., X_n)'$ не коллинеарен вектору $\iota = (1,...,1)'$.

3a. $E\varepsilon_t = 0$, $E(\varepsilon_t^2) = V(\varepsilon_t) = \sigma^2$ — не зависит от t.

3b. $E(\varepsilon_t \varepsilon_s) = 0$ при $t \neq s$, некоррелированность ошибок для разных наблюдений

3c. Ошибки ε_t, $t = 1,...,n$, имеют совместное нормальное распределение: $\varepsilon_t \sim N(0,\sigma^2)$.

Все выводы далее для условия нормальности. Т.е.

Пусть выполняется условие нормальной линейной регрессионной модели $\varepsilon \sim N(0,\sigma^2 I_n)$,
или, что то же самое, Y_t имеют совместное нормальное распределение.
Тогда МНК-оценки коэффициентов регрессии $\hat{a}$, $\hat{b}$ имеют совместное нормальное распределение,

$$\hat{a} \sim N\left(a, \sigma^2 \frac{\sum X_t^2}{n\sum x_t^2}\right), \quad \hat{b} \sim N\left(b, \sigma^2 \frac{1}{\sum x_t^2}\right)$$

Если гипотеза нормальности ошибок не выполняется, то
при некоторых условиях регулярности на поведение X_t при росте n
оценки $\hat{a}$, $\hat{b}$ имеют асимптотически нормальное распределение, т. е.

$$\hat{a} \underset{n\to\infty}{\text{к}} N\left(a, \sigma^2 \frac{\sum X_t^2}{n\sum x_t^2}\right), \quad \hat{b} \underset{n\to\infty}{\text{к}} N\left(b, \sigma^2 \frac{1}{\sum x_t^2}\right)$$

Распределение оценки дисперсии ошибок s^2

в случае когда ε — многомерная нормально распределенная случайная величина, т.е. $\varepsilon \sim N(0,\sigma^2 I_n)$, выполняется

$$\frac{(n-2)s^2}{\sigma^2} \sim \chi^2(n-2),\ s^2 = \hat{\sigma}^2$$

Независимость s^2 и МНК-оценок $\hat{a},\ \hat{b}$

Так как оценка дисперсии ошибок s^2 является функцией от остатков регрессии e_t. то для того чтобы доказать независимость s^2 и ($\hat{a}, \hat{b}$), достаточно доказать независимость e_t и ($\hat{a}, \hat{b}$).

Известно

два случайных вектора, имеющие совместное нормальное распределение, независимы тогда и только тогда, когда они некоррелированы.

некоррелированность e_t и ($\hat{a}, \hat{b}$).

$\mathrm{Cov}(e_t, \hat{b})=0$. ***Д/з*** ***$Corr(e_t, \hat{b})=0$***

$\mathrm{Cov}(e_t, \hat{a})=0$ ***$Corr(e_t, \hat{a})=0$***

Проверка гипотезы $b = b_0$

величина

$t = \dfrac{\hat{b}-b_0}{s_{\hat{b}}} \sim t(n-2)$ распределена по закону Стьюдента если $\varepsilon \sim N(0,\sigma^2 I_n)$ **(*)**

$t = \dfrac{\hat{a}-a_0}{s_{\hat{a}}} \sim t(n-2)$.

Статистику **(*)** можно использовать для проверки гипотезы H_0: $b=b_0$ против альтернативной гипотезы H_1: $b \neq b_0$.

Предположим, что верна гипотеза H_0, и зададимся, например, 95%-квантилью t-распределения с *(n-2)* степенями свободы t_c (при 40 степенях свободы t_c=2.021), т.е.

$$P\left\{-t_c < \frac{\hat{b}-b_0}{s_{\hat{b}}} < t_c\right\} = 0.95$$

т.е. H_0 принимаем, если $|t| \le t_c$ на 5% уровне значимости

Мы отвергаем гипотезу H_0 (и принимаем H_1) на 5%-уровне значимости, если $|t| > t_c$ («редкое» событие с точки зрения гипотезы H_0).

Разрешив неравенство в $P\left\{\left|(\hat{b}-b)/s_{\hat{b}}\right| < t_c\right\} = 0.95$ относительно b, получим

$$P\left\{\hat{b} - t_c s_{\hat{b}} < b < \hat{b} + t_c s_{\hat{b}}\right\} = 0.95,$$

т.е. $\left[\hat{b} - t_c s_{\hat{b}}\ ,\ \hat{b} + t_c s_{\hat{b}}\right]$— 95%-*доверительный интервал* для b.

при гипотезе H_0: $b=0$, ***- значим регрессор при b или нет?***

t-статистика $t = \hat{b} / s_{\hat{b}}$

Малые значения t-статистики соответствуют отсутствию достоверной статистической связи объясняющей переменной X и зависимой переменной Y.

Анализ вариации зависимой переменной в регрессии. Коэффициент детерминации R^2

Анализ вариации зависимой переменной в регрессии

Рассмотрим вариацию (разброс) $\Sigma(Y_t - \bar{Y})^2$ значений Y_t вокруг среднего значения. Разобьем эту вариацию на две части: объясненную регрессионным уравнением и не объясненную (т.е. связанную с ошибками ε_t).

Обозначим через $\hat{Y}_t = \hat{a} + \hat{b}X_t$ предсказанное значение Y_t, тогда $Y_t - \bar{Y} = (Y_t - \hat{Y}_t) + (\hat{Y}_t - \bar{Y})$ (см. рис. 2.5) и вариация Y_t представляется в виде трех слагаемых:

$$\sum (Y_t - \bar{Y})^2 = \sum (Y_t - \hat{Y}_t)^2 + \sum (\hat{Y}_t - \bar{Y})^2 + 2\sum (Y_t - \hat{Y}_t)(\hat{Y}_t - \bar{Y}) \qquad (2.25)$$

Третье слагаемое в (2.25) равно нулю, так как $y - \hat{y} = e_t$ — вектор остатков регрессии, ортогонален константе ι и вектору x (см.(2.7)).

В самом деле $\sum e_t(\hat{Y}_t - \bar{Y}) = \sum e_t(\hat{a} + \hat{b}X_t - \bar{Y}) = (\hat{a} + \hat{b}\bar{X} - \bar{Y})\sum e_t + \hat{b}\sum e_t x_t = 0$.

Поэтому верно равенство

$$\underset{TSS}{\sum (Y_t - \bar{Y})^2} = \underset{ESS}{\sum (Y_t - \hat{Y}_t)^2} + \underset{RSS}{\sum (\hat{Y}_t - \bar{Y})^2} \qquad (2.26)$$

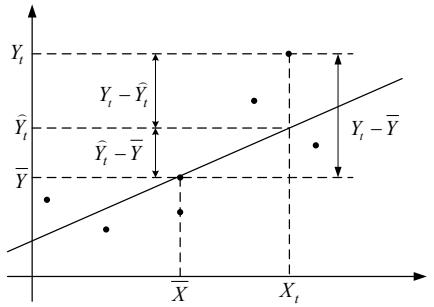


Рис. 2.5

Обозначим левую часть в (2.26) через TSS *(total sum of squares)* — вся

дисперсия, первое слагаемое в правой части, соответствующее не объясненной дисперсии, через ESS *(error sum of squares*), второе слагаемое в правой части — RSS *(regression sum of squares)* — объясненная часть всей дисперсии[1].

Статистика R^2 — коэффициент детерминации

Определение. Коэффициентом детерминации, или долей объясненной дисперсии, называется

$$R^2 = 1 - \frac{ESS}{TSS} = \frac{RSS}{TSS} \qquad (2.27)$$

Заметим, что второе равенство в (2.27) верно лишь в том случае, если верно (2.26), т.е. когда константа включена в уравнение регрессии. Только в этом случае имеет смысл рассматривать статистику R^2.

В силу определения R^2 принимает значения между . 0 и 1, $0 \le R^2 \le 1$.

Если $R^2 = 0$, то это означает, что регрессия ничего не дает, т. е. X_t не улучшает качество предсказания Y_t по сравнению с тривиальным предсказанием $\hat{Y}_t = \overline{Y}$.

Другой крайний случай R^2=1 означает точную подгонку: все точки наблюдений лежат на регрессионной прямой (все $e_t = 0$).

Чем ближе к 1 значение R^2, тем лучше качество подгонки, $\hat{y}$ более точно аппроксимирует y.

F-статистика

Снова предположим, что мы находимся в рамках нормальной линейной регрессионной модели.

$$F = \frac{\left(\frac{\hat{b}-b}{\sigma_{\hat{b}}}\right)^2 \frac{1}{1}}{\frac{\sum e_t^2}{\sigma^2}\frac{1}{n-2}} = \frac{(\hat{b}-b)^2 \sum x_t^2}{\sum e_t^2/(n-2)} \sim \frac{\frac{1}{1}\chi^2(1)}{\frac{1}{n-2}\chi^2(n-2)} = F(1, n-2) \qquad (2.28)$$

Полученную F-статистику можно использовать для проверки нулевой гипотезы H_0: b-b_0=0. При этой гипотезе статистика (2.28) выглядит следующим образом:

$$F = \frac{(\hat{b}-b_0)^2 \sum x_t^2}{\sum e_t^2/(n-2)} \sim F(1, n-2) \qquad (2.29)$$

Если нулевая гипотеза справедлива, то значение F в (2.29) мало.

***F* ≤ F_α(1,n-2) –H_0 приним.**

Таким образом, мы отвергаем нулевую гипотезу, если F превосходит критическое значение $F_a(1, n-2)$ распределения Фишера с параметрами (1, n-2) для выбранного уровня значимости α.

Статистика (2.29) особенно просто выглядит для гипотезы H_0: b=0 (случай отсутствия линейной функциональной связи между X и Y).

$$F = \frac{\hat{y}_\bullet' \hat{y}_\bullet}{e'e/(n-2)} \qquad \hat{y}_\bullet' \hat{y}_\bullet = \sum (\hat{y}_t)^2 \qquad e'e = \sum e_t^2$$

связывающее R^2 и F-статистики соотношение

$$F = (n-2)\frac{R^2}{1-R^2}.$$

Не удивительно, что малым значениям F (отсутствие значимой функциональной связи X и Y) соответствуют малые значения R^2 (плохая аппроксимация данных).

Оценка максимального правдоподобия коэффициентов регрессии

Оценка максимального правдоподобия

Рассмотрим его применение к оцениванию парной регрессии. Предположим, что мы ищем параметры нормальной линейной регрессионной модели

$$Y_i = a + bX_i + \varepsilon_i. \qquad (1)$$

Ошибки регрессии ε_i независимы и распределены по нормальному закону:

$$\varepsilon_i \sim N(0, \sigma^2), \qquad (2)$$

или, что является эквивалентной записью,

$$Y_i \sim N(a + b, \sigma^2)$$

Имея набор наблюдений $(X_i, Y_i), t = 1, \dots, n$, мы можем попытаться ответить на вопрос: при каких значениях параметров a, b, σ^2 модели (1)-(2) вероятность получить этот набор наблюдений наибольшая? Другими словами, каковы наиболее вероятные значения параметров модели для данного набора наблюдений?

Чтобы ответить на этот вопрос составим функцию правдоподобия, равную произведению плотностей вероятности отдельных наблюдений (мы считаем все ε_t независимыми):

$$L(Y_1, \dots, Y_n, a, b, \sigma^2) = p(Y_1, \dots, Y_n | X_1, \dots, X_n, a, b, \sigma^2) = \prod_{t=1}^{n} p(Y_t)$$

$$= (2\pi)^{-\frac{n}{2}} (\sigma^2)^{-\frac{n}{2}} exp\left(-\frac{1}{2\sigma^2}\sum (Y_i - a - bX_t)^2\right), \quad (3)$$

где p обозначает плотность вероятности, зависящую от X_t, Y_t и параметров a, b, σ^2.

Для того чтобы найти наиболее правдоподобные значения параметров, нам необходимо найти такие же значения, при которых функция правдоподобия L достигает своего максимума.

$$L(Y_1, \dots, Y_n, a, b, \sigma^2) \to \max_{a,b,\sigma^2}$$

Так как функции L и $\ln L$ одновременно достигают своего максимума, достаточно искать максимум логарифма функции правдоподобия

$$\ln L\,(Y_1, \dots, Y_n, a, b, \sigma^2) \to \max_{a,b,\sigma^2}$$

$$\ln L\,(Y_1, \dots, Y_n, a, b, \sigma) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum (Y_i - a - bX_t)^2.$$

Необходимые условия экстремума функции $\ln L$ имеют вид:

$$\frac{\partial \ln L}{\partial a} = \frac{1}{\sigma^2}\sum (Y_t - a - bX_t) = 0, \quad (4)$$

$$\frac{\partial \ln L}{\partial b} = \frac{1}{\sigma^2}\sum X_t (Y_t - a - bX_t) = 0, \quad (5)$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{\sigma^4}\sum (Y_t - a - bX_t)^2 = 0. \quad (6)$$

Решением системы уравнений (4), (5), (6) являются оценки максимального правдоподобия

$$\hat{b}_{ML} = \frac{\sum x_t y_t}{\sum x_t^2};\ \hat{a}_{ML} = \bar{Y} - \hat{b}_{ML}\bar{X};\ \hat{\sigma}^2_{ML} = \frac{1}{n}\sum e_t^2.$$

Отметим, что оценки максимального правдоподобия параметров a, b совпадают с оценками метода наименьших квадратов $\hat{a}_{ML} = \hat{a}_{OLS}, \hat{b}_{ML} = \hat{b}_{OLS}$.

Оценка максимального правдоподобия для σ^2 не совпадает с $\sigma^2_{OLS} = \sum \frac{e_t^2}{n-2}$, которая, является несмещенной оценкой дисперсии ошибок. Таким образом, $\hat{\sigma}^2_{ML} = ((n-2)/n)\hat{\sigma}^2_{OLS}$ является смещенной, но тем не менее состоятельной оценкой σ^2.
То есть при увеличении объема выборки гарантированно приближается к истинному значению параметра.

Семинар №2
№2.1 (2)
№2.9 (а) (б)
Д/З №2.1 (2) проверить гипотезу, но a=0.
Доказательства по тексту
+ тема 1 №1
Тема 3 №5

Тема 1

П. 1.1. Количественные задачи

1. Найдете методом максимального оценку параметра p (p - вероятность наступления события A) биномиального распределения $P_n(k) = C_n^k p^k (1-p)^{n-k}$, если в n независимых испытаниях событие A появилось x=m раз. Запишите оценку параметра p для случая n=50. m = 41.

   Сумма баллов: 10

   Решение.
   Функция правдоподобия:

   $$L(x, p) = P_n(m) = C_n^m p^m (1-p)^{(n-m)}.$$

Логарифм функции правдоподобия:

$$\ln L(x,p) = lnC_n^m + m\ln p + (n-m)\ln(1-p).$$

Первая частная производная по p:

$$\frac{\partial \ln L(x,p)}{\partial p} = \frac{m}{p} - \frac{n-m}{1-p}.$$

Приравняем первую частную производную нулю, чтобы найти точку экстремума: $p = \frac{m}{n}$.

Проверим, что эта точка является точкой максимума функции правдоподобия:

$$\frac{\partial^2 \ln L(x,p)}{\partial p^2} = -\frac{m}{p^2} - \frac{n-m}{(1-p)^2} < 0$$

Следовательно, оценка максимального правдоподобия параметра p:

$$\hat{p} = \frac{m}{n} = \frac{41}{50}$$

2. Гипотеза о том, свободный член уравнения регрессии равен 6, не может быть отвергнута.
3. Имеются данные о годовых объемых продаж товара (Y) и расходах на рекламу товара (X) по 24 магазинам. На основании этих данных с помощью метода наименьших квадратов была построена парная регрессионная модель:

$$\widehat{Y_t} = 6{,}331 + 0{,}336X_t$$

Объясненная регриссонным уравнением часть дисперсии $RSS = 8235{,}69$, не объясненная регрессионным уравнением часть дисперсии $ESS = 95{,}408$. Найдите коэффициент детерминации R^2, и протестируйте значимость регрессии с уровнем значимости 5%.

Сумма баллов: 5

Решение.

1) $TSS = RSS + ESS$

$$R^2 = \frac{RSS}{TSS} = \frac{8235{,}69}{8331{,}098} = 0{,}989$$

2) $H_0: \hat{\beta} = 0$.

$$F = (n-2)\frac{R^2}{1-R^2} = 22 \times \frac{0{,}989}{1-0{,}989} = 1889{,}056$$

95% критическое значение $F-$ статистики c $(1, n-2)$ степенями свободы:

$$F_{0,95}(1{,}22) = 4{,}30.$$

$F > F_{0,95}(1{,}22)$. Следовательно, регрессия статически значима на 5% уровне значимости.

Ответ:
1. $R^2 = 0{,}989$
2. Регрессия статически значима.
3. Вопросы качественного характера
   Дана модель парной регрессии $Y_t = a + bX_t + \varepsilon_t, t = \overline{1,n}$, для которой выполнены условия теоремы Гаусса-Маркова: X_t - детерминированные

Модель множественной регрессии

Перейдем к модели n-мерной регрессии:

$$y_t = \beta_1 + \beta_2 x_{t2} + ... + \beta_k x_{tk} + \varepsilon_t, \;\; t = 1, ..., n$$

или

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + ... + \beta_k x_{tk} + \varepsilon_t, \;\; t = 1, ..., n$$

где x_{tp} — значения регрессора x_p в наблюдении t, а $x_{t1} = 1, t = 1, ..., n$. С учетом этого замечания мы не будем далее различать модели со свободным членом и без свободного члена.

Основные гипотезы

1. $y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + ... + \beta_k x_{tk} + \varepsilon_t, \;\; t = 1, ..., n$ — спецификация модели.
2. $x_{t1}, ..., x_{tk}$ — детерминированные величины. Векторы $x_s = (x_{1s}, ..., x_{ns})', s = 1, ..., k$ линейно независимы в R^n.
3. a. $\mathbb{E}\varepsilon_t = 0, \mathbb{E}(\varepsilon_t^2) = V(\varepsilon_t) = \sigma^2$ — не зависит от t.
4. b. $\mathbb{E}(\varepsilon_t \varepsilon_s) = 0$ при $t \neq s$ — статистическая независимость (некоррелированность) ошибок для разных наблюдений. Часто добавляется следующее условие.
5. c. Ошибки $\varepsilon_t, t = 1, ..., n$ имеют совместное нормальное распределение: $\varepsilon_t \; N(0, \sigma^2)$.

В этом случае модель называется *нормальной линейной регрессионной.*

Основные гипотезы в матричном виде:

Пусть y обозначает $n \times 1$ матрицу (вектор-столбец) $(y_1, ..., y_n)', \beta = (\beta_1, ..., \beta_k)'$ — $k \times 1$ вектор коэффициентов; $\varepsilon = (\varepsilon_1, ..., \varepsilon_n)'$ — $n \times 1$ вектор ошибок; X — $n \times k$ матрица объясняющих переменных.

Столбцами матрицы X являются векторы регрессоров $x_s = (x_{1s}, ..., x_{ns})', s = 1, ..., k$. Условия 1–3 в матричной записи выглядят следующим образом:

1. $y = X\beta + \varepsilon$ — спецификация модели;
2. X — детерминированная матрица, имеет максимальный ранг k;
3. a,b. $\mathbb{E}(\varepsilon) = 0; V(\varepsilon) = \mathbb{E}(\varepsilon\varepsilon') = \sigma^2 I_n$;
4. c. (дополнительное условие) $\varepsilon \; N(0, \sigma^2 I_n)$, т.е. ε — нормально распределенный случайный вектор со средним 0 и матрицей ковариаций $\sigma^2 I_n$ (нормальная линейная регрессионная модель).

Практическое применение условия 2: $n \geq k$ т.е. нужно иметь достаточное количество наблюдений (данных) для проведения регрессии: количество наблюдений должно быть не меньше количества регрессоров.

Метод наименьших квадратов. Теорема Гаусса-Маркова

Целью метода является выбор вектора оценок $\widehat{\beta}$, минимизирующего сумму квадратов остатков e_t (не объясненную часть дисперсии).

$$e = y - \widehat{y} = y - X\widehat{\beta},$$

$$ESS = \sum e_t^2 = e'e \to \min_{\widehat{\beta}}$$

Выразим $e'e$ через X и β:

$$e'e = y'y - 2\widehat{\beta}'X'y + \widehat{\beta}'X'X\widehat{\beta}.$$

Необходимые условия минимума ESS получаются дифференцированием по вектору $\widehat{\beta}$

$$\frac{\partial ESS}{\partial \widehat{\beta}} = -2X'y + 2X'X\widehat{\beta} = 0, (*)$$

откуда, учитывая обратимость матрицы $X'X$ находим оценку метода наименьших кважратов:

$$\widehat{\beta}_{OLS} = (X'X)^{-1}X'y.$$

Из (*) следует, что вектор остатков e ортогонален всем независимым пересенным $x_1, ..., x_k$ (столбцам матрицы X).

$$X'e = 0.$$

$X'e = X'(y - X\widehat{\beta}) = X'y - X'X\widehat{\beta} = X'y - X'X(X'X)^{-1}X'y = 0.$

Получим еще одну формулу для суммы квадратов остатков:

$$e'e = y'y - 2\widehat{\beta}'X'y + \widehat{\beta}'X'X\widehat{\beta} = y'y - \widehat{\beta}'(2X'y - X'X(X'X)^{-1}X'y) = y'y - \widehat{\beta}'X'y,$$

т.е.

$$e'e = y'y - \widehat{\beta}'X'y.$$

Доказательство:

1) $\widehat{\beta}'$ — несмещенная оценка $\widehat{\beta}$

$\mathbb{E}\widehat{\beta} = \mathbb{E}[(X'X)^{-1}X'y] = (X'X)^{-1}X'\mathbb{E}y = (X'X)^{-1}X'\mathbb{E}(X\beta + \varepsilon) = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'\mathbb{E}\varepsilon = \beta.$

Пусть $A = (X'X)^{-1}X'$

Тогда $\widehat{\beta} = Ay$

Пусть b — любая другая несмещенная оценка β. Тогда можно представить b в виде

$$b = (A + C)y, C\ k \times n$$

Так как b — несмещенная оценка β, то $\beta = \mathbb{E}b = (A + C)\mathbb{E}y = (A + C)\mathbb{E}(X\beta + \varepsilon) = (A + C)X\beta + (A + C)\mathbb{E}\varepsilon = (A + C)X\beta = (AX + CX)\beta = (I + CX)\beta \Rightarrow CX = 0$.

2) Вычислим матрицу ковариаций $\widehat{\beta}$:

$V(\widehat{\beta}) = V(Ay) = AV(y)A' = A\sigma^2 IA' = \sigma^2(X'X)^{-1}X'X(X'X)^{-1} = \sigma^2(X'X)^{-1}$

3) $V(b) = V(\widehat{\beta}) + \sigma^2 CC', CC' \geq 0$.

$\Rightarrow V(b) \geq V(\widehat{\beta})$.

Анализ вариации зависимой переменной в регрессии. Коэффициенты R^2 и скорректированный R^2_{adj}

Как и в случае регрессионной модели с одной независимой переменной, вариацию $\sum(y_t - \overline{y})^2$ можно разбить на две части: объясненную регрессионным уравнением и необясненную (т.е. связанную с ошибками ϵ)

$$\sum(y_t - \overline{y})^2 = \sum(y_t - \widehat{y}_t)^2 + \sum(\widehat{y}_t - \overline{y})^2 + 2\sum(y_t - \widehat{y}_t)(\widehat{y}_t - \overline{y}),$$

или в векторной форме:

$$(y - \overline{y}i)'(y - \overline{y}i) = (y - \widehat{y})'(y - \widehat{y}) + (\widehat{y} - \overline{y}i)'(\widehat{y} - \overline{y}i) + 2(y - \widehat{y})'(\widehat{y} - \overline{y}i)$$

Третье слагаемое равно нулю в случае, если константа, т.е. вектор $i = (1, ..., 1)'$, принадлежит линейной оболочке векторов $x_1, ..., x_k$, т.е.

$$2\sum(y_t - \widehat{y}_t)(\widehat{y}_t - \overline{y}) = 0.$$

Таким образом,

$$\underset{\text{TSS}}{||y - \overline{y}i||^2} = \underset{\text{ESS}}{||y - \widehat{y}||^2} + \underset{\text{RSS}}{||\widehat{y} - \overline{y}i||^2}.$$

в отклонениях: $y'_{\cdot}y_{\cdot} = e'e + \widehat{y}'_{\cdot}\widehat{y}_{\cdot}$, где $y_{\cdot} = y - \overline{y}i$; $\widehat{y}_{\cdot} = \widehat{y} - \overline{y}i$;

Как и ранее определим коэффициент детерминации R^2 как

$$R^2 = 1 - \frac{\text{ESS}}{\text{TSS}} = 1 - \frac{e'e}{y'_{\cdot}y_{\cdot}} = \frac{\widehat{y}'_{\cdot}\widehat{y}_{\cdot}}{y'_{\cdot}y_{\cdot}} = \frac{\text{RSS}}{\text{TSS}}.$$

Свойства R^2

$$R^2 \in [0; 1],$$

если вектор $i = (1, ..., 1)'$ принадлежит линейной оболочке векторов $x_1, ..., x_k$.

Коэффициент R^2 показывает качество подгонки регрессионной модели к наблюденным значениям y_t.

Если $R^2 = 0$, то регрессия y на $x_1, ..., x_k$ не улучшает качество предсказания y_t по сравнению с тривиальным предсказанием $\widehat{y}_t = \overline{y}$.

Другой крайний случай $R^2 = 1$ означает точную подгонку: все $e_t = 0$, т.е. все точки наблюдений удовлетворяют уравнению регрессии.

1. R^2, вообще говоря, возрастает при добавлении еще одного регрессора.
2. R^2 изменяется даже при простейшем преобразовании зависимой переменной, поэтому сравнивать по значению R^2 можно только регрессии с одинаковыми зависимыми переменными:

Пример.

1. $y = X\beta + \varepsilon$.
2. $z = y - x_1 = X\gamma + \varepsilon$.

обоим уравнениям соответствует одна и та же геометрическая картинка и экономически содержательная ситуация. Однако коэффициенты R^2 не совпадают только потому, что зависимость сформулирована в разных координатах.

Если взять число регрессоров равным числу наблюдений, всегда можно добиться того, что $R^2 = 1$, но это вовсе не будет означать наличие содержательной (имеющей экономический смысл) зависимости y от регрессоров.

Скорректированный коэффициент детерминации R^2_{adj}

Попыткой устранить эффект, связанный с ростом R^2 при возрастании числа регрессоров, является коррекция R^2 на число регрессоров. *Скорректированным (adjusted)* R^2 называется

$$R^2_{adj} = 1 - \frac{e'e/(n - k)}{y'_{\cdot}y_{\cdot}/(n - 1)}.$$

Заметим, что нет никакого существенного оправдания именно такого способа коррекции.

Свойства скорректированного R^2:

1. $R^2_{adj} = 1 - (1 - R^2)\frac{(n-1)}{(n-k)}$.
2. $R^2 \geq R^2_{adj}, k > 1$.
3. $R^2_{adj} \leq 1$, но может принимать значения < 0.

В определенной степени использование скорректированного коэффициента детерминации R^2_{adj} более корректно для сравнения регрессий при изменении количества регрессоров.

Что "лучше": y или $\widehat{y}$?

В качестве значений зависимой переменной в момент t мы можем использовать y_t или, например, прогноз y_t. Так как

$$V(y) \geq V(\widehat{y}),$$

или

$$V(y_t) \geq V(\widehat{y}_t),$$

то в качестве значения зависимой переменной зачастую лучше брать предсказанное по модели значение, а не фактически наблюденное. При этом, естественно, предполагается, что наблюдаемые значения y_t действительно удовлетворяют соотношению $y = X\beta + \varepsilon$, т.е. порождаются рассматриваемой моделью.

Проверка статистических гипотез. Доверительные интервалы и доверительные области

Проверка гипотезы $\mathrm{H}_0 : \beta_i = \beta_{i0}$.

В рамках нормальной модели $\varepsilon \sim N(0; \sigma^2 I_n)$ регрессор x_i несущественный.

Результаты:

1. $\widehat{\beta}_{OLS} \sim N(\beta; \sigma^2 (X'X)^{-1})$
2. Случайная величина $(n-k)\frac{s^2}{\sigma^2} \sim \chi^2(n-k)$
3. Оценки $\widehat{\beta}_{OLS}$, s^2 и вектор e независимы

Отсюда получаем

$$t = \frac{\widehat{\beta}_{OLS,i} - \beta_{i0}}{s_{\widehat{\beta}_i}} = \frac{(\widehat{\beta}_{OLS,i} - \beta_{i0})/\sigma_{\widehat{\beta}_i}}{s_{\widehat{\beta}_i}/\sigma_{\widehat{\beta}_i}} \sim t(n-k),$$

где $s^2_{\widehat{\beta}_i} = \sigma^2_{\widehat{\beta}_i} = \widehat{\sigma}^2 q^{ii} = s^2 q^{ii}$, где q^{ii} — i-ый диагональный элемент матрицы $(X'X)^{-1}$, $s^2 = \frac{\sum e_i^2}{n-k}$.

$[\widehat{\beta}_{OLS,i} - t_c s_{\widehat{\beta}_i};\ \widehat{\beta}_{OLS,i} + t_c s_{\widehat{\beta}_i}]$ является 95%-*доверительным интервалом* для истинного значения коэффициента β_i, где t_c — двусторонняя 95%-квантиль распределения Стьюдента с $n-k$ степенями свободы.

Нулевая гипотеза $\mathrm{H}_0 : \beta_i = \beta_{i0}$ отклоняется на 5%-уровне значимости, если $|t| = \left|\frac{\widehat{\beta}_{OLS,i} - \beta_i}{s_{\widehat{\beta}_i}}\right| > t_c(n-k)$

Проверка гипотезы $H_0 : \beta_2 = \beta_3 = ... = \beta_k = 0$.

Пусть дана модель регрессии со свободным членом

$$y_t = \beta_1 + \beta_2 x_{t2} + ... + \beta_k x_{tk} + \varepsilon_t.$$

Нулевая гипотеза состоит в том, что коэффициенты при всех регрессорах равны нулю.

Статистика F имеет распределение Фишера

$$F = \frac{R^2}{1-R^2}\frac{n-k}{k-1} = \frac{\mathrm{RSS}}{\mathrm{ESS}}\frac{n-k}{k-1} = \frac{\widehat{y}'\widehat{y}./(k-1)}{e'e/(n-k)} \sim F(k-1, n-k)$$

и ее можно использовать для проверки гипотезы H_0.

Гипотеза H_0 отвергается на 5%-уровне значимости, если $F > F_c$, где F_c — односторонняя 95%-квантиль распределения Фишера $F(k-1, n-k)$.

Линейное ограничение общего вида $\mathrm{H}_0 : H\beta = r$. Пусть H — $q \times k$ матрица, β — $k \times 1$ вектор коэффициентов, r — $q \times 1$ вектор.

Естественно считать, что число ограничений не превосходит числа параметров и ограничения линейно независимы, т.е. $q \leq k$ и матрица H имеет максимальный ранг: $\mathrm{rank}(H) = q$.

Пример. В качестве примера рассмотрим следующие матрицы H, r для $k = 3, q = 2$:

$$H\beta = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix} = r.$$

Это условие соответствует системе двух линейных ограничений

$$\begin{cases} \beta_1 = 2 \\ \beta_2 - \beta_3 = 0 \end{cases}$$

$$F = \frac{(H\widehat{\beta} - r)'(H(X'X)^{-1}H')^{-1}(H\widehat{\beta} - r)/q}{e'e/(n-k)} \sim F(q, n-k).$$

Если справедлива гипотеза $\mathrm{H}_0 : H\beta - r = 0$, то статистика F не должна принимать слишком больших значений, а именно, с 95%-вероятностью $F < F_c(q, n-k)$, где $F_c(q, n-k)$ есть 95%-квантиль распределения Фишера $F(q, n-k)$.

Другой вид F-статистики:

$$F = \frac{(\widehat{\beta} - \beta)'H'(H(X'X)^{-1}H')^{-1}H(\widehat{\beta} - \beta)/q}{e'e/(n-k)} \sim F(q, n-k)$$

Условие $F < F_c(q, n-k)$ задает 95%-*доверительную область* для коэффициентов β (выпуклую).

В случае $H = I$ статистика F выглядит следующим образом:

$$F = \frac{(\widehat{\beta} - \beta)'(X'X)(\widehat{\beta} - \beta)/k}{e'e/(n-k)} \sim F(k, n-k).$$

В этом случае доверительная область является эллипсоидом в k-мерном пространстве коэффициентов β.

$\mathrm{H}_0 : \beta_{k-q+1} = \beta_{k-q+2} = ... = \beta_k = 0$. Гипотеза является, конечно, частным случаем общей линейной гипотезы $H\beta = r$.

$$F = \frac{(e_\cdot'e_\cdot - e'e)/q}{e'e/(n-k)} = \frac{(\mathrm{ESS_R} - \mathrm{ESS_{UR}})/q}{\mathrm{ESS_{UR}}/(n-k)} \sim F(q, n-k).$$

Здесь $\mathrm{ESS_R}$ — сумма квадратов остатков "короткой"(*restricted*) регрессии; $\mathrm{ESS_{UR}}$ — сумма квадратов остатков "длинной"(*unrestricted*) регрессии.

Как и ранее, F-статистику можно выразить через коэффициенты детерминации R^2 для "короткой"и "длинной"регрессий:

$$F = \frac{(R^2_{\mathrm{UR}} - R^2_{\mathrm{R}})/q}{(1 - R^2_{\mathrm{UR}})/(n-k)} \sim F(q, n-k)$$

При $F < F_c(q, n-k) \Rightarrow \mathrm{H_0}$ о "короткой"регрессии принимается, т.е. использование q дополнительных регрессоров ничего не добавляет к качеству оценивания модели.

$\mathrm{H_0} : \beta' = \beta''; \sigma' = \sigma''$ *(тест Чоу (Chow))*. Предположим, у нас есть две выборки данных. По каждой выборке мы строим регрессионную модель. Вопрос, который нас интересует: верно ли, что эти две модели совпадают? Рассмотрим модели:

$$y_t = \beta'_1 x_{t1} + \beta'_2 x_{t2} + ... + \beta'_k x_{tk} + \varepsilon'_t, \quad t = 1, ..., n, \tag{1}$$

$$y_t = \beta''_1 x_{t1} + \beta''_2 x_{t2} + ... + \beta''_k x_{tk} + \varepsilon''_t, \quad t = n+1, ..., n+m, \tag{2}$$

в первой выборке n наблюдений и m наблюдений во второй. Например, y — заработная плата, x_t — регрессоры (возраст, стаж, уровень образования и т.п.), и пусть первая выборка относится к женщинам, вторая — к мужчинам. Вопрос: следует ли из оценки моделей (1), (2), что модель зависимости зарплаты от регрессоров одна и та же для мужчин и женщин?

Сведем эту ситуацию к общей схеме проверки линейных ограничений на параметры модели. Регрессией без ограничений здесь является объединение двух регрессий (1), (2), т.е. $\mathrm{ESS_{UR}} = \mathrm{ESS_1} + \mathrm{ESS_2}$, число степеней свободы при этом равно $(n-k)+(m-k) = n+m-2k$. Предположим теперь, что верна нулевая гипотеза. Тогда регрессия с ограничениями записывается одним уравнением

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + ... + \beta_k x_{tk} + \varepsilon_t, \quad t = 1, ..., n+m. \tag{3}$$

Оценивая (3), получаем $\mathrm{ESS_R}$. Тогда, учитывая, что наложено k ограничений на параметры модели, получаем

$$F = \frac{(\mathrm{ESS_R} - \mathrm{ESS_{UR}})/k}{\mathrm{ESS_{UR}}/(n+m-2k)} \sim F(k, n+m-2k). \tag{4}$$

Если F-статистика (4) больше критического значения F_c, то нулевая гипотеза отвергается. В этом случае мы не можем объединять две выборки в одну.

В противном случае $(F < F_c(k, n+m-2k)) \Rightarrow$ принимается $\mathrm{H_0}$ и можно использовать объединенную регрессию (3).

Лекция № 4

Октябрь 2020

1 Различные аспекты множественной регрессии

Рассмотрим некоторые проблемы, которые часто возникают при практическом использовании регрессионных моделей.

На практике исследователю нередко приходится сталкиваться с ситуацией, когда полученная им регрессия является “плохой“, т.е. t-статистики большинства оценок малы, что свидетельствует о незначимости соответствующих независимых переменных (регрессоров). В то же время F-статистика может быть достаточно большой, что говорит о значимости регрессии в целом. Одна из возможных причин такого явления носит название *мультиколлинеарности* и возникает при наличии высокой корреляции между регрессорами.

Регрессионные модели являются достаточно гибким инструментом, позволяющим, в частности, оценивать влияние качественных признаков (пол, профессия, наличие детей и т.п.) на изучаемую переменную. Это достигается введением в число регрессоров так называемых *фиктивных переменных*, принимающих, как правило, значения 1 или 0 в зависимости от наличия или отсутствия соответствующего признака в очередном наблюдении. С формальной точки зрения фиктивные переменные ничем не отличаются от других регрессоров. Однако следует обратить особое внимание на правильное их использование и интерпретацию оценок.

1.1 Мультиколлинеарность

Одним из условий классической регрессионной модели является предположение о линейной независимости объясняющих переменных, что означает *линейную независимость столбцов матрицы регрессоров* X или (эквивалентно) что *матрица* $(X'X)^{-1}$ *имеет полный ранг* k. При нарушении этого условия, т.е. когда один из столбцов матрицы X есть линейная комбинация остальных столбцов, говорят, что имеет место *полная коллинеарность*. В этой ситуации нельзя построить МНК-оценку параметра β, что формально следует из сингулярности матрицы $X'X$ и невозможности решить нормальные уравнения. Нетрудно также понять и содержательный смысл этого

явления. Рассмотрим следующий простой пример регрессии (Greene, 1997):

$$C = \beta_1 + \beta_2 S + \beta_3 N + \beta_4 T + \epsilon$$

, где C - потребление, S - зарплата, N - доход, получаемый вне работы, T - полный доход. Поскольку выполнено равенство $T = S + N$, то для произвольного числа h исходную регрессию можно переписать в следующем виде:

$$C = \beta_1 + \beta_2' S + \beta_3' N + \beta_4' T + \epsilon$$

, где $\beta_2' = \beta_2 + h$, $\beta_3' = \beta_3 + h$, $\beta_4' = \beta_4 - h$. Таким образом, одни и те же наблюдения могут быть объяснены различными наборами коэффициентов β. Эта ситуация тесно связана с проблемой *идентифицируемости* системы, о чем более подробно будет говориться позднее. Кроме того, если с учетом равентва $T = S + N$ переписать исходную систему в виде

$$C = \beta_1 + (\beta_2 + \beta_4)S + (\beta_3 + \beta_4)N + \epsilon$$

, то становится ясно, что оценить можно лишь три параметра β_1, $(\beta_2 + \beta_4)$ и $(\beta_3 + \beta_4)$, а не четыре исходных. В общем случае можно показать, что если $rank(X'X) = l < k$, то оценить можно только l линейных комбинаций исходных коэффициентов. Если есть полная коллинеарность, то можно выделить в матрице X максимальную линейно независимую систему столбцов и, удалив остальные столбцы, провести новую регрессию.

На практике полная коллинеарность встречается исключительно редко. Гораздно чаще приходится сталкиваться с ситуацией, когда матрица X имеет полный ранг, но между регрессорами имеется высокая степень корреляции, т.е., говоря нестрого, *когда матрица $X'X$ близка к вырожденной.* Тогда говорят о наличии мультиколлинеарности. В этом случае МНК-оценка формально существует, но обладает "плохими свойствами".

Мультиколлинеарность может возникать в силу разных причин. Например несколько независимых переменных могут иметь общий временной тренд, относительно которого они совершают малые колебания. В частности, так может случиться, когда значения одной независимой переменной являются лагированными значениями другой.

Выделим некоторые наиболее характерные признаки мультиколлинеарности:

1. Небольшое изменение исходных данных (например, добавление новых наблюдений) приводит к существенному изменению оценок коэффициентов модели.

2. Оценки имеют большие стандартные ошибки, малую значимость, в то время как модель в целом является значимой (высокое значение коэффициента детерминации R^2 и соответствующей F-статистики).

3. Оценки коэффициентов имеют неправильные с точки зрения теории знаки или неоправданно большие значения.

Что же делать, если по всем признакам имеется мультиколлинеарность? Однозначного ответа на этот вопрос нет, и среди эконометристов есть разные мнения на этот счет. Существует даже такая школа, представители которой считают, что и не нужно ничего делать, посколько "так устроен мир" (см. Kennedy, 1992). Мы здесь не ставим цель дать достаточно полное описание методов борьбы с мультиколлинеарностью. Более подробно об этом можно прочесть, например, в (Greene, 1997, глава 9).

У неискушенного исследователя при столкновении с проблемой мультиколлинеарности может возникнуть естественное желание отбросить "лишние" независимые переменные, которые, возможно, служат ее причиной. Однако следует помнить, что при этом могут возникнуть новые трудности. Во-первых, далеко не всегда ясно, какие переменные являются лишними в указанном смысле. Мультиколлинеарность означает лишь приблизительную линейную зависимость между столбцами матрицы X, но это не всегда выделяет "лишние" переменные. Во-вторых, во многих ситуациях удаление каких-либо независимых переменных может значительно отразиться на содержательном смысле модели. Наконец, отбрасывание так называемых существенных переменных, т.е. независимых переменных, которые реально влияют на изучаемую зависимую переменную, приводит к смещенности МНК-оценок.

Что можно предпринять в случае мультиколлинеарности?

1. Фактор - σ_E^2.

   Можно попытаться уменьшить σ_ϵ^2. Случайный член отражает воздействие на переменную Y всех влияющих на нее переменных, не включенных в регрессионное уравнение. Таким образом если найти важную переменную,не включенную в модель и, следовательно, вносящую вклад в ϵ, то при ее включении дисперсия случайного члена уменьшится.

2. Фактор - число наблюдений

   Можно увеличить объем выборки

   Можно применить метод группировки. Например, страну можно поделить на области, отдельные города и зоны. Далее используя случайный отбор географических зон и случайной выборки, обеспечивают должное представительство различных регионов. При работе с временными рядами можно увеличить выборку перейдя от длинных к коротким временным интервалам: от годовых к квартальным, месячным и т.д.

3. Увеличение дисперсии объясняющих переменных

   Это можно сделать лишь на стадии планирования проводимого опроса. Например, при планировании опроса семей по доходам и расходам следует включить в опрос относительно богатые и бедные семьи наряду с семьями со средним доходом.

4. На стадии планирования опроса следует получить такую выборку, в которой оюъясняющие переменные были бы как можно меньше связаны между собой.

5. Если коррелированные переменные связаны между собой концептуально, то может быть рационально объединить их в совокупный индекс.

6. Можно исключить некоторые из коррелированных переменных если их коэффициенты незначительны. Это может быть опасно, т.к. некоторые из исключаемы исключаемых коэффициентов могли принадлежать модели и их невключение может служить причиной смещения оценки.

7. Использовать внешнюю информацию относительно коэффициента одной из переменных в уравнении (например, экспертная оценка).

1.2 Фиктивные переменные

Как правило, независимые переменные в регрессионных моделях имеют "непрерывные" области изменения (нац. доход, уровень безработицы, размер зарплаты и т.п.). Однако теория не накладывает никаких ограничений на характер регрессоров, в частности, некоторые переменные могут принимать всего два значения или, в более общей ситуации, дискретное множество значений. Необходимость рассматривать такие переменные возникает довольно часто в тех случаях, когда требуется принимать во внимание какой-либо *качественный признак*. Например, при исследовании зависимости зарплаты от различных факторов может возникнуть вопрос, влияет ли на ее размер, и если да, то в какой степени, наличие у работника высшего образования. Также можно задать вопрос, существует ли дискриминация в оплате труда между мужчинами и женщинами. В принципе можно оценивать соответствующие уравнения внутри каждой категории, а затем изучать различия между ними, но введение дискретных переменных позволяет оценивать одно уравнение сразу по всем категориям.

Покажем, как это можно сделать в примере с зарплатой. Пусть $x_t = (x_{t1}, ..., x_{tk})'$ - набор объясняющих переменных, т.е. первоначальная модель описывается уравнениями

$$y_t = x_{t1}\beta_1 + .. + x_{tk}\beta_k + \epsilon_t = x_t'\beta + \epsilon_t, t = 1, ..., n, \quad (1)$$

где y_t - размер зарплаты t-го работника. Теперь мы хотим включить в рассмотрение такой фактор, как наличие или отсутствие высшего образования. Введем новую, бинарную переменную d, полагая $d_t = 1$, если в t-м наблюдении индивидуум имеет высшее образование, и $d_t = 0$ в противном случае, и рассмотрим новую систему

$$y_t = x_{t1}\beta_1 + ... + x_{tk}\beta_k + d_t\delta + \epsilon_t = z_t'\gamma + \epsilon_t, t = 1, ..., n, \quad (2)$$

где $z = (x_1, ..., x_k, d)' = (x', d)'$, $\gamma = (\beta_1, ..., \beta_k, \delta)'$. Иными словами, принимая модель 1, мы считаем, что средняя зарплата есть $x'\beta$ при отсутствии высшего образования и $x'\beta + \delta$ - при его наличии. Таким образом, величина δ интерпретируется как среднее изменение зарплаты при переходе из одной категории (без высшего образования) в другую (с высшим образованием) при неизменных значениях остальных параметров. К системе 1 можно применить метод наименьших квадратов и получить оценки соответствующих коэффициентов. Легко понять, что, тестируя гипотезу $\delta = 0$, мы проверяем предположение о несущественном различии в зарплате между категориями.

Замечание. В англоязычной литературе по эконометрике переменные указанного типа называются *dummy variables*, что на русский язык часто переводится как "фиктивные переменные" (см., например, Джонстон, 1980). Следует, однако, ясно понимать, что d такая же "равноправная" переменная, как и любой из регрессоров $x_j, j = 1, ..., k$. Ее "фиктивность" состоит только в том, что она количественным образом описывает качественный признак.

Качественное различие можно формализовать с помощью любой переменной, принимающей два значения, а не обязательно значения 0 или 1. Однако на практике почти всегда используют фиктивные переменные типа "0-1", поскольку в этом случае интерпретация выглядит наиболее просто. Если бы в рассмотренном выше примере переменная d принимала значение, скажем, 5 для индивидуума с высшим образованием и 2 для индивидуума без высшего образования, то коэффициент при этом регрессоре равнялся бы трети среднего изменения зарплаты при получении высшего образования.

Если включаемый в рассмотрение качественный признак имеет не два, а несколько начений, то в принципе можно было бы ввести дискретную переменную, принимающую такое же количество значений. Но этого фактически никогда не делают, так как тогда трудно дать содержательную интерпретацию соответствующему коэффициенту. *В этих случаях целесообразнее использовать несколько бинарных переменных.* Типичным примером подобной ситуации является исследование сезонных колебаний. Пусть, например, y_t - объем потребления некоторого продукта в месяц t, и есть все основания считать, что потребление зависит от времени года. Для выявления влияния сезонности можно ввести три бинарные переменные d_1, d_2, d_3:

1. $d_{t1} = 1$, если месяц t является зимним, $d_{t1} = 0$ в остальных случаях;

2. $d_{t2} = 1$, если месяц t является весенним, $d_{t2} = 0$ в остальных случаях;

3. $d_{t3} = 1$, если месяц t является летним, $d_{t3} = 0$ в остальных случаях,

и оценивать уравнение

$$y_t = \beta_0 + \beta_1 d_{t1} + \beta_2 d_{t2} + \beta_3 d_{t3} + \epsilon_t. \quad (3)$$

Отметим, что мы не вводим четвертую бинарную переменную d_4, относящуюся к осени, иначе тогда для любого месяца t выполнялось бы тождество $d_{t1} + d_{t2} + d_{t3} + d_{t4} = 1$, что означало бы линейную зависимость регрессоров

в 3 и, как следствие, невозможность получения МНК-оценок. Такая ситуация, когда сумма фиктивных переменных тождественно равна константе, также включенной в регрессию, называется *"dummy trap"*. Иначе говоря, среднемесячный объем потребления есть β_0 для осенних месяцев, $\beta_0 + \beta_1$ - для зимних, $\beta_0 + \beta_2$ - для весенних и $\beta_0 + \beta_3$ - для летних. Таким образом, оценки коэффициентов $\beta_i, i = 1, 2, 3$, показывают средние сезонные отклонения в объеме потребления по отношению к осенним месяцам. Тестируя, например, стандартную гипотезу $\beta_3 = 0$, мы проверяем предположение о несущественном различии в объеме потребления между летним и осенним сезоном, гипотеза $\beta_1 = \beta_2$ эквивалентна предположению об отсутствии различия в потреблении между зимой и весной и т.д: $H_0 : \beta_3 = 0, H_0 : \beta_1 = \beta_2$.

Фиктивные переменные, несмотря на свою внешнюю простоту, являются весьма гибким инструментом при исследовании влияния качественных признаков. Рассмотрим еще один пример. В предыдущей модели мы интересовались сезонными различиями лишь для среднемесячного объема потребления. Модифицируем ее, введя новую независимую переменную i - доход, используемый на потребление. Как известно, в регрессии

$$y_t = \beta_0 + \beta_1 i_t + \epsilon_t \tag{4}$$

коэффициент β_1 носит название "склонность к потреблению". Поэтому естественно поставить задачу исследовать влияние сезона на склонность к потреблению. Для этого можно рассмотреть модель

$$y_t = \beta_0 + \beta_1 d_{t1} + \beta_2 d_{t2} + \beta_3 d_{t3} + \beta_4 d_{t1} i_t + \beta_5 d_{t2} i_t + \beta_6 d_{t3} i_t + \beta_7 i_t + \epsilon_t, \tag{5}$$

согласно которой склонность к потреблению зимой, весной, летом и осенью есть $\beta_4 + \beta_7, \beta_5 + \beta_7, \beta_6 + \beta_7$, и β_7 соответственно. Как и в предыдущей модели, можно тестировать гипотезы об отсутствии сезонных влияний на склонность к потреблению.

Фиктивные переменные позволяют строить и оценивать так называемые кусочно-линейные модели, которые можно применять для исследования структурных изменений. Как и раньше, проще всего это продемонстрировать на примере.

Пусть y - зависимая переменная и пусть для простоты есть только две независимые переменные: x и постоянный член. Предположим, что x и y представлены в виде временных рядов $\{(x_t, y_t), t = 1, ..., n\}$ (например, x_t - размер основного фонда некоторого предприятия в период t, y_t - объем продукции, выпущенной в этот же период). Из некоторых априорных соображений исследователь считает, что в момент времени t_0 произошла структурная перестройка и линия регрессии будет отличаться от той, что была до момента t_0, но общая линия остается непрерывной (рис. 1).

Чтобы оценить такую модель, введем бинарную переменную $r_t = 0$, если $t \leq t_0$ и $r_t = 1$, если $t > t_0$ и запишем следующее регрессионное уравнение:

$$y_t = \beta_1 + \beta_2 x_t + \beta_3 (x_t - x_{t_0}) r_t + \epsilon_t. \tag{6}$$

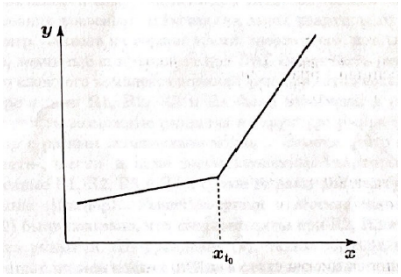


Рис. 1: Рис. 1

Нетрудно проверить, что регрессионная линия, соответствующая 6, имеет коэффициент наклона β_2 для $t \leq t_0$ и $\beta_2 + \beta_3$ для $t > t_0$, и разрыва в точке x_{t_0} не происходит. Таким образом, тестируя гипотезу $\beta_3 = 0$, мы проверяем предположение о том, что *фактически структурного изменения не произошло*, $H_0 : \beta_3 = 0$.

Выводы:

- для исследования влияния качественных признаков в модель можно вводить бинарные (фиктивные) переменные, которые, как правило, принимают значение 1, если данный качественный признак присутствует в наблюдении, и значение 0 при его отсутствии;
- способ включения фиктивных переменных зависит от априорной информации относительно влияния соответствующих качественных признаков на зависимую переменную и от гипотез, которые проверяются с помощью модели;
- от способа включения фиктивной переменной зависит и интерпретации оценки коэффициента при ней.

1.3 Частная корреляция

В том случае, когда имеются одна независимая и одна зависимая переменные, естественной мерой зависимости (в рамках линейного подхода) является (выборочный) коэффициент корреляции между ними. Использование множественной регрессии позволяет обобщить это понятие на случай, когда имеется несколько независимых переменных. Корректировка здесь необходима по следующим очевидным соображениям. Высокое значение коэффициента может означать высокую степень зависимости или то, что существует третья переменная, от которой зависят и первая и вторая. Поэтому

возникает естественная задача найти “чистую“ корреляцию между двумя переменными, исключив (линейное) влияние других факторов. Это можно сделать с помощью коэффициента частной корреляции. Для простоты предположим, что имеется регрессионная модель

$$y = \beta_0 + x_1\beta_1 + x_2\beta_2 + \epsilon,$$

где, как обычно, $y - n \times 1$ вектор наблюдений зависимой переменной, $x_1, x_2 - n \times 1$ векторы независимых переменных, $\beta_0, \beta_1, \beta_2$ – скалярные параметры, $\epsilon - n \times 1$ вектор ошибок. Наша цель - определить корреляцию между y и, например, первым регрессором x_1 после исключения влияния x_2.

Соответствующая процедура устроена следующим образом:

1. Осуществим регрессию y на x_2 и константу и получим прогнозные значения $\hat{y} = \hat{\alpha}_1 + \hat{\alpha}_2 x_2$.

2. Осуществим регрессию x_1 на x_2 и константу и получим прогнозные значения $\hat{x}_1 = \hat{\gamma}_1 + \hat{\gamma}_2 x_2$.

3. Удалим влияние x_2, взяв остатки $e_y = y - \hat{y}$ и $e_{x_1} = x_1 - \hat{x}_1$.

4. Определим (выборочный) *коэффициент частной корреляции* между y и x_1 при исключении влияния x_2 как (выборочный) коэффициент корреляции между e_y и e_{x_1}:

$$r(y, x_1|x_2) = r(e_y, e_{x_1}). \tag{7}$$

   Напомним, что из свойств метода наименьших квадратов следует, что e_y и e_{x_1} не коррелированы с x_2: $x_1^T e_{x_1} = 0$, $x_2^T e_y = 0$. Именно в этом смысле указанная процедура соответствует интуитивному представлению об “исключении (линейного) влияния переменной x_2“

Прямыми вычислениями можно показать (см. упражнение 2), что справедлива следующая формула, связывающая коэффициенты частной и обычной корреляции:

$$r(y, x_1|x_2) = \frac{r(y, x_1) - r(y, x_2)r(x_1, x_2)}{\sqrt{1 - r^2(x_1, x_2)}\sqrt{1 - r^2(y, x_2)}}. \tag{8}$$

Значения $r(y, x_1|x_2)$ лежат в интервале $[-1, 1]$, как у обычного коэффициента корреляции. Равенство коэффициента $r(y, x_1|x_2)$ нулю означает, говоря нестрого, отсутствие *прямого линейного влияния переменной x_1 на y.*

Существует тесная связь между коэффициентом частной корреляции $r(y, x_1|x_2)$ и коэффициентом детерминации R^2, а именно

$$r^2(y, x_1|x_2) = \frac{R^2 - r^2(y, x_2)}{1 - r^2(y, x_2)}, \tag{9}$$

или

$$1 - R^2 = (1 - r^2(y, x_2))(1 - r^2(y, x_1|x_2)).$$

Описанная выше процедура очевидным образом обобщается на случай, когда исключается влияние не одной, а нескольких переменных: достаточно переменную x_2 заменить на набор переменных X_2, сохраняя определение 7. Формула 8, естественно, усложнится. Подробнее об этом можно прочесть в книге (Айвазян и др., 1986).

Проиллюстрируем приведенное выше понятие частных коэффициентов корреляции и их отличие от обычных коэффициентов корреляции на следующем примере.

Пример. Рынки валютных фьючерсов. Рассмотрим вопрос о связи российского и западных рынков валютных фьючерсов.

В настоящее время несколько российских бирж ведут торговлю срочными контрактами на поставку доллара США: МТБ, МЦФБ, РТСБ и др. Однако (см. Яковлев, Бессонов, 1995а, 1995б) в течение периода наблюдений (ноябрь 1992 г. - сентябрь 1995 г.) на МТБ приходилось от 75 до 85% общего объема торговли. Поэтому в качестве цен фьючерсных контрактов на поставку доллара США мы выбрали котировки контрактов на МТБ.

Динамика цен валютных фьючерсов на Западе не сильно зависит от биржи. Для анализа мы взяли биржу с наибольшим объемом торговли - IMM (International Monetary Market, Chicago).

Мы используем ежедневные данные - цена закрытия для IMM и котировочная цена для МТБ - показатели, которые используют торговые палаты этих бирж для ежедневного перерасчета позиций инвесторов (вариационной маржи).

В качестве параметров для сравнения мы взяли не сами цены контрактов, а “доходности“, приведенные к годичному базису, определяемые как

$$y_{t,T} = (\ln F_t^T - \ln S_t)/(T - t) \cdot 365,$$

где F_t^T - цена контракта в момент времени t на поставку 1 доллара в момент времени T (т.е. со сроком до поставки $T - t$); S_t - спот-курс доллара в момент t. (Для рубля - данные ММВБ, для немецкой марки DM, британского фунта BP, японской иены JY - данные IMM.) $y_{t,T}^{RU}, y_{t,T}^{DM}, y_{t,T}^{BP}, y_{t,T}^{JY}$ обозначают доходности контрактов на поставку 1 доллара в рублях, DM, BP, JY. На наш взгляд, этот показатель в меньшей мере зависит от темпа инфляции, чем сама цена контракта. Время t изменяется в днях.

Рассмотрим *таблицу коэффициентов корреляции доходностей* $y_{t,T}^{RU}, y_{t,T}^{DM}$, $y_{t,T}^{BP}, y_{t,T}^{JY}$:

	RU	DM	BP	JY
RU	1			
DM	0.626	1		
BP	0.380	0.775	1	
JY	0.615	0.919	0.602	1

Из таблицы видны высокие (0.602, 0.775, 0.919) значения коэффициентов корреляции показателей для западных валют, что неудивительно ввиду высокой степени интегрированности западных финансовых рынков. Удивле-

ние вызывают высокие 0.615 (0.626) значения коэффициентов корреляции показателей для рубля и японской иены (немецкой марки).

Рассмотрим теперь таблицу коэффициентов частной корреляции между доходностями $y_{t,T}^{XX}$ для XX = RU,DM,BP,JY (устранено влияние временного тренда t).

	RU	DM	BP	JY
RU	1			
DM	0.024	1		
BP	0.008	0.807	1	
JY	-0.003	0.488	0.276	1

Теперь мы видим картину более реалистичную! Наиболее тесно связаны между собой европейские валюты (BP,DM), слабее связь европейских валют и японской иены и практически отсутствует связь российской валюты с западными.

Таким образом, высокие коэффициенты корреляции в первой таблице, например 0.626 для RU-DM, были лишь следствием того, что на интервале наблюдений (ноябрь 1992 г. - сентябрь 1995 г.) отмечалось падение курса рубля по отношению к доллару и падение курса доллара по отношению к немецкой марке, т.е. эта корреляция является следствием наличия временного тренда в $y_{t,T}^{RU}$ и $y_{t,T}^{DM}$. Наш вывод подтверждается также тем, что коэффициенты корреляции $y_{t,T}^{RU}$ и $y_{t,T}^{DM}$ с t достаточно высоки (-0.673, 0.920).

1.4 Процедура пошагового отбора переменных

Коэффициент частной корреляции часто используется при решении проблемы *спецификации модели*.

Иногда исследователь заранее знает характер зависимости исследуемых величин, опираясь, например, на экономическую теорию, предыдущие результаты, априорные знания и т.п., и задача состоит лишь в оценивании неизвестных параметров. (По существу, во всех наших предыдущих рассуждениях мы неявно предполагали, что имеется именно такая ситуация.) Классический пример - оценивание параметров производственной функции Кобба-Дугласа $Y = AK^{\alpha}L^{\beta}$, где Y - совокупный выпуск, K - капиталовложения и L - трудозатраты. Логарифмируя это выражение получаем линейное относительно $\ln A, \alpha, \beta$ уравнение, из которого, например, с помощью метода наименьших квадратов можно получить оценки этих параметров, проверять те или иные гипотезы и т.д.

Однако на практике довольно часто приходится сталкиваться с ситуацией, когда имеется большое число наблюдений различных параметров (независимых переменных), но нет априорной модели изучаемого явления. Возникает естественная проблема, какие переменные включить в регрессионную схему. Теоретические вопросы, связанные с этой проблемой, будут изложены далее.

В компьютерные пакеты включены различные эвристические процедуры *пошагового отбора регрессоров*. Основными пошаговыми процедурами

являются *1) процедура последовательного присоединения, 2) процедура присоединения-удаления и 3) процедура последовательного удаления.* Опишем кратко одну из таких процедур, использующую понятие коэффициента частной корреляции.

Процедура присоединения-удаления

На первом шаге из исходного набора объясняющих переменных выбирается (включается в число регрессоров) переменная, имеющая наибольший по модулю коэффициент корреляции с зависимой переменной y.

$$x_i : |r(y, x_i)| = \max_j |r(y, x_j)|.$$

Второй шаг состоит из двух подшагов. На первом из них, который выполняется, если число регрессоров уже больше двух, делается попытка исключить один из регрессоров. Ищется тот регрессор x_s, удаление которого приводит к наименьшему уменьшению коэффициента детерминации. Затем сравнивается значение F-статистики для проверки гипотезы $H_0 : \beta_s = 0$ о незначимости этого регрессора с некоторым заранее заданным пороговым значением $F_{\text{искл}}$. Если $F < F_{\text{искл}}$, то x_s удаляется из списка регрессоров. Второй подшаг состоит в попытке включения нового регрессора из исходного набора предсказывающих переменных. Ищем переменную x_p с наибольшим по модулю частным коэффициентом корреляции (исключается влияние ранее включенных в уравнение регрессоров) и сравниваем значение F-статистики для проверки гипотезы $H_0 : \beta_p = 0$ о незначимости этого регрессора с некоторым заранее заданным пороговым значением $F_{\text{вкл}}$.

$$x_p : |r(y, x_p|x_i)| = max_j r(y, x_j|x_i)|.$$

Если $F > F_{\text{вкл}}$, то x_p включается в список регрессоров. Обычно выбирают $F_{\text{искл}} < F_{\text{вкл}}$. Второй шаг повторяется до тех пор, пока происходит изменение списка регрессоров. Конечно, ни одна из пошаговых процедур не гарантирует получение оптимального по какому-либо критерию набора регрессоров.

Подробное описание пошаговых процедур содержится в книге (Айвазян и др., 1985).

Билет 5. Различные аспекты множественной регрессии. Тест Чоу. Мультиколлинеарность. Фиктивные переменные. Коэффициент частной корреляции. Процедура пошагового отбора переменных. Спецификация модели. Короткая и длинная регрессии. Замещающие переменные.

Модель множественной регрессии

Модель n-мерной регрессии:

$$y_t = \beta_1 + \beta_2 x_{t2} + ... + \beta_k x_{tk} + \varepsilon_t, \quad t = 1, ..., n$$

или

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + ... + \beta_k x_{tk} + \varepsilon_t, \quad t = 1, ..., n$$

где x_{tp} — значения регрессора x_p в наблюдении t, а $x_{t1} = 1, t = 1, ..., n$.

Статистики

t-статистику можно использовать для проверки гипотезы H_0: $b = b_0$ против альтернативной гипотезы H_1: $b \neq b_0$. В множественной регрессии как b берем β_i.

Предположим, что верна гипотеза H_0, и зададимся, например, α-квантилью t-распределения с $(n-2)$ степенями свободы t_c , т.е H_0 принимаем, если $|t| \leq t_c$ на заданном уровне значимости $\frac{\alpha}{2}$. Мы отвергаем гипотезу H0 (и принимаем H1) на уровне значимости $\frac{\alpha}{2}$, если $|t| > t_c$ («редкое» событие с точки зрения гипотезы H0).

$$t = \frac{b - b_0}{s_b}, \quad \text{где} \quad s_b = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}} \frac{\sqrt{\sum x_i^2}}{n\sqrt{DX}}$$

$t_c(n-2; \frac{\alpha}{2})$ считаем в Excel или находим здесь: таблица Стьюдента.

Малые значения t-статистики соответствуют отсутствию достоверной статистической связи объясняющей переменной X и зависимой переменной Y.

f-статистику используют для проверки гипотезы H_0 о том, что коэффициенты регрессии одновременно равны нулю, $\beta = (\beta_1, ..., \beta_k) = 0$. Другими словами, суть расчетов - ответить на вопрос: можно ли анализируемое уравнение регрессии использовать для дальнейшего анализа и прогнозов?

$$F = \frac{R^2}{1-R^2} \frac{n-k}{k-1} = \frac{\text{RSS}}{\text{ESS}} \frac{n-k}{k-1} \sim F(k-1, n-k)$$

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \overline{y})^2} = 1 - \frac{\text{ESS}}{\text{TSS}} = \frac{\sum (\hat{y}_i - \overline{y})^2}{\sum (y_i - \overline{y})^2} = \frac{\text{RSS}}{\text{TSS}}.$$

Гипотеза H_0 отвергается на 5%-уровне значимости, если $F > F_c$, где F_c — односторонняя 95%-квантиль распределения Фишера $F(k-1, n-k)$. Считаем ее через Excel: FРАСПОБР(0,95;k-1;n-k).

5.1 Тест Чоу (Chow)

Предположим, у нас есть две выборки данных. По каждой выборке мы строим регрессионную модель. Вопрос, который нас интересует: верно ли, что эти две модели совпадают? Рассмотрим модели:

$$y_t = \beta'_1 x_{t1} + \beta'_2 x_{t2} + ... + \beta'_k x_{tk} + \varepsilon'_t, \quad t = 1, ..., n, \tag{5.1}$$

$$y_t = \beta''_1 x_{t1} + \beta''_2 x_{t2} + ... + \beta''_k x_{tk} + \varepsilon''_t, \quad t = n+1, ..., n+m, \tag{5.2}$$

в первой выборке n наблюдений и m наблюдений во второй. Например, y — заработная плата, x_t — регрессоры (возраст, стаж, уровень образования и т.п.), и пусть первая выборка относится к женщинам, вторая — к мужчинам. Вопрос: следует ли из оценки моделей (5.1), (5.2), что модель зависимости зарплаты от регрессоров одна и та же для мужчин и женщин?

Сведем эту ситуацию к общей схеме проверки линейных ограничений на параметры модели. Регрессией без ограничений здесь является объединение двух регрессий (5.1), (5.2), т.е. $\text{ESS}_{\text{UR}} = \text{ESS}_1 + \text{ESS}_2$, число степеней свободы при этом равно $(n-k) + (m-k) = n + m - 2k$. Предположим теперь, что верна нулевая гипотеза. Тогда регрессия с ограничениями записывается одним уравнением

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + ... + \beta_k x_{tk} + \varepsilon_t, \quad t = 1, ..., n+m. \tag{5.3}$$

Оценивая (5.3), получаем ESS_{R}. Тогда, учитывая, что наложено k ограничений на параметры модели, получаем

$$F = \frac{(\text{ESS}_{\text{R}} - \text{ESS}_{\text{UR}})/k}{\text{ESS}_{\text{UR}}/(n+m-2k)} \sim F(k, n+m-2k). \tag{5.4}$$

Если F-статистика (5.4) больше критического значения F_c, то нулевая гипотеза отвергается. В этом случае мы не можем объединять две выборки в одну.

В противном случае $(F < F_c(k, n+m-2k)) \Rightarrow$ принимается H_0 и можно использовать объединенную регрессию (5.3).

5.2 Мультиколлинеарность

На практике исследователю нередко приходится сталкиваться с ситуацией, когда полученная им регрессия является "плохой", т.е. t-статистики большинства оценок малы, что свидетельствует о незначимости соответствующих независимых переменных (регрессоров). В то же время F-статистика может быть достаточно большой, что говорит о значимости регрессии в целом. Одна из возможных причин такого явления носит название *мультиколлинеарности* и возникает при наличии высокой корреляции между регрессорами.

Одним из условий классической регрессионной модели является предположеие о линейной независимости объясняющих переменных, что означает *линейную независимость столбцов матрицы регрессоров* X или (эквивалентно) что *матрица* $(X'X)^{-1}$ *имеет полный ранг* k. При нарушении этого условия, т.е. когда один из столбцов матрицы X есть линейная комбинация остальных столбцов, говорят, что имеет место *полная коллинеарность*. В этой ситуации нельзя построить МНК-оценку параметра β, что формально следует из сингулярности матрицы $X'X$ и невозможности решить нормальные уравнения.

Выделим некоторые **наиболее характерные признаки мультиколлинеарности**:

1. Небольшое изменение исходных данных (например, добавление новых наблюдений) приводит к существенному изменению оценок коэффициентов модели.

2. Оценки имеют большие стандартные ошибки, малую значимость, в то время как модель в целом является значимой (высокое значение коэффициента детерминации R^2 и соответствующей F-статистики).

3. Оценки коэффициентов имеют неправильные с точки зрения теории знаки или неоправданно большие значения.

Что можно предпринять в случае мультиколлинеарности?

1. Можно попытаться уменьшить σ_ϵ^2

2. Фактор - число наблюдений

   Можно увеличить объем выборки

   Можно применить метод группировки. Например, страну можно поделить на области, отдельные города и зоны.

3. Увеличение дисперсии объясняющих переменных.

4. На стадии планирования опроса следует получить такую выборку, в которой объясняющие переменные были бы как можно меньше связаны между собой.

5. Если коррелированные переменные связаны между собой концептуально, то может быть рационально объединить их в совокупный индекс.

6. Можно исключить некоторые из коррелированных переменных если их коэффициенты незначительны. Это может быть опасно, т.к. некоторые из исключаемы исключаемых коэффициентов могли принадлежать модели и их невключение может служить причиной смещения оценки.

7. Использовать внешнюю информацию относительно коэффициента одной из переменных в уравнении (например, экспертная оценка).

5.3 Фиктивные переменные

Как правило, независимые переменные в регрессионных моделях имеют "непрерывные" области изменения (нац. доход, уровень безработицы, размер зарплаты и т.п.). Однако теория не накладывает никаких ограничений на характер регрессоров, в частности, некоторые переменные могут принимать всего два значения или, в более общей ситуации, дискретное множество значений. Необходимость рассматривать такие переменные возникает довольно часто в тех случаях, когда требуется принимать во внимание какой-либо *качественный признак*.

Следует, ясно понимать, что речь идет о такой же "равноправной" переменной, как и любой из регрессоров $x_j, j = 1, ..., k$. Ее "фиктивность" состоит только в том, что она количественным образом описывает качественный признак.

Качественное различие можно формализовать с помощью любой переменной, принимающей два значения, а не обязательно значения 0 или 1. Однако на практике почти всегда используют фиктивные переменные типа "0-1", поскольку в этом случае интерпретация выглядит наиболее просто.

5.4 Частная корреляция

В том случае, когда имеются одна независимая и одна зависимая переменные, естественной мерой зависимости (в рамках линейного подхода) является (выборочный) коэффициент корреляции между ними. Использование множественной регрессии позволяет обобщить это понятие на случай, когда имеется несколько независимых переменных. Корректировка здесь необходима по следующим очевидным соображениям. Высокое значение коэффициента может означать высокую степень зависимости или то, что существует третья переменная, от которой зависят и первая и вторая. Поэтому возникает естественная задача найти "чистую" корреляцию между двумя переменными, исключив (линейное) влияние других факторов. Это можно сделать с помощью коэффициента частной корреляции.

Для простоты предположим, что имеется регрессионная модель

$$y = \beta_0 + x_1\beta_1 + x_2\beta_2 + \epsilon,$$

где, как обычно, $y - n \times 1$ вектор наблюдений зависимой переменной, $x_1, x_2 - n \times 1$ векторы независимых переменных, $\beta_0, \beta_1, \beta_2$ – скалярные параметры, $\epsilon - n \times 1$ вектор ошибок. Наша цель - определить корреляцию между y и, например, первым регрессором x_1 после исключения влияния x_2.

Соответствующая процедура устроена следующим образом:

1. Осуществим регрессию y на x_2 и константу и получим прогнозные значения $\hat{y} = \hat{\alpha}_1 + \hat{\alpha}_2 x_2$.

2. Осуществим регрессию x_1 на x_2 и константу и получим прогнозные значения $\hat{x}_1 = \hat{\gamma}_1 + \hat{\gamma}_2 x_2$.

3. Удалим влияние x_2, взяв остатки $e_y = y - \hat{y}$ и $e_{x_1} = x_1 - \hat{x}_1$.

4. Определим (выборочный) *коэффициент частной корреляции* между y и x_1 при исключении влияния x_2 как (выборочный) коэффициент корреляции между e_y и e_{x_1}:

$$r(y, x_1|x_2) = r(e_y, e_{x_1}). \tag{5.5}$$

   Напомним, что из свойств метода наименьших квадратов следует, что e_y и e_{x_1} не коррелированы с x_2: $x_1^T e_{x_1} = 0, x_2^T e_y = 0$. Именно в этом смысле указанная процедура соответствует интуитивному представлению об "исключении (линейного) влияния переменной x_2"

Прямыми вычислениями можно показать, что справедлива следующая формула, связывающая коэффициенты частной и обычной корреляции:

$$r(y, x_1|x_2) = \frac{r(y, x_1) - r(y, x_2)r(x_1, x_2)}{\sqrt{1 - r^2(x_1, x_2)}\sqrt{1 - r^2(y, x_2)}}. \tag{5.6}$$

Значения $r(y, x_1|x_2)$ лежат в интервале $[-1, 1]$, как у обычного коэффициента корреляции. Равенство коэффициента $r(y, x_1|x_2)$ нулю означает, говоря нестрого, отсутствие *прямого линейного влияния переменной x_1 на y*.

Существует тесная связь между коэффициентом частной корреляции $r(y, x_1|x_2)$ и коэффициентом детерминации R^2, а именно

$$r^2(y, x_1|x_2) = \frac{R^2 - r^2(y, x_2)}{1 - r^2(y, x_2)}, \tag{5.7}$$

или

$$1 - R^2 = (1 - r^2(y, x_2))(1 - r^2(y, x_1|x_2)).$$

5.5 Процедура пошагового отбора переменных

Коэффициент частной корреляции часто используется при решении проблемы *спецификации модели*.

Иногда исследователь заранее знает характер зависимости исследуемых величин, опираясь, например, на экономическую теорию, предыдущие результаты, априорные знания и т.п., и задача состоит лишь в оценивании неизвестных параметров. Однако на практике довольно часто приходится сталкиваться с ситуацией, когда имеется большое число наблюдений различных параметров (независимых переменных), но нет априорной модели изучаемого явления. Возникает естественная проблема, какие переменные включить в регрессионную схему.

В компьютерные пакеты включены различные эвристические процедуры *пошагового отбора регрессоров*. Основными пошаговыми процедурами являются *1) процедура последовательного присоединения, 2) процедура присоединения-удаления и 3) процедура последовательного удаления*. Опишем кратко одну из таких процедур, использующую понятие коэффициента частной корреляции.

Процедура присоединения-удаления

На первом шаге из исходного набора объясняющих переменных выбирается (включается в число регрессоров) переменная, имеющая наибольший по модулю коэффициент корреляции с зависимой переменной y.

$$x_i : |r(y, x_i)| = \max_j |r(y, x_j)|.$$

Второй шаг состоит из двух подшагов. На первом из них, который выполняется, если число регрессоров уже больше двух, делается попытка исключить один из регрессоров. Ищется тот регрессор x_s, удаление которого приводит к наименьшему уменьшению коэффициента детерминации. Затем сравнивается значение F-статистики для проверки гипотезы $H_0 : \beta_s = 0$ о незначимости этого регрессора с некоторым заранее заданным пороговым значением $F_{\text{искл}}$. Если $F < F_{\text{искл}}$, то x_s удаляется из списка регрессоров. Второй подшаг состоит в попытке включения нового регрессора из исходного набора предсказывающих переменных. Ищем переменную x_p с наибольшим по модулю частным коэффициентом корреляции (исключается влияние ранее включенных в уравнение регрессоров) и сравниваем значение F-статистики для проверки гипотезы $H_0 : \beta_p = 0$ о незначимости этого регрессора с некоторым заранее заданным пороговым значением $F_{\text{вкл}}$.

$$x_p : |r(y, x_p|x_i)| = max_j r(y, x_j|x_i).$$

Если $F > F_{\text{вкл}}$, то x_p включается в список регрессоров. Обычно выбирают $F_{\text{искл}} < F_{\text{вкл}}$. Второй шаг повторяется до тех пор, пока происходит изменение списка регрессоров. Конечно, ни одна из пошаговых процедур не гарантирует получение оптимального по какому-либо критерию набора регрессоров.

5.6 Спецификация модели

		Истинная модель	
		$y = Xb + \varepsilon$	$y = Xb + Z\gamma + \varepsilon$
Оценочна модель	$y = Xb + \varepsilon$	Правильная спецификация. Все оценки и их статистические свойства соответствуют нашим выводам.	Коэффициенты регрессии смещены. Стандартные ошибки рассчитаны некорректно. I
	$y = Xb + Z\gamma + \varepsilon$	Коэффициенты регрессии не смещены. Стандартные ошибки, в целом, рассчитаны корректно. II	Правильная спецификация. Все в порядке.

Случай I. Из модели исключены существенные переменные

- Процесс, порождающий данные("истинная модель"):
  $(*)$ $y = Xb + Z\gamma + \varepsilon$ - "длинная" регрессия
- Модель, описывающая данные:
  $(**)$ $y = Xb + \varepsilon$ - "короткая" регрессия

$y - (n \times 1)$ - вектор наблюдений
$X - (n \times k)$ - матрица
$Z - (n \times l)$ - матрица
$b - (k \times 1)$ - вектор коэффициентов
$\gamma - (l \times 1)$ - вектор коэффициентов

МНК-оценка вектора параметров b имеет вид для модели $(**)$:

$$\hat{b} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y$$

.

Так как истинная модель имеет вид $(*)$, то
$E\hat{b} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Ey = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}(Xb + Z\gamma) = b + (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Z\gamma.$
$V(\hat{b}) = V((X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y) = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}V(y)((X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}})^{\mathrm{T}} = \sigma^2(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}X(X^{\mathrm{T}}X)^{-1} = \sigma^2(X^{\mathrm{T}}X)^{-1}.$
$\Rightarrow$ оценка $\hat{b}$ смещена, кроме двух случаев:
1) $\gamma = 0$.
2) $X^{\mathrm{T}}Z = 0$. (ортогональность регрессоров X и Z).

Выводы:
1. Если включить как можно больше переменных в регрессию уравнения, то при этом уменьшится точность оценок, а качество подготовки возрастет.
2. Увеличение числа регрессоров приводит к сильной корреляции между регрессорами и, следовательно, к неустойчивости модели (мультиколлинеарность).

5.7 Замещающие переменные

Часто бывает, что нельзя найти данные по переменной, которую хотелось бы включить в модель, например:
1. Из-за расплывчатого определения переменную нельзя измерить,
2. Измерить можно, но это требует большого количества времени и средств.

В этом случае вместо переменной можно использовать ее заменитель. Это лучше, чем пренебрегать переменной. Например:
1. Качество образования - отношение числа студентов к числу преподавателей,
2. Социально-экономическое положение человека - уровень его дохода.

Причины, по которым следует включить в регрессионное уравнение замещающие переменные:
1. Если опустить важную переменную, то можно получить смещенные оценки,
2. Оценка регрессии можно дать косвенную информацию о замещаемой переменной.

5 лекция

Содержание

1 Общие понятия

Под временным рядом понимается последовательность наблюдений значений некоторой переменной, произведенных через равные промежутки времени. Если принять длину такого промежутка за единицу времени (год, квартал, день и т.п), то можно считать, что последовательные наблюдения $x_1, ..., x_n$ произведены в моменты $t = 1, ..., n$.

Основная отличительная особенность статистического анализа временных рядов состоит в том, что последовательность наблюдений $x_1, ..., x_n$ рассматривается как реализация последовательности, вообще говоря, *статистически зависимых* случайных величин $X_1, ..., X_n$, имеющих некоторое совместное распределение с функцией распределения

$$F(v_1, v_2, ..., v_n) = P\{X_1 < v_1, X_2 < v_2, ..., X_n < v_n\}$$

Мы будем рассматривать в основном временные ряды, у которых совместное распределение случайных величин $X_1, ..., X_n$ имеет совместную плотность распределения $p(x_1, x_2, ..., x_n)$.

Чтобы сделать задачу статистического анализа временных рядов доступной для практического решения, приходится так или иначе ограничивать класс рассматриваемых моделей временных рядов, вводя те или иные предположения относительно структуры ряда и структуры его вероятностных характеристик. Одно из таких ограничении предполагает *стационарность* временного ряда.

Ряд x_t, $t = 1, ...n$ называется **строго стационарным** (или **стационарным в узком смысле**), если для любого m ($m < n$) совместное распределение вероятностей случайных величин $X_{t_1}, ..., X_{t_m}$, такое же как и для $X_{t_1+\tau}, ..., X_{t_m+\tau}$, при любых $t_1, ..., t_m$ и τ, таких, что $1 \le t_1, ..., t_m \le n$ и $1 \le t_1 + \tau, ..., t_m + \tau \le n$, т.е функция плотности при сдвиге во времени не изменится

$$P\{x_{t_1}, ..., x_{t_m}\} = P\{x_{t_1+\tau}, ..., x_{t_m+\tau}\}$$

В частности, при $m = 1$ из предположения о строгой стационарности временного ряда x_t следует, что закон распределения вероятностей

случайной величины X_t не зависит от t, а значит, не зависят от t и все его основные числовые характеристики (если, конечно, они существуют), в том числе: математическое ожидание $E(X_t) = \mu$ и дисперсия $D(X_t) = \sigma^2$.

Значение μ определяет постоянный уровень, относительно которого колеблется анализируемый временной ряд x_t, а постоянная σ характеризует размах этих колебаний.

Как мы уже говорили, одно из главных отличий последовательности наблюдений, образующих временной ряд, заключается в том, что члены временного ряда являются, вообще говоря, статистически взаимозависимыми. Степень тесноты статистической связи между случайными величинами X_t и $X_{t+\tau}$ может быть измерена парным коэффициентом корреляции

$$Corr(X_t, X_{t+\tau}) = \frac{Cov(X_t, X_{t+\tau})}{\sqrt{D(X_t)}\sqrt{D(X_{t+\tau})}}$$

где

$$Cov(X_t, X_{t+\tau}) = E[(X_t - E(X_t)(X_{t+\tau} - E(X_{t+\tau}))].$$

Если ряд стационарный, то значение $Cov(X_t, X_{t+\tau})$ не зависит от t и является функцией только от τ; мы будем использовать для него обозначение $\gamma(\tau)$:

$$\gamma(\tau) = Cov(X_t, X_{t+\tau}).$$

В частности,

$$D(X_t) = Cov(X_t, X_t) \equiv \gamma(0).$$

Соответственно, для стационарного ряда и значение коэффициента корреляции $Corr(X_t, X_{t+\tau})$ зависит только от τ; мы будем использовать для него обозначение $\rho(\tau)$, так что

$$\rho(\tau) = Corr(X_t, X_{t+\tau}) = \gamma(\tau)/\gamma(0).$$

В частности, $\rho(0) = 1$.

Практическая проверка строгой стационарности ряда x_t на основании наблюдения значений $x_1, x_2, ..., x_n$ в общем случае затруднительна.

В связи с этим под стационарным рядом на практике часто подразумевают временной ряд x_t, у которого

- $E(X_t) \equiv \mu$
- $D(X_t) \equiv \sigma^2$
- $Cov(X_t, X_{t+\tau} = \gamma(\tau)$ для любых t и τ.

Ряд, для которого выполнены указанные три условия, называют **стационарным в широком смысле (слабо стационарным, стационарным второго порядка** или **ковариационно стационарным)**.

Если ряд является стационарным в широком смысле, то он не обязательно является строго стационарным. В то же время, и строго стационарный ряд может не быть стационарным в широком смысле просто потому, что у него могут не существовать математическое ожидание и/или дисперсия. (В отношении последнего примером может служить случайная выборка из распределения Коши.) Кроме того, возможны ситуации, когда указанные три условия выполняются, но, например $E(X_t^3)$ зависит от t.

Ряд x_t $t = 1, ..., n$ называется **гауссовским**, если совместное распределение случайных величин $X_1, ..., X_n$ является n-мерным нормальным распределением. $X_t \sim N(\mu, \sigma^2)$. Для гауссовского ряда понятия стационарности в узком и широком смысле совпадают.

В дальнейшем, говоря о стационарности некоторого ряда x_t, мы (если не оговаривается противное) будем иметь в виду, что этот ряд стационарен в широком смысле (так что у него существуют математическое ожидание и дисперсия).

Итак, пусть x_t - стационарный ряд с $E(X_t) \equiv \mu, D(X_t) \equiv \sigma^2$ и $\rho(\tau) = Corr(X_t, X_{t+\tau})$. Поскольку в данном случае коэффициент $\rho(\tau)$ измеряет корреляцию между членами одного и того же временного ряда, его принято называть **коэффициентом автокорреляции** (или

просто **автокорреляцией**). По той же причине о ковариациях $\gamma(\tau) = Cov(X_t, X_{t+\tau})$ говорят как об **автоковариациях**. При анализе изменения величины $\rho(\tau)$ в зависимости от значения τ принято говорить об **автокорреляционной функции** $\rho(\tau)$. Автокорреляционная функция безразмерна, т.е. не зависит от масштаба измерения анализируемого временного ряда. Ее значения могут изменяться в пределах от -1 до +1; при этом $\rho(0) = 1$. Кроме того, из стационарности ряда x_t следует, что $\rho(-\tau) = \rho(\tau)$, так что при анализе поведения автокорреляционных функций обычно ограничиваются рассмотрением только неотрицательных значений τ.

График зависимости $\rho(\tau)$ от τ часто называют **коррелограммой**. Он может использоваться для характеризации некоторых свойств механизма, порождающего временной ряд. Для дальнейшего заметим, что если x_t - стационарный временной ряд и c - некоторая постоянная, то временные ряды x_t и $(x_t + c)$ имеют одинаковые коррелограммы.

Если предположить, что временной ряд описывается моделью стационарного гауссовского процесса, то полное описание совместного распределения случайных величин $X_1, ..., X_n$ требует задания $n + 1$ параметров: $\mu, \gamma(0), \gamma(1), ..., \gamma(n-1)$ или $(\mu, \gamma(0), \rho(1), ..., \rho(n-1))$. Это намного меньше, чем без требования стационарности, но все же больше чем, количество наблюдений. В связи с этим, даже для стационарных гауссовских временных рядов приходится производить дальнейшее упрощение модели с тем, чтобы ограничить количество параметров, подлежающих оцениванию по имеющимся наблюдениям. Мы переходим теперь к рассмотрению некоторых простых по структуре временных рядов, которые, в то же время, полезны для описания эволюции во времени многих реальных экономических показателей.

2 Процесс белого шума

Процессом беого шума ("белым шумом "чисто случайным временным рядом") называют стационарный временной ряд x_t, для

которого

$$E(X_t) \equiv 0, D(X_t) \equiv \sigma^2 > 0$$

и

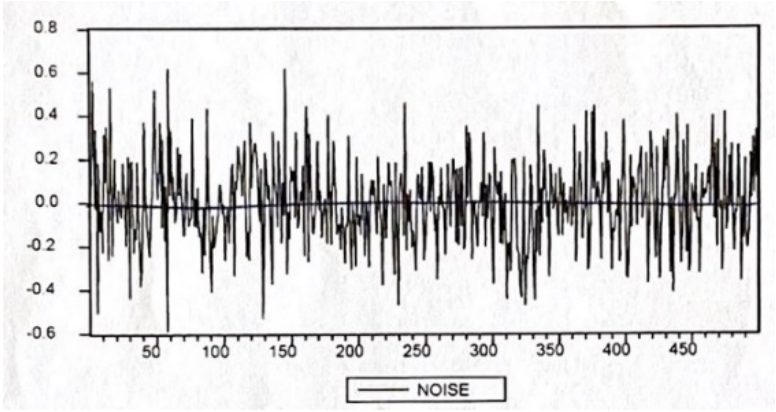
$$\rho(\tau) = 0, \text{ при } \tau \neq 0.$$

Последнее означает, что при $t \neq s$ случайные величины X_t и X_s, соответствующие наблюдениям процесса белого шума в моменты t и s, некоррелированы.

В случае, когда X_t имеет нормальное распределение, случайное величины $X_1, ...X_n$ взаимно независимы и имеют одинаковое нормальное распределение $N(0, \sigma^2)$, образуя случайную выборку из этого распределения, т.е. $X_t \sim i.i.d.N(0, \sigma^2)$. Такой ряд называют **гауссовским белым шумом**.

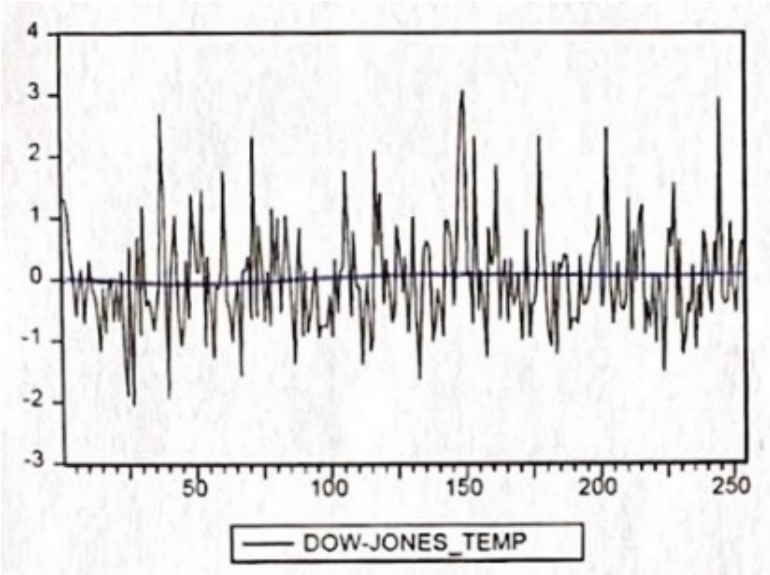
В то же время, в общем случае, даже если некоторые случайные величины $X_1, ..., X_n$ взаимно независимы и имеют одинаковое распределение, то это еще не означает, что они образуют процесс белого шума, т.к. случайная величина X_t может просто не иметь математического ожидания и/или дисперсии (в качестве примера мы опять можем указать на распределение Коши).

Временной ряд, соответствующий процессу белого шума, ведет себя крайне нерегулярным образом из-за некоррелированности при $t \neq s$ случайны величин X_t и X_s. Это иллюстрирует приводимый ниже график смоделированной реализации гауссовского процесса белого шума (NOISE) с $D(X_t) \equiv 0.04$.



В связи с этим процесс белого шума не годится для непосредственного моделирования эволюции большинства временных рядов, встречающихся в экономике. В то же время, как мы увидим ниже, такой процесс является базой для построения более реалистичных моделей временных рядов, порождающих "более гладкие"траектории ряда. В связи с частым использованием процесса белого шума в дальнейшем изложении, мы будем отличать этот процесс от других моделей временных рядов, используя для него обозначение ε_t.

В качестве примера ряда, траектория которого похожа на реализацию процесса белого шума, можно указать, например, на ряд, образованный значениями темпов изменения (прироста) индекса Доу - Джонса в течение 1984 года (дневные данные). График этого ряда имеет вид



Заметим, однако, что здесь наблюдается некоторая асимметрия распределения вероятностей значений x_t (скошенность этого распределения в сторону положительных значений), что исключает описание модели этого ряда как гауссовского белого шума.

3 Процесс авторегрессии

Одной из широко используемых моделей временных рядов является **процесс авторегрессии (модель авторегрессии)**. В своей простейшей форме модель авторегрессии описывает механизм порождения ряда следующим образом:

$$X_t = aX_{t-1} + \varepsilon_t, t = 1, ..., n,$$

где ε_t - процесс белого шума, для которого $E(\varepsilon_t) = 0$ и $D(\varepsilon_t) = \sigma^2 > 0$, X_0 - некоторая случайная величина, а $a \neq 0$ - некоторый постоянный

коэффициент.

При этом

$$E(X_t) = aE(X_{t-1}),$$

так что рассматриваемый процесс может быть стационарным только если $E(X_t) = 0$ для всех $t = 0, 1, ..., n$.

Далее,

$$X_t = aX_{t-1} + \varepsilon_t = a(aX_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = a^2X_{t-2} + a\varepsilon_{t-1} + \varepsilon_t = ...$$
$$= a^tX_0 + a^{t-1}\varepsilon_1 + a^{t-2}\varepsilon_2 + ... + \varepsilon_t,$$
$$X_{t-1} = aX_{t-2} + \varepsilon_{t-1} = a^{t-1}X_0 + a^{t-2}\varepsilon_1 + a^{t-3}\varepsilon_2 + ... + \varepsilon_{t-1},$$
$$X_{t-2} = aX_{t-3} + \varepsilon_{t-2} = a^{t-2}X_0 + a^{t-3}\varepsilon_1 + a^{t-4}\varepsilon_2 + ... + \varepsilon_{t-2}$$
$$...$$
$$X_1 = aX_0 + \varepsilon_1.$$

Если случайная величина X_0 не коррелирована со случайными величинами $\varepsilon_1, \varepsilon_2, ..., \varepsilon_n$, то отсюда следует, что

$$Cov(X_0, \varepsilon_1) = 0, Cov(X_1, \varepsilon_2) = 0, ..., Cov(X_{t-1}, \varepsilon_t) = 0$$

и

$$D(X_t) = D(aX_{t-1} + \varepsilon_t) = a^2D(X_{t-1}) + D(\varepsilon_t), t = 1, ..., n.$$

Предполагая, наконец, что
$D(X_0) = D(X_t) = \sigma_X^2$ для всех $t = 1, ..., n$,

находим:

$$\sigma_X^2 = a^2\sigma_X^2 + \sigma_\varepsilon^2.$$

Последнее может выполняться только при выполнии условия $a^2 < 1$, т,е. $a < 1$.

При этом получаем выражение для σ_X^2

$$\sigma_X^2 = \frac{\sigma_\varepsilon^2}{1 - a^2}$$

Что касается автокорреляций и автоковариаций, то

$$Corr(X_t, X_{t+\tau}) = a^\tau$$

и

$$\begin{aligned} Cov(X_t, X_{t+\tau} = Cov(a^t X_0 + a^{t-1}\varepsilon_1 + a^{t-2}\varepsilon_2 + ... + \varepsilon_t, \\ a^{t+\tau} X_0 + a^{t+\tau-1}\varepsilon_1 + a^{t+\tau-2}\varepsilon_2 + ... + \varepsilon_{t+\tau}) = \\ = a^{2t+\tau} D(X_0) + a^\tau (1 + a^2 + ... + a^{2(t-1)})\sigma_\varepsilon^2 = \\ = a^\tau \left[\frac{a^{2t}\sigma_\varepsilon^2}{(1-a^2)} + \frac{(1-a^{2t})\sigma_\varepsilon^2}{(1-a^2)}\right] = \left[\frac{a^\tau}{(1-a^2)}\right]\sigma_\varepsilon^2. \end{aligned}$$

т.е при сделанных предположениях автоковариации и автокорреляции зависят только от того насколько разнесены по времени соответствующие наблюдения.

Таким образом, механизм порождения последовательных наблюдений, заданный соотношениями
$X_t = aX_{t-1} + \varepsilon_t, t = 1, ..., n,$
порождает стационарный временной ряд, если

- $a < 1$;
- случайная величина X_0 не коррелирована со случайными величинами $\varepsilon_1, \varepsilon_2, ..., \varepsilon_n$;
- $E(X_0) = 0$;
- $D(X_0) = \sigma_\varepsilon^2/(1-a^2)$.

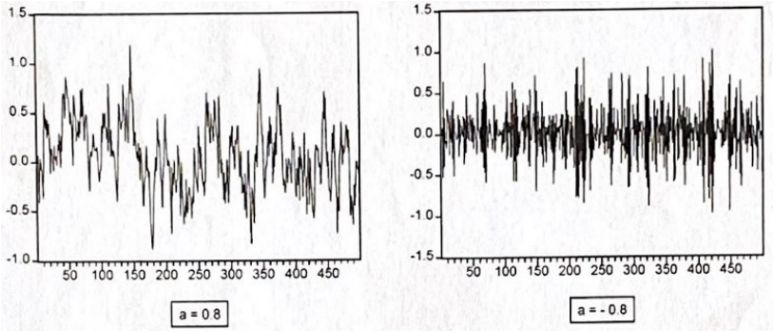
При этом

- $E(X_t) = 0$;
- $Corr(X_t, X_{t+\tau} = \rho(\tau) = a^\tau$;
- $Cov(X_t, X_{t+\tau} = \left[\frac{a^\tau}{(1-a^2)}\right]\sigma_\varepsilon^2$;
- $D(X_t) = \sigma_x^2 = \frac{\sigma_\varepsilon^2}{1-a^2}$

Рассмотренная модель порождает (при указанных условиях) стационарный ряд, имеющий нулевое математическое ожидание. Однако ее можно легко распространить и на временные ряды y_t с ненулевым математическим ожиданием $E(Y_t) = \mu$, полагая, что указанная модель относится к центрированному ряду $X_t = Y_t - \mu$:
$Y_t - \mu = a(Y_{t-1} - \mu) + \varepsilon_t, t = 1, ..., n,$
так что
$Y_t = aY_{t-1} + \sigma + \varepsilon_t, t = 1, ..., n,$
где
$\sigma = \mu(1-a).$

Поэтому без ограничения общности можно обойтись в текущем рассмотрении моделями авторегрессии, порождающими стационарный процесс с нулевым средним.

Продолжая рассмотрение для ранее определенного процесса X_t (с нулевым математическим ожиданием), заметим, что для него
$\gamma(1) = E(X_t \cdot X_{t-1}) = E[(aX_{t-1} + \varepsilon_t) \cdot X_{t-1}] = a\gamma(0),$
так что
$\rho(1) = \gamma(1)/\gamma(0) = a,$
и при значениях $a > 0$, близких к 1, между соседними наблюдениями имеется сильная положительная корреляция, что обеспечивает более гладкий характер поведения траекторий ряда по сравнению с процессом белого шума. При $a < 0$ процесс авторегрессии, напротив, имеет менее гладкие реализации, поскольку в этом случае проявляется тенденция чередования знаков последовательных наблюдений. Следующие два графика демонстрируют поведение смоделированных реализаций временных рядов, порожденных моделями авторегрессии $X_t = aX_{t-1} + \varepsilon_t$ с $\sigma_\varepsilon^2 = 0.2$ при $a = 0.8$ (первый график) и $a = -0.8$ (второй график).



Теперь мы должны обратить внимание на следующее важное обстоятельство. В практических ситуациях "стартовое"значение $X_0 = x_0$, на основе которого в соответствии с соотношением $X_t = aX_{t-1} + \varepsilon_t$ строятся последующие значения ряда X_t, может относиться к концу предыдущего периода, на котором просто в силу других эконометрических условий эволюция соответствующего экономического показателя следует иной модели, например, модели $X_t = aX_{t-1} + \varepsilon_t$ с другими значениями a и σ_ε^2. Более того, статистические данные о поведении ряда до момента $t = 0$ могут отсутствовать вовсе, так что значение x_0 является просто некоторой наблюдаемой числовой величиной. В обоих случаях ряд X_t уже не будет стационарным даже при $a < 1$. Рассмотрим подробнее характеристики и поведение ряда в таких ситуациях.

Если не конкретизировать модель, в соответствии с которой порождались наблюдения до момента $t = 1$, то значение x_0 можно рассматривать как фиксированное. При этом

$X_t = a^t x_0 + a^{t-1}\varepsilon_1 + a^{t-2}\varepsilon_2 + ... + \varepsilon_t,$

$E(X_t) = a^t x_0 + a^{t-1}E(\varepsilon_1) + a^{t-2}E(\varepsilon_2) + ... + E(\varepsilon_t) = a^t x_0,$

$D(X_t) = (a^{2(t-1)} + a^{2(t-2)} + ... + 1)\sigma_\varepsilon^2 = \left[\frac{(1-a^{2t})}{(1-a^2)}\right]\sigma_\varepsilon^2,$

$Cov(X_t, X_{t+\tau}) = Cov(X_t - a^t x_0, X_{t+\tau} - a^{t+\tau}x_0) = a^\tau(1 + a^2 + ... + a^{2(t-1)})\sigma_\varepsilon^2$

$$= a^\tau\left[\frac{(1-a^{2t})}{(1-a^2)}\right]\sigma_\varepsilon^2,$$

так что и математическое ожидание и дисперсия случайной величины X_t, а также ковариация $Cov(X_t, X_{t+\tau})$ зависят от t.

В то же время, если $a < 1$, то при $t \longrightarrow \infty$ получаем
$E(X_t) \longrightarrow 0, D(X_t) \longrightarrow \sigma_\varepsilon^2/(1-a^2), Cov(X_t, X_{t+\tau}) \longrightarrow a^\tau[\sigma_\varepsilon^2/(1-a^2)]$, т.е. при $t \longrightarrow \infty$ значения математического ожидания и дисперсии случайной величины X_t, а также автоковариации $Cov(X_t, X_{t+\tau})$ стабилизируются, приближаясь к своим предельным значениям.

С этой точки зрения, условие $a < 1$ можно трактовать как условие стабильности ряда, порождаемого моделью $X_t = aX_{t-1} + \varepsilon_t$ при фиксированном значении $X_0 = x_0$. Рассмотрим в этой ситуации наряду с только что исследованным рядом X_t,

$$X_t = a^t x_0 + \sum_{k=0}^{t-1} a^k \varepsilon_{t-k},\ a < 1,$$

ряд порождаемый моделью

$$\widetilde{X}_t = \sum_{k=0}^{\infty} a^k \varepsilon_{t-k}.$$

Имеем:

$$\widetilde{X}_t - X_t = -a^t x_0 + \sum_{k=0}^{\infty} a^k \varepsilon_{t-k};$$

при $t \longrightarrow \infty$

$$a^t x_0 \longrightarrow 0 \text{ и } E\left|\sum_{k=0}^{\infty} a^k \varepsilon_{t-k}\right|^2 = \sigma_\varepsilon^2 \sum_{k=t}^{\infty} a^{2k} \longrightarrow 0.$$

$$\lim_{t\longrightarrow\infty} X_t = \widetilde{X}_t$$

Таким образом, ряд $\widetilde{X}_t$ является предельным для X_t; ряд X_t "выходит на режим"$\widetilde{X}_t$ при $t \longrightarrow \infty$. При этом выход ряда X_t на режим $\widetilde{X}_t$ происходит быстрее, чем ближе X_0 и a к нулю.

Для ряда $\widetilde{X}_t$

$$E(\widetilde{X}_t) = E\left(\sum_{k=0}^{\infty} a^k \varepsilon_{t-k}\right) = \sum_{k=0}^{\infty} a^k E(\varepsilon_{t-k}) = 0,$$

$$D(\widetilde{X}_t) = D\left(\sum_{k=0}^{\infty} a^k \varepsilon_{t-k}\right) = \sum_{k=0}^{\infty} a^k D(\varepsilon_{t-k}) = \sigma_\varepsilon^2 \sum_{k=0}^{\infty} a^{2k} = \frac{\sigma_\varepsilon^2}{1-a^2},$$

$$Cov(\widetilde{X}_t, \widetilde{X}_{t+\tau}) = E\left[\left(\sum_{k=0}^{\infty} a^k \varepsilon_{t-k}\right), \left(\sum_{k=0}^{\infty} a^k \varepsilon_{t+\tau-k}\right)\right] = a^\tau \left(\sum_{k=0}^{\infty} a^{2k} E(\varepsilon_{t-k}^2)\right) =$$

$$= a^\tau \frac{\sigma_\varepsilon^2}{1-a^2}$$

так что $\widetilde{X}_t$ - стационарный ряд (в широком смысле). Кроме того,

$$\widetilde{X}_{t-1} = \frac{1}{a} \sum_{k=1}^{\infty} a^k \varepsilon_{t-k},$$

так что

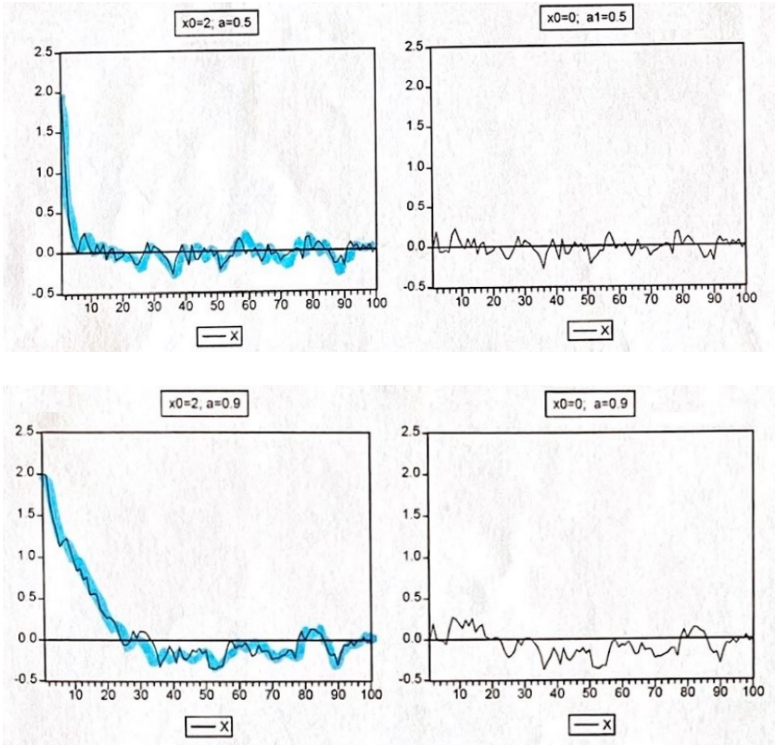
$$a\widetilde{X}_{t-1} + \varepsilon_t = \sum_{k=1}^{\infty} a^k \varepsilon_{t-k} = \widetilde{X}_t,$$

т.е. $\widetilde{X}_t$ удовлетворяет соотношению

$$\widetilde{X}_t = a\widetilde{X}_{t-1} + \varepsilon_t.$$

Поскольку ε_t не входит в правую часть выражений для $\widetilde{X}_{t-1}, \widetilde{X}_{t-2}, ...$, то случайная величина ε_t не коррелирована с $\widetilde{X}_{t-1}, \widetilde{X}_{t-2}, ...$, т.е. ε_t является инновацией (обновлением). В итоге получаем, что $\widetilde{X}_t$ - стационарный процесс авторегрессии певрого порядка, и фактически именно этот процесс имеется в виду, когда говорят о стационарном процессе AR(1).

Проиллюстрируем сказанное выше с помощью смоделированных реализацией ряда x_t, порожденных моделью $X_t = aX_{t-1} + \varepsilon_t$, с $\sigma_\varepsilon^2 = 0.2$ и различными значениями коэффициента a и стартового значения x_0.



Рассмотренную только что модель $X_t = aX_{t-1} + \varepsilon_t$ называют процессом авторегрессии первого порядка. **Процесс авторегрессии порядка p** (в кратком обозначении - **AR(p)**) определяется соотношениями

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + ... + a_p X_{t-p} + \varepsilon_t, a_p \neq 0,$$

где ε_t - процесс белого шума с $D(\varepsilon_t) = \sigma_\varepsilon^2$. Для простоты мы будем теперь сразу полагать, что $Cov(X_{t-s}, \varepsilon_t) > 0$ для всех $s > 0$; при этом говорят, что случайные величины ε_t образуют **инновационную (обновляющую) последовательность**, а случайная величина ε_t называется **инновацией** для наблюдения в момент t. Такая терминология

объясняется тем, что наблюдаемое значение ряда в момент t получается здесь как линейная комбинация p предшествующих значений этого ряда плюс не коррелированная с этими предшествующими значениями случайная составляющая ε_t, отражающая обновленную информацию, скажем, о состоянии экономики, на момент t, влияющую на наблюдаемое значение X_t.

При рассмотрении процессов авторегрессии и некоторых других моделей удобно использовать **оператор запаздывания L (lag operator)**, который воздействует на временной ряд и определяется соотношением

$LX_t = X_{t-1}$

в некоторых руководствах его называют оператором обратного сдвига и используют для него обозначение **B (backshift operator)**.

Если оператор запаздывания применяется k раз, что обозначается как L^k, то это дает в результате

$$L^k X_t = X_{t-k}$$

Выражение

$$a_1 X_{t-1} + a_2 X_{t-2} + ... + a_p X_{t-p} =$$

можно записать теперь в виде

$$= (a_1 L + a_2 L^2 + ... + a_p L^p) X_t$$

а соотношение, определяющее процесс авторегрессии p-го порядка, в виде

$$a(L) X_t = \varepsilon_t$$

где

$$a(L) = 1 - (a_1 L + a_2 l^2 + ... + a_p L^p).$$

Для того, чтобы такой процесс был стационарным, все корни алгребраического уравнения

$$a(z) = 0$$

т.е. корни z: $|z| > 1$ (вещественные и комплексные) должны лежать вне единичного круга $|z| \leq 1$. (В частности, для процесса AR(1) имеем

$a(z) = 1 - az$, уравнение $a(z) = 0$ имеет корень $z = 1/a$, и условие стационарности $z > 1$ равносильно уже знакомому нам условию $a < 1$.) При этом **решение уравнения** $a(L)X_t = \varepsilon_t$, можно представить в виде

$$X_t = \frac{1}{a(L)} \varepsilon_t = \sum_{j=0}^{\inf} b_j \varepsilon_{t-j}$$

где $\sum_{j=0}^{\inf} |b_j| < \inf$ откуда, в частности, следует, что

$$E(X_t) = E\left(\sum_{j=0}^{\inf} b_j \varepsilon_{t-j} \right) = \sum_{j=0}^{\inf} b_j E(\varepsilon_{t-j}) = 0$$

Стационарный процесс $AR(p)$ с нулевым математическим ожидаением μ удовлетворяет соотношению

$$a(L)(X_t - \mu) = \varepsilon_t$$

или

$$a(L) X_t = \delta + \varepsilon_t$$

где

$$\delta = a(L)\mu = \mu(1 - a_1 - a_2 - ... - a_p) = \mu a(1)$$

При этом **решение уравнения** $a(L)(X_t - \mu) = \varepsilon_t$ имеет вид

$$X_t = \mu + \frac{1}{a(L)} \varepsilon_t$$

Таким образом, если стационарный процесс $AR(p)$ задан в виде $a(L)X_t = \delta + \varepsilon_t$, то следует помнить о том, что в этом случае математическое ожидание этого процесса равно не δ, а

$$E(x_t) = \mu = \frac{\delta}{(1 - a_1 - a_2 - ... - a_p)}$$

Для процесса $AR(1)$ имеем $a(L) = 1 - aL$, так что (**вне зависимости от того, равно μ нулю или нет**).

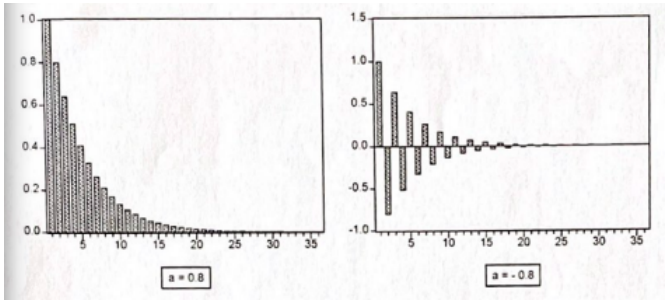
$$X_t - \mu = (\frac{1}{aL} \varepsilon_t) = (1 + aL + a^2 L^2 + ...) \varepsilon_t = \varepsilon_t + a\varepsilon_{t-1} + a^2 \varepsilon_{t-2} + ...$$

Из последнего выражения сразу видно, что

$$\rho(k) = Corr(X_t, X_{t+k}) = a^k \ , \ k = 0, 1, 2, ...$$

При $0 < a < 1$ коррелограмма (график функции $\rho(k)$ для $k = 0, 1, 2, ...$) отражает **показательное убывание** корреляции с возрастанием интервала между наблюдениями; при $-1 < a < 0$ (коррелограмма имеет **характер затухающей косинусоиды**).

Сравним поведение коррелограммы стационарного процесса $AR(1)$ при $a = \pm 0.8$



Коррелограмма процесса $AR(p)$ при $p > 1$ имеет более сложную форму, зависящую от расположения (на комплексной плоскости) корней уравнения $a(z) = 0$. Однако для больших значений k автокорреляции $\rho(k)$ хорошо апроксимируется значением $A\theta^k$, где $\theta = 1/z_{\min}$ и $z_{\min}$ - наименьший по абсолютной величине корень уравнения $a(z) = 0$, если этот корень являетя вещественным и положительным, или заключена в интервале $\pm A\theta^k$ в противном случае. Здесь $A > 0$ - некоторая постоянная, определяемая коэффицентами $a_1, ..., a_p$

Если умножить на $X_{t-k}(k > 0)$ обе части соотношения, определяющего процесс $AR(p)$, и после этого взять от обеих частей математическое ожидание, то получим соотношение

$$\gamma(k) = a_1\gamma(k-1) + a_2\gamma(k-2) + ... + a_p\gamma(k-p) \ , \ k > 0$$
$$\gamma(k) = Cov(X_t, X_{t-k})$$

Разделив обе части последнего на $\gamma(0)$, приходим к **системе Юла-Уокера**

$$\rho(k) = a_1\rho(k-1) + a_2\rho(k-2) + ... + a_p\rho(k-p) \ , \ k > 0$$

Эта система позволяет последовательно находить значения автокорреляции и дает возможность, используя первые p уравнений, выразить коэффициенты a_j через значения первых p автокорреляций, что можно непосредственно использовать при подборе модели авторегрессии к реальным статистическим данным (см. разделы 3.1 и 3.2)

4 Пример

Рассмотрим процесс авторегрессии $AR(2)$

$$X_r = 4.375 + 0.25X_{t-1} - 0.125X_{t-2} + \varepsilon_t$$

Уравнение $a(z) = 0$ принимает в этом случае вид

$$1 - 0.25z + 0.125z^2 = 0 \ \textbf{или} \ z^2 - 2z + 8 = 0$$

и имеет корни $z_{1,2} = 1 \pm i\sqrt{7}$. Оба корня по абсолютной величине больше единицы, так что процесс стационарный. Математичесое ожидание этого процесса равно

$$\mu\delta/(1 - a_1 - a_2) = 4.375/(1 - 0.25 - 0.125) = 5$$

так что траектории этого процесса флуктурирует вокруг уровня 5.

Для построения коррелограммы воспользуемся уравнением Юла-Уокера. У нас $p = 2$, так что

$$\rho(k) = 0.25\rho(k-1) - 0.125\rho(k-2) \ , \ k > 0$$

По определению, $\rho(0) = 1$. Для $\rho(1)$ имеем

$$\rho(1) = 0.25\rho(0) - 0.125\rho(-1) = 0.25 - 0.125\rho(1)$$

откуда находим:

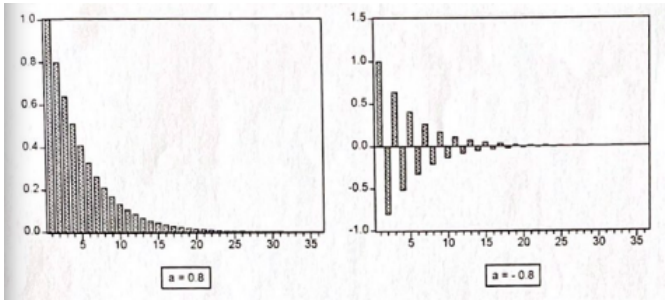
$$\rho(1) = 0.25/(1 + 0.125) = 2/9 = 0.222$$

Далее последовательно находим:

$$\rho(2) = 0.25\rho(1) - 0.125\rho(0) = 0.25 * 0.222 - 0.125 = -0.069$$

$$\rho(3) = -0.045, \ \rho(4) = -0.003, \ \rho(5) = 0.005 \ \textbf{и т.д.}$$

Корреляции даже между соседними наблюдениями очени малы, и можно ожидать, что поведение траекторий такого ряда не очень существенно от поведения реализаций процесса белого шума. Теоретическая коррелограмма расматриваемого процесса и смоделированная реализация этого процесса приведены ниже.



Лекция № 6.

1. Процесс скользящего среднего.

Еще одной простой моделью порождения временного ряда является процесс **скользящего среднего порядка** q $(MA(q))$. Согласно этой модели,

$$X_t = \varepsilon_t + b_1\,\varepsilon_{t-1} + b_2\,\varepsilon_{t-2} + \ ... \ + b_q\,\varepsilon_{t-q}, \quad b_q \neq 0,$$

где ε_t — процесс белого шума.

Такой процесс имеет нулевое математическое ожидание. Модель можно обобщить до процесса, имеющего ненулевое математическое ожидание μ, полагая

$$X_t - \mu = \varepsilon_t + b_1\,\varepsilon_{t-1} + b_2\,\varepsilon_{t-2} + \ ... \ + b_q\,\varepsilon_{t-q}\,,$$

т.е.

$$X_t = \mu + \varepsilon_t + b_1\,\varepsilon_{t-1} + b_2\,\varepsilon_{t-2} + \ ... \ + b_q\,\varepsilon_{t-q}.$$

Для процесса скользящего среднего порядка q используется обозначение $MA(q)$ (скользящее среднее — **moving average**).

При $q = 0$ и $\mu = 0$ получаем процесс белого шума. Если $q = 1$, то

$$X_t - \mu = \varepsilon_t + b\,\varepsilon_{t-1}$$

— скользящее среднее первого порядка. В последнем случае

$$D(X_t) = (1 + b^2)\,\sigma_\varepsilon^2,$$

$$\gamma(1) = Cov(X_t, X_{t-1}) = E[(X_t - \mu)(X_{t-1} - \mu)] = b\,\sigma_\varepsilon^2$$

$$\gamma(k) = Cov(X_t, X_{t-k}) = E[(X_t - \mu)(X_{t-k} - \mu)] = 0, k > 1,$$

так что процесс X_t является стационарным с $E(X_t) = \mu$, $D(X_t) = (1 + b^2)\,\sigma_\varepsilon^2$,

Автоковариации:

$$\gamma(k) = \begin{cases} (1 + b^2)\,\sigma_\varepsilon^2, & k = 0 \\ b\,\sigma_\varepsilon^2, & k = 1 \\ 0, & k > 1 \end{cases}$$

Автокорреляции этого процесса равны

$$\rho(k) = \begin{cases} 1, & k = 0 \\ b/(1 + b^2), & k = 1 \\ 0, & k > 1 \end{cases}$$

т.е коррелограмма процесса имеет весьма специфический вид. Коррелированными оказываются только соседние наблюдения. Корреляция между ними положительна, если $b > 0$, и отрицательна при $b < 0$. Соответственно, процесс $MA(1)$ с $b > 0$ имеет более гладкие, по сравнению с белым шумом, реализации, а процесс $MA(1)$ с $b < 0$ имеет менее гладкие, по сравнению с белым шумом, реализации. Заметим, что для любого процесса $MA(1)$ $\rho(1) \leq 0.5$, т.е. корреляционная связь между соседними наблюдениями невелика, тогда как у процесса $AR(1)$ такая связь может быть сколь угодно сильной (при значениях a, близких к 1). $\rho(k) = a^k$ для $AR(1)$

Модель $MA(q)$ кратко можно записать в виде

$$X_t - \mu = b(L)\,\varepsilon_t,$$

где

$$b(L) = 1 + b_1\,L + \ldots + b_q\,L^q.$$

Для автоковариации имеем:

$$\gamma(k) = E[(X_t - \mu)(X_{t-k} - \mu)] = \begin{cases} \left(\sum\limits_{j=0}^{q-k} b_j b_{j+k}\right)\sigma_\varepsilon^2, & 0 \leq k \leq q,\ b_0 = 1, \\ 0, & k > q, \end{cases}$$

так что $MA(q)$ является стационарным процессом с ненулевым математическим ожиданием, дисперсией

$$\sigma_X^2 = (1 + b_1^2 + \ldots + b_q^2)\sigma_\varepsilon^2$$

и автокорреляциями

$$\rho_k = \begin{cases} \left(\sum\limits_{j=0}^{q-k} b_j b_{j+k}\right) \Big/ \left(\sum\limits_{j=0}^{q} b_j^2\right), & k = 0, 1, \ldots, q, \\ 0, & k = q+1, q+2, \ldots. \end{cases}$$

Здесь статистическая связь между наблюдениями сохраняется в течение q единиц времени (т.е. "длительность памяти"процесса равна q).

Подобного рода временные ряды соответствуют ситуации, когда некоторый экономический показатель находится в равновесии, но отклоняется от положения равновесия в силу последовательно возникающих непредсказуемых событий, причем система такова, что влияние таких событий отмечается на протяжении некоторого периода времени.

2. Обратимость процессов $AR(p)$ и $MA(q)$.

Если влияние прошлых событий ослабевает с течением времени показательным образом, так что $b_j = a^j$, $0 < a < 1$, то искусственное предположение о том, что ряд ε_t начинается в "бесконечном прошлом приводит к модели бесконечного скользящего среднего $MA(\infty)$

$$X_t = \sum_{j=0}^{\infty} a^j \, \varepsilon_{t-j} = \sum_{j=0}^{\infty} b_j \, \varepsilon_{t-j}, \quad \text{где} \sum_{j=0}^{\infty} |b_j| < \infty$$

Ранее мы видели, что такое же представление допускает стационарный процесс авторегрессии первого порядка $AR(1)$

$$X_t = aX_{t-1} + \varepsilon_t, \quad a < 1$$

т.е. в рассматриваемом случае процесс $MA(\infty)$ эквивалентен процессу $AR(1)$.

Вообще всякий стационарный процесс $AR(p)$ можно записать в форме процесса $MA(\infty)$:

$$X_t = \mu + \frac{1}{a(L)}\varepsilon_t = \mu + \sum_{j=0}^{\infty} b_j \varepsilon_{t-j} = \mu + b(L)\varepsilon_t,$$

где

$$b(L) = \sum_{j=0}^{\infty} b_j L^j = \frac{1}{a(L)} \quad \text{и} \quad \sum_{j=0}^{\infty} |b_j| < \infty$$

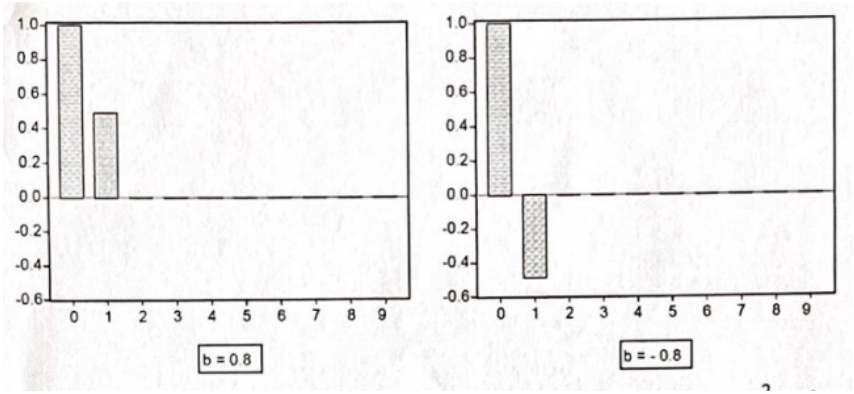
Примеры.

a) Рассмотрим процесс $MA(1)$ с $b = 0.8$ и $\mathbb{E}(X_t) = 6$, т.е. $X_t = +\varepsilon_t + 0.8\varepsilon_{t-1}$.

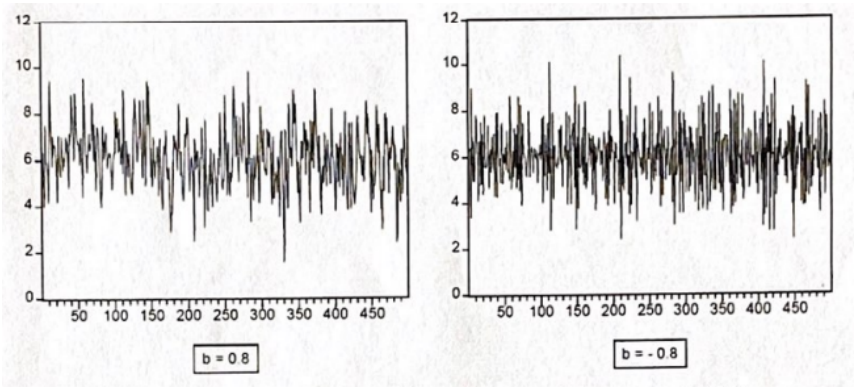
Для него $\rho(1) = 0.8/(1 + 0.8^2) = 0.488$.

b) Для процесса $MA(1)$ с $b = -0.8$ и $\mathbb{E}(X_t) = 6$ имеем $\rho(1) = -0.8/(1 + 0.8^2) = -0.488$.

Коррелограммы этих двух процессов имеют вид:



Смоделированные реализации этих двух процессов с $\sigma_\varepsilon^2 = 1$:

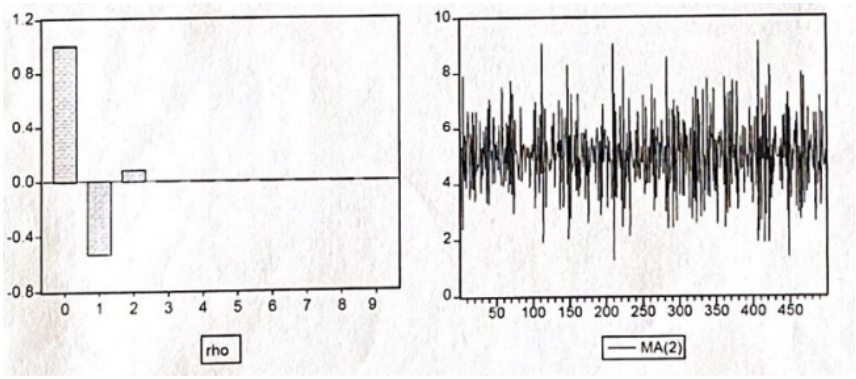


c) Для $MA(2)$ процесса $X_t = 5 + \varepsilon_t - 0.75\,\varepsilon_{t-1} + 0.125\,\varepsilon_{t-2}$ имеем:

$$\rho(1) = (b_0\,b_1{+}b_1\,b_2)/({b_0}^2{+}{b_1}^2{+}{b_2}^2) = (-0.75{-}0.75{\cdot}0.125)/(1{+}0.75^2{+}0.125^2) = -0.535$$

$$\rho(2) = 0.125/1.578 = 0.079$$

Ниже приводятся коррелограмма и смоделированная реализация этого процесса.



$$AR(p) : X_t = \delta + a_1 x_{t-1} + a_2 x_{t-2} + ... + a_p x_{t-p} + \epsilon_t, a_p \neq 0$$

$$E\,(X_t) = \mu = \frac{\delta}{1 - a_1 - a_2 - \ldots - a_p}$$

$$\rho(k) = a_1^k - AR(1), k = 0, 1, 2, \ldots$$

$$\rho(k) = a_1\rho(k-1) + a_2\rho(k-2) + \ldots + a_p\rho(k-p), k > 0$$

$$MA(q) : X_t = \mu + \epsilon_t + b_1\epsilon_{t-1} + \ldots + b_q\epsilon_{t-q}$$

$$\sigma_x^2 = (1 + b_1^2 + \ldots + b_q^2)\sigma_\epsilon^2$$

$$\rho(k) = \begin{cases} \left(\sum_{j=0}^{q-k} b_j b_{j+k}\right) / \left(\sum_{j=0}^{q} b_j^2\right), & k = 0, 1, \ldots q \\ 0, & k = q+1, q+2, \ldots \end{cases}$$

$$X_t = aX_{t-1} + \epsilon_t$$

1)$X_t = \phi_{11}X_{t-1} + \epsilon_t$. Умножаем на X_{t-1}:

$$\frac{E\left(X_t, X_{t-1}\right)}{Cov\left(X_t, X_t\right)} = \phi_{11}\frac{E\left(X_{t-1}^2\right)}{Cov\left(X_t, X_t\right)} + \frac{E\left(X_{t-1}, \epsilon_t\right)}{Cov\left(X_t, X_t\right)}$$

$$\rho(1) = \phi_{11}\cdot 1 + 0 = a$$

2) a)$X_t = \phi_{21}X_{t-1} + \phi_{22}Xt - 2 + \epsilon_t$. Умножаем на X_{t-1}:

$$\frac{E\left(X_t, X_{t-1}\right)}{Cov\left(X_t, X_t\right)} = \phi_{21}\frac{E\left(X_{t-1}^2\right)}{Cov\left(X_t, X_t\right)} + \phi_{22}\frac{E\left(X_{t-2}, X_{t-1}\right)}{Cov\left(X_t, X_t\right)} + \frac{E\left(\epsilon_t, X_{t-1}\right)}{Cov\left(X_t, X_t\right)}$$

$$\rho(1) = \phi_{21} + \phi_{22}\cdot\rho(1)$$

b)$X_t = \phi_{21}X_{t-1} + \phi_{22}X_{t-2} + \epsilon_t$. Умножаем на X_{t-2}:

$$\frac{E\left(X_{t-2}, X_t\right)}{Cov\left(X_t, X_t\right)} = \phi_{21}\frac{E\left(X_{t-1}, X_{t-2}\right)}{Cov\left(X_t, X_t\right)} + \phi_{22}\frac{E\left(X_{t-2}^2\right)}{Cov\left(X_t, X_t\right)} + \frac{E\left(\epsilon_t, X_{t-2}\right)}{Cov\left(X_t, X_t\right)}$$

$$\rho(2) = \phi_{21}\cdot\rho(1) + \phi_{22}$$

$$\begin{cases} \phi_{21} + \phi_{22}\cdot\rho(1) = \rho(1), \\ \phi_{21}\cdot\rho(1) + \phi_{22} = \rho(2). \end{cases}$$

Нужно найти ϕ_{22} –?

$$\begin{bmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{bmatrix}\begin{bmatrix} \phi_{21} \\ \phi_{22} \end{bmatrix} = \begin{bmatrix} \rho(1) \\ \rho(2) \end{bmatrix}$$

Решаем данное уравнение методом Крамера:

$$\phi_{22} = \frac{\Delta_2}{\Delta} = \frac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2}$$

где Δ_1, Δ_2 и Δ определяется следующим образом

$$\Delta_1 = \begin{vmatrix} \rho(1) & \rho(1) \\ \rho(2) & 1 \end{vmatrix}, \quad \Delta_2 = \begin{vmatrix} 1 & \rho(1) \\ \rho(1) & \rho(2) \end{vmatrix}, \quad \Delta = \begin{vmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{vmatrix}$$

Т.к $\rho(k) = a^k$ для любого k, то

$$\phi_{22} = \frac{a^2 - a^2}{1 - a^2} = 0$$

$ARMA(1,1)$:

$$X_t = aX_{t-1} + \epsilon_t + b\epsilon_{t-1}$$

$$X_t = \frac{1 + bL}{1 - aL}\epsilon_t = (1 + bL)\left[1 + aL + a^2L^2 + a^3L^3 + \cdots\right]\epsilon_t =$$

$$= \left[1 + (a + b)L + a(a + b)L^2 + a^2(a + b)L^3 + \cdots\right]\epsilon_t$$

где $\frac{1}{1-aL}$ – бесконечно убывающая геометрическая прогрессия

1) Посчитаем дисперсию от X_t:

$$Var(X_t) = \left[1 + (a + b)^2a^2(a + b)^2 + a^4(a + b)^2 + \cdots\right]\sigma_\epsilon^2 =$$

$$= \left(1 + \frac{(a + b)^2}{1 - a^2}\right)\sigma_\epsilon^2 = \frac{1 - a^2 + a^2 + 2ab + b^2}{1 - a^2}\sigma_\epsilon^2 = \frac{1 + 2ab + b^2}{1 - a^2}\sigma_\epsilon^2$$

2)$X_t = aX_{t-1} + \epsilon_t + b\epsilon_{t-1}$. Умножаем на X_{t-1}:

$$E\left(X_t, X_{t-1}\right) = aE\left(X_{t-1}, X_{t-1}\right) + bE\left(\epsilon_{t-1}, X_{t-1}\right)$$

$$\gamma(1) = a\gamma(0) + b\cdot Cov\left(\epsilon_{t-1}, X_{t-1}\right)$$

3)$X_{t-1} = aX_{t-2} + \epsilon_{t-1} + b\epsilon_{t-2}$. Умножаем на ϵ_{t-1}:

$$E\left(X_{t-1}, \epsilon_{t-1}\right) = aE\left(X_{t-2}, \epsilon_{t-1}\right)\sigma_\epsilon^2 + bE\left(\epsilon_{t-2}, \epsilon_{t-1}\right)$$

$$Cov\left(X_{t-1}, \epsilon_{t-1}\right) = \sigma_\epsilon^2$$

$$\gamma(1) = a\gamma(0) + b\sigma_\epsilon^2.$$

$$\frac{\gamma(1)}{\gamma(0)} = a + \frac{b\sigma_\epsilon^2}{(1 + b^2 + 2ab)\sigma_\epsilon^2}(1 - a^2) = \frac{a + a^2b + 2ab^2 + b - a^2b}{1 + b^2 + 2ab} = \frac{(a + b)(1 + ab)}{1 + b^2 + 2ab}.$$

4)$X_t = aX_{t-1} + \epsilon_t + b\epsilon_{t-1}$. Умножаем на X_{t-2}:

$$E\left(X_t, X_{t-2}\right) = aE\left(X_{t-1}, X_{t-2}\right) + E\left(\epsilon_t, X_{t-2}\right) + bE\left(\epsilon_{t-1}, X_{t-2}\right)$$

$$\gamma(2) = a\cdot\gamma(1)$$

Лекция №7

2.5. Смешанный процесс авторегрессии – скользящего среднего (процесс авторегрессии с остатками в виде скользящего среднего)

ARMA(p,q):

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + ... + a_p X_{t-p} + \epsilon_t + b_1 \epsilon_{t-1} + ... + b_q \epsilon_{t-q}, \quad a_p \neq 0, \;\; b_q \neq 0, \;\; \epsilon_t - \textit{белый шум} \quad (1)$$

Процесс X_t с нулевым математическим ожиданием, принадлежащий такому классу процессов, характеризуется порядками p и q его AR и MA составляющих и обозначается как процесс **ARMA(p, q) (autoregressive moving average, mixed autoregressive moving average)**. Более точно, процесс X_t с нулевым математическим ожиданием принадлежит классу $ARMA(p,q)$, если:

$$X_t = \sum_{j=1}^{p} a_j X_{t-j} + \sum_{j=0}^{q} b_j \epsilon_{t-j}, \quad a_p \neq 0, \;\; b_q \neq 0, \;\; \mathbb{E}\, X_t = 0 \quad (2)$$

где ϵ_t – процесс белого шума и $b_0 = 1$. В операторной форме последнее соотношение имеет вид

$$a(L) = 1 - (a_1 L + a_2 L^2 + ... + a_p L^p) \quad (3)$$

$$b(L) = 1 + b_1 L + b_2 L^2 + ... + b_q L^q \quad (4)$$

$$a(L) X_t = b(L) \epsilon_t, \quad (5)$$

т.е. $a(L)$ и $b(L)$ имеют тот же вид, что и в определенных ранее моделях $AR(p)$ и $MA(q)$. Если процесс имеет постоянное математическое ожидание μ , то он является процессом типа $ARMA(p,q)$, если

$$X_t - \mu = \sum_{j=1}^{p} a_j (X_{t-j} - \mu) + \sum_{j=0}^{q} b_j \epsilon_{t-j} \quad (\mathbb{E}\, X_t = \mu \neq 0) \quad (6)$$

Отметим следующие свойства процесса $ARMA(p,q)$ в широком смысле с $\mathbb{E}(X_t) = \mu$.

- Процесс стационарен, если все корни уравнения $a(z) = 0$ лежат вне единичного круга $|z| \leq 1$
- Если процесс стационарен, то существует эквивалентный ему процесс $MA(\infty)$

$$X_t - \mu = \sum_{j=0}^{\infty} c_j \epsilon_{t-j}, \quad c_0 = 1, \;\; \sum_{j-0}^{\infty} |c_j| < \infty \quad (7)$$

или

$$X_t - \mu = c(L) \epsilon_t, \quad \textit{где} \quad c(z) = \sum_{j=0}^{\infty} c_j z^j = \frac{b(z)}{a(z)} \quad (8)$$

- Если все корни уравнения $b(z) = 0$ лежат вне единичного круга $|z| \leq 1$ **(условие обратимости)**, то существует эквивалентное представление процесса X_t в виде процесса авторегрессии бесконечного порядка $AR(\infty)$

$$X_t - \mu = \sum_{j=1}^{\infty} d_j (X_{t-j} - \mu) + \epsilon_t, \quad (9)$$

или

$$d(L)(X_t - \mu) = \epsilon_t, \quad \textit{где} \quad d(z) = 1 - \sum_{j=1}^{\infty} d_j z^j = \frac{a(z)}{b(z)} \quad (10)$$

Отсюда вытекает, что стационарный процесс $ARMA(p,q)$ всегда можно аппроксимировать процессом скользящего среднего достаточно высокого порядка, а при выполнении условия обратимости его можно также аппроксимировать процессом авторегрессии достаточно высокого порядка.

Специфику формы коррелограммы процесса $ARMA(p,q)$ в общем случае указать труднее, чем для моделей $AR(p)$ и $MA(q)$. Отметим только, что для значений $k > q$ коррелограмма процесса $a(L)X_t = b(L)\epsilon_t$ выглядит так же, как и коррелограмма процесса авторегрессии $a(L)X_t = \epsilon_t$. Так, для процесса $AR(1)$:

$$\rho(k) = a^k, \quad k = 0, 1, 2, ... \quad (11)$$

Для процесса $ARMA(1,1)$:

$$\rho(k) = a_1 \rho(k-1) \quad k = 2, 3, ... \quad (12)$$

как и у процесса $X_t = a_1 X_{t-1} + \epsilon_t$. При этом, однако, $\rho(1) \neq a_1$.

Предпосылкой для обоснования использования моделей $ARMA$ является следующий факт. Если $ARMA(p_1, q_1)$ ряд X_t и $ARMA(p_2, q_2)$ ряд Y_t статистически независимы между собой, и $Z_t = X_t + Y_t$, то типичным является положение, когда Z_t является $ARMA(p,q)$ рядом, у которого:

$$p = p_1 + p_2, \quad (13)$$

$$q = p_1 + q_2, \quad p_1 + q_2 > p_2 + q_1 \quad (14)$$

$$q = p_2 + q_1, \quad p_2 + q_1 > p_1 + q_2 \quad (15)$$

Возможны также ситуации, когда значения p и q оказываются меньше указанных значений. (Такие ситуации возникают в случаях, когда многочлены $a_X(z)$ и $a_Y(z)$, соответствующие авторегрессионным частям процессов X_t и Y_t , имеют общие корни.)

В частном случае, когда оба ряда имеют тип $AR(1)$, но с различными параметрами, их сумма имеет тип $ARMA(2,1)$.

$$X_t \sim AR(1), Y_t \sim AR(1) \Rightarrow z_t = X_t + Y_t \sim ARMA(1,1) \quad (16)$$

В экономике многие временные ряды являются агрегированными. Из указанного выше факта вытекает, что если каждая из компонент отвечает простой модели AR, то при независимости этих компонент их сумма будет $ARMA$ процессом. Такого же рода процесс мы получим, если часть компонент имеет тип AR, а остальные компоненты имеют тип MA. Единственным исключением является случай, когда все компоненты являются MA процессами – в этом случае в результате получаем MA процесс.

Предположим, наконец, что "истинный" экономический ряд отвечает $AR(p)$ модели, но значения этого ряда измеряются со случайными ошибками, образующими процесс белого шума (т.е. $MA(0)$). Тогда наблюдаемый ряд имеет тип $ARMA(p,p)$.

$$X_t \sim AR(p), Y_t \sim AR(0) \Rightarrow z_t = X_t + Y_t \sim ARMA(p,p) \tag{17}$$

Замечание

$$\sum AR \to ARMA \tag{18}$$

$$\sum AR + \sum MA \to ARMA \tag{19}$$

$$\sum MA \to MA \tag{20}$$

Ранее мы уже говорили о том, что если $ARMA(p,q)$ процесс X_t удовлетворяет условию обратимости, то его можно представить в виде стационарного процесса $AR(\infty)$. Последний, в свою очередь, можно аппроксимировать стационарным процессом $AR(p)$, быть может, достаточно высокого порядка.

Таким образом, в практических задачах можно было бы и вовсе обойтись без использования моделей $ARMA$, ограничиваясь либо AR либо MA моделями. При этом, однако, количество коэффициентов, подлежащих оцениванию, может оказаться слишком большим (что снижает точность оценивания) и даже превосходить количество имеющихся наблюдений. В этом смысле модели $ARMA$ могут быть "**более экономными**".

2.6. Модели ARMA, учитывающие наличие сезонности

$$X_t = \sum_{j=1}^{p} a_j X_{t-j} + \sum_{j=0}^{q} b_j \epsilon_{t-j}, \quad \textit{где} \quad b_0 = 1, a_p \neq 0, b_q \neq 0, \epsilon_1 - \textit{белый шум} \tag{21}$$

Если наблюдаемый временной ряд обладает выраженной сезонностью, то модель $ARMA$, соответствующая этому ряду, должна содержать составляющие, обеспечивающие проявление сезонности в порождаемой этой моделью последовательности наблюдений. Для квартальных данных чисто сезонными являются стационарные модели ***сезонной авторегрессии первого порядка (SAR(1))***.

$$X_t = a_4 X_{t-4} + \epsilon_t, \quad a_4 < 1 \tag{22}$$

и ***сезонного скользящего среднего первого порядка (SMA(1))***.

$$X_t = \epsilon_t b_4 \epsilon_{t-4}. \tag{23}$$

Пропущен +, то есть eps_t + b_4*eps_(t-4)

В первой модели автокорреляционные функции:

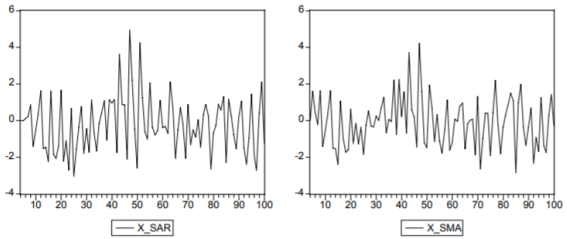
$$\rho(k) = a_4^{k/4}, \quad \textit{для} \quad k = 4m, m = 0, 1, 2, \ldots, \tag{24}$$

$$\rho(k) = 0 \quad \textit{для остальных} \quad k > 0. \tag{25}$$

Во второй модели

$$\rho(0) = 1, \quad \rho(4) = b_4, \quad \rho(k) = 0 \quad \textit{для остальных} \quad k > 0. \tag{26}$$

Ниже приведены смоделированные реализации модели SAR(1) с $a_4 = 0.8$ и модели SMA(1) с $b_4 = 0.8$.



Комбинации несезонных и сезонных изменений реализуются, например, в моделях $ARMA((1,4),1)$ и $ARMA(1,(1,4))$. Такие модели носят название аддитивных сезонных моделей.

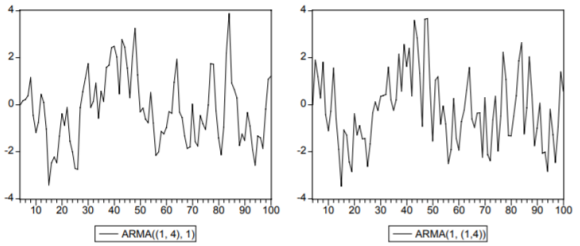
$ARMA((1,4),1)$

$$X_t = a_1 X_{t-1} + a_4 X_{t-4} + \epsilon_t + b_1 \epsilon_{t-1} \tag{27}$$

$ARMA(1,(1,4))$

$$X_t = a_1 X_{t-1} + \epsilon_t + b_1 \epsilon_{t-1} + b_4 \epsilon_{t-4}. \tag{28}$$

Следующие два графика показывают поведение смоделированных реализаций таких рядов при $a_1 = 2/3$, $a_4 = \check{}1/48$, $b_4 = 1/5$ у первого ряда и при $a_1 = 0.4$, $b_1 = 0.3$, $b_4 = 0.8$ у второго ряда. Таким образом можно так подобрать коэф-ты моделей, что эти реализации будут очень близки.



Заметим, что для первой модели уравнение $a(z) = 0$ принимает вид $1 - 2/3z + 1/48z^4 = 0$, т.е. $z^4 - 32z + 48 = 0$; корни этого уравнения $z_1 = 2$, $z_2 = 2$, $z_3 = -2 + i\sqrt{8}$, $z_4 = -2 - i\sqrt{8}$ лежат вне единичного круга, что обеспечивает стационарность рассматриваемого процесса. Во второй модели $a(z) = 0$ принимает вид $1 - 0.4, z = 0$; корень этого уравнения $z = 2.5 > 1$, так что и эта модель стационарна.

Кроме рассмотренных примеров аддитивных сезонных моделей, употребляются также и мультипли-

кативные спецификации, например,

$$(1 - a_1 L)X_t = (1 + b_1 L)(1 + b_4 L_4)\epsilon_t, \tag{29}$$

$$(1 - a_1 L)(1 - a_4 L^4)X_t = (1 + b_1 L)\epsilon_t. \tag{30}$$

Первая дает

$$X_t = a_1 X_{t-1} + \epsilon_t + b_1\epsilon_{t-1} + b_4\epsilon_{t-4} + b_1 b_4\epsilon_{t-5}, \tag{31}$$

а вторая

$$X_t = a_1 X_{t-1} + a_4 X_{t-4} - a_1 a_4 X_{t-5} + \epsilon_t + b_1\epsilon_{t-1} \tag{32}$$

В первой модели допускается взаимодействие составляющих скользящего среднего на лагах 1 и 4 (т.е. значений ϵ_{t-1} и ϵ_{t-4}), а во второй – взаимодействие авторегрессионных составляющих на лагах 1 и 4 (т.е. значений X_{t-1} и X_{t-4}). Конечно, эти две модели являются частными случаями аддитивных моделей

$$X_t = a_1 X_{t-1} + \epsilon_t + b_1\epsilon_{t-1} + b_4\epsilon_{t-4} + b_5\epsilon_{t-5} \tag{33}$$

$$X_t = a_1 X_{t-1} + a_4 X_{t-4} - a_5 X_{t-5} + \epsilon_t + b_1\epsilon_{t-1} \tag{34}$$

с $b_5 = b_1 b_4$, $a_5 = -a_1 a_4$. При приближенном выполнении последних соотношений (по крайней мере, если гипотезы о наличии таких соотношений не отвергаются), естественно перейти от оценивания аддитивной модели к оцениванию мультипликативной модели, опять следуя ***принципу "экономности"*** модели ***("parsimony model")***. Впрочем, каких-либо теоретических оснований, ведущих к предпочтению одной формы сезонности перед другой (мультипликативной или аддитивной), не существует.

Замечание 7.1.3. Пусть X_t - процесс типа $ARMA(p, q)$, $a(L)X_t = b(L)\epsilon_t$. Выше отмечалось, что если все корни алгебраического уравнения $a(z) = 0$ лежат вне единичного круга $|z| \le 1$ на комплексной плоскости, то X_t - стационарный процесс. Но такой процесс может быть стационарным и в случае, когда уравнение $a(z) = 0$ имеет корень z с $|z| = 1$. Поясним это следующим простым примером.

Пусть $X_t = X_{t-1}$:

$$X_t - X_{t-1} = \epsilon_t - \epsilon_{t-1}. \tag{35}$$

Последнее выражение записывается в виде:

$$(1 - L)X_t = (1 - L)\epsilon_t, \tag{36}$$

т.е. $a(L)X_t = b(L)\epsilon_t$, где $a(L) = 1 - L$ и $b(L) = 1 - L$. Иными словами, для процесса X_t получили $ARMA(1, 1)$ представление

$$X_t = X_{t-1} + \epsilon_t - \epsilon_{t-1}, \tag{37}$$

для которого уравнение $a(z) = 0$ имеет корень $|z| = 1$.

Замечание 7.1.4. В общем случае, если у $ARMA(p, q)$ процесса X_t, $a(L)X_t = b(L)\epsilon_t$, многочлены $a(z)$ и $b(z)$ *не имеют общих корней*, условие нахождения всех корней уравнения $a(z) = 0$ вне единичного круга $|z| \le 1$ является *необходимым и достаточным* для стационарности процесса X_t.

Замечание 7.1.5. Представить в виде процесса скользящего среднего бесконечного порядка можно не только стационарный процесс X_t типа $ARMA(p, q)$, но фактически и любой стационарный процесс, встречающийся на практике. Это вытекает из так называемого **разложения Вольда** (*Wold's decomposition*). Вольд в работе (*Wold*, 1938) показал, что любой стационарный в широком смысле процесс X, с нулевым математическим ожиданием может быть представлен в виде:

$$X_t = \sum_{j=0}^{\infty} c_j\epsilon_{t-j} + Z_t, \tag{38}$$

где $c_0 = 1$ и $\sum_{j=0}^{\infty} c_j^2 < \infty$, ϵ_t - процесс белого шума, Z_t - стационарный случайный процесс с нулевым математическим ожиданием, $Cov(Z_t, \epsilon_{t-j}) = 0$ для всех j, и значение Z_t можно сколь угодно точно предсказать на основании линейной функции от прошлых значений процесса X_t : $X_{t-1}, X_{t-2}, ...$

Тем самым стационарный процесс X_t представляется в виде суммы двух компонент: **линейно недетерминированной** (*linearly indeterministic*) компоненты $\sum_{j=0}^{\infty} c_j\epsilon_{t-j}$ и **линейно детерминированной** (*linearly deterministic*) компоненты Z_t. Если вторая компонента в разложении Вольда отсутствует, т.е. $Z_t \equiv 0$, то процесс называется **чисто линейно недетерминированным** (*purely linearly indeterministic*).

Таким образом, если стационарный случайный процесс с нулевым математическим ожиданием является чисто линейно недетерминированным, то он представим в виде процесса скользящего среднего бесконечного порядка

$$X_t = \sum_{j=1}^{\infty} c_j\epsilon_{t-j}, \qquad \textit{где} \qquad c_0 = 1, \sum_{j=0}^{\infty} c_j^2 < \infty. \tag{39}$$

В качестве тривиального примера линейно детерминированного стационарного процесса с нулевым средним можно указать на модель случайного уровня:

$$X_t = X_0, \qquad t = 1, 2, ..., \tag{40}$$

где X_0 - случайная величина, имеющее нулевое математическое ожидание и конечную дисперсию.

В практических исследованиях обычно сразу предполагают отсутствие линейно детерминированной компоненты.

Подбор стационарной модели ARMA для ряда наблюдений

$ARMA(p,q): x_t = \sum_{j=1}^{p} \alpha_j x_{t-j} + \sum_{j=0}^{q} \beta_j \varepsilon_{t-j}, \quad \beta_0 = 1, \quad \alpha_p \neq 0, \quad \beta_q \neq 0, \quad \varepsilon_t$ - белый шум.

Задача: При предположении, что некоторый временной ряд $x_1, x_2, \ldots, x_T$ порождается моделью ARMA, необходимо подобрать конкурентную модель из этого класса моделей.

Решение этой задачи предусматривает три этапа:

1.**идентификация** модели;

2.**оценивание** модели;

3.**диагностика** модели.

На этапе идентификации производится выбор некоторой частной модели из всего класса ARMA, т.е. выбор значений p и q. Используемые при этом процедуры являются не вполне точными, что может при последующем анализе привести к выводу о непригодноси идентифицированной модели и необходимости замены ее альтернативной моделью. На этом же этапе делаются предварительные грубые оценки коэффициентов $\alpha_1, \alpha_2, \ldots, \alpha_p, \beta_1, \beta_2, \ldots, \beta_q$ идентифицированной модели.

На втором этапе производится вычисление оценок коэффициентов модели с использованием эффективных статистических методов. Для оцененных коэффициентов вычисляются приближенные стандартные ошибки, дающие возможность, при дополнительных предположениях о распределениях случайных величин $X_1, X_2, \ldots$, строить доверительные интервалы для коэффициентов и проверять гипотезы об их истинных значениях с целью уточнения спецификации модели.

На третьем этапе применяются различные диагностические процедуры про-верки адекватности выбранной модели имеющимся данным. Неадекватности, обнаруженные в процессе такой проверки, могу указать на необходимую корректировку модели, после чего производится новый цикл подбора, и т.д. до тех пор, пока не будет получена удовлетворительная модель.

В случае, если мы имеем дело с ситуацией, когда уже имеется достаточно отработанная и разумно интерпретируемая модель эволюции того или иного показателя, можно обойтись и без идентификации.

Идентификация стационарной модели ARMA

Основной отправной точкой для идентификации стационарной модели ARMA является различие поведения **автокореляционных** и **частных автокореляционных** функций временных рядов, соответствующих различным моделям ARMA. Ранее было рассмотрено поведение автокорреляционных функций для различных моделей ARMA. В частности, было показано, что:

1. автокорреляционные функции ρ_k процесса $AR(p)$, начиная с некоторого $k \geq p$, являются экспоненциально затухающими;
2. автокорреляционные функции ρ_k процесса $MA(q)$, начиная с некоторого $k \geq q$, обрываются ($\rho_k = 0$).

Ясно, что порядок авторегресионной модели имеет большее значение, однако по поведению только пвтокорреляционной фукнции трудно идентифицировать даже порядок чистого (без MA составляющей) процесса авторегресии. Решению этого вопроса помогает рассмотрение поведения частной автокорреляционной функции.

В определение автокорреляционной функции

$$\rho_\tau = \frac{1}{Var(X_t)} E\{(X_t - \mu)(X_{t-\tau} - \mu)\}$$

входит ковариация между значениями процесса, отстоящими на τ шагов о времени друг от друга. Однако, на поведение процесса $AR(p)$ статистически влияет не только его значеие в момент, отстоящий на τ единиц назад, но и все промежуточные значения процесса между моментами t и $t - \tau$.

В этом случае вознкикает вопрос: какова линейная статистическая зависимость между значениями процесса в эти моменты, если мы устраним влияние вех промежуточных значений?

Коэффициент корреляции при исключении промежуточных значений называется **частным коэффициентом корреляции**.

Обозначим через φ_{kk} - k-ое значение частной автокорреляционной функции. По определению φ_{kk} - это коэффициент корреляции между x_{t-k} и x_t за вычетом той части x_t, которая линейно объяснена промежуточными шагами $x_{t-1}, x_{t-2}, \ldots, x_{t-k+1}$.

Более строго, нас интересует коэффициент корреляции

$$Corr(x_t - \varphi_{k1}x_{t-1} - \varphi_{k2}x_{t-2} - \ldots - \varphi_{kk-1}x_{t-k+1}),$$

пропущено (...,x_{t-k}) в корреляции
написано только значение первого элемента

где $\varphi_{k1}, \varphi_{k2}, \ldots, \varphi_{kk-1}$ - коэффициенты линейной комбинации, обеспечивающей минимальную среднеквадратичную ошибку предсказания

$$\min E\{(x_t - \varphi_{k1}x_{t-1} - \varphi_{k2}x_{t-2} - \ldots - \varphi_{kk-1}x_{t-k+1})^2\}.$$

Если исследуемый процесс принадлежит к виду $AR(p)$, то $\varphi_{pp} = \alpha_p$. Более того, для такого процесса текущее значение выражается через ровно p предыдущих значений, и учет более ранних значений уже не может улучшить прогноз.

Это означает, что для процесса $AR(p) \quad \varphi_{kk} = 0 \quad$ при $\quad k > p$.

Рассмотрим несколько простых примеров расчета значений частной автокорреляционной функции.

Процесс $AR(1)$. $X_t = \alpha X_{t-1} + \varepsilon_t$.

1. Рассмотрим регрессию: $X_t = \varphi_{11} X_{t-1} + \varepsilon_t$.

Умножаем уравнение на X_{t-1}, берем математическо ожидание от обеих частей и делим результат на ~~автокорреляционную~~ функцию $\gamma(0) = Cov(X_t, X_t)$.

автоковариационную

Получаем:

$$\gamma(0) = Cov(X_t, X_t) = E\big[(X_t - E(X_t))(X_t - E(X_t))\big] = E(X_t, X_t), \text{так как} \quad E(X_t) = 0$$

$$\frac{E(X_t, X_{t-1})}{\gamma(0)} = \varphi_{11} \underbrace{\frac{E(X_{t-1}^2)}{\gamma(0)}}_{=1} + \frac{E(X_{t-1}, \varepsilon_t)}{\gamma(0)} \quad \Rightarrow \quad \rho_1 = \varphi_{11} = \alpha$$

$P(k) = \alpha_k$

2. Чтобы получить φ_{22}, построим теоретическую регрессию

$X_t = \varphi_{21} X_{t-1} + \varphi_{22} X_{t-2} + \varepsilon_t$. Как и раньше, умножаем обе части на X_{t-1}, берем математическое ожидание, делим на $\gamma(0)$.

Получаем:

$$\frac{E(X_t, X_{t-1})}{\gamma(0)} = \varphi_{21} \underbrace{\frac{E(X_{t-1}^2)}{\gamma(0)}}_{=1} + \varphi_{22} \frac{E(X_t, X_{t-1})}{\gamma(0)} + \frac{E(X_{t-1}, \varepsilon_t)}{\gamma(0)} \quad \Rightarrow \quad \rho_1 = \varphi_{21} + \varphi 22 \rho_1.$$

$$\frac{E(X_t, X_{t-2})}{\gamma(0)} = \varphi_{21} \frac{E(X_{t-1} X_{t-2})}{\gamma(0)} + \varphi_{22} \frac{E(X_{t-1}^2)}{\gamma(0)} + \frac{E(X_{t-2}, \varepsilon_t)}{\gamma(0)} \quad \Rightarrow \quad \rho_2 = \varphi_{21}\rho_1 + \varphi 22.$$

В эти соотношение входят 2 неизвестных: φ_{21} и φ_{22}, а коэффициенты этой системы уравнений выражены через ρ_1 и ρ_2, связь которых с коэффициентами исходного уравнения процесса нам уже известна. Другими словами, наш подход дает связь между коэффициентами автокорреляционной функции и частной автокорреляционной функции. В данном случае у нас всего 2 уравнения, можно решать их по-разному. Но поскольку нас интересует только один единственный коэффициент φ_{22}, наиболее экономно будет использовать правило Крамера.

$$\begin{cases} \varphi_{21} + \varphi_{22}\rho_1 &= \rho_1 \\ \varphi_{21}\rho_1 + \varphi_{22} &= \rho_2 \end{cases}$$

$$\begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix} \times \begin{bmatrix} \varphi_{21} \\ \varphi_{22} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} \quad \Rightarrow \quad \varphi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}}{1 - \rho_1^2}. \quad \left(\frac{\Delta_2}{\Delta}\right)$$

Кстати, поскольку ρ_1 - это значение автокорреляционной функции, то оно меньше 1, то есть в знаменателе 0 не будет. Так как для процесса $AR(1)$ можно показать, что $\rho_k = \alpha^k \quad \forall k$, то

$$\varphi_{22} = \frac{\alpha^2 - \alpha^2}{1 - \rho^2} = 0 \quad \text{, как и ожидалось.}$$

Для того чтобы посчитать φ_{33}, запишем теоретическую регрессию $X_t = \varphi_{31}X_{t-1} + \varphi_{32}X_{t-2} + \varphi_{33}X_{t-3} + \varepsilon_t$. Аналогично предыдущему получаем систему уравнений:

$$\begin{cases} \rho_1 &= \varphi_{31} \cdot 1 + \varphi_{32}\rho_1 + \varphi_{33}\rho_2 \\ \rho_2 &= \varphi_{31}\rho_1 + \varphi_{32} \cdot 1 + \varphi_{33}\rho_1 \\ \rho_3 &= \varphi_{31}\rho_2 + \varphi_{32}\rho_1 + \varphi_{33} \cdot 1 \end{cases}$$

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{bmatrix} \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \end{bmatrix}$$

Вновь нас инетересует только коэффициент φ_{33}. Восопользуемся правилом Крамера, но, используя предыдущий опыт, считаем только определитель в числителе:

$$\Delta_3 = \begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix} = \begin{vmatrix} 1 & \alpha & \alpha \\ \alpha & 1 & \alpha^2 \\ \alpha^2 & \alpha & \alpha^3 \end{vmatrix} = 0$$

(последний столбец пропорционален первому стобцу, причем коэффициент пропорциональности равен коэффициенту уравнения процесса) $\quad \Rightarrow \quad \varphi_{33} = 0.$

Теперь можно выразить определитель в числителе для произвольного k. Система линейных уравнений, связывающая значения автокорреляционной и частной автокорреляционной функций, фактически является системой уравнений Юла-Уокера для теоретической регрессии с неизвестными и подлежащими определению значениями частной автокорреляционной функции:

$$\begin{cases} \rho_1 &= \varphi_{k1} \cdot 1 + \varphi_{k2}\rho_1 + \varphi_{k3}\rho_2 + \ldots + \varphi_{kk}\rho_{k-1} \\ \rho_2 &= \varphi_{k1}\rho_1 + \varphi_{k2} \cdot 1 + \varphi_{k3}\rho_1 + \ldots + \varphi_{kk}\rho_{k-2} \\ \ldots & \\ \rho_k &= \varphi_{k1}\rho_{k-1} + \varphi_{k2}\rho_{k-2} + \varphi_{k3}\rho_{k-3} + \ldots + \varphi_{kk} \cdot 1 \end{cases}$$

Искомый определитель имеет вид:

$$\Delta_k = \begin{vmatrix} 1 & \rho_1 & \ldots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \ldots & \rho_1 & \rho_2 \\ \vdots & \ldots & \ldots & 1 & \rho_{k-1} \\ \rho_{k-1} & \rho_{k-2} & \ldots & 1 & \rho_k \end{vmatrix}.$$

$$\Delta = \begin{vmatrix} 1 & \rho_1 & \rho_2 & \ldots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \ldots & \rho_{k-2} \\ \vdots & \ldots & \ldots & \ldots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \ldots & 1 \end{vmatrix}.$$

На главной диагонали все элементы кроме последнего будут равны 1. Связь между коэффициентами φ_{11} и $\rho_1, \ldots, \rho_k$, то есть между автокорреляционной и частной автокорреляционной функциями, выражается следующими соотношениями:

$\varphi_{kk} = \dfrac{\Delta_k}{\Delta}$, $\quad$ где $\quad \Delta$ - определитель системы уравнений

$$\Delta = \begin{vmatrix} 1 & \rho_1 & \ldots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \ldots & \ldots & \rho_{k-2} \\ \vdots & \ldots & \ldots & \ldots & \vdots \\ \rho_{k-1} & \ldots & \ldots & \ldots & 1 \end{vmatrix}.$$

Таким образом, зная коэффициенты автокорреляционной функции, можно по уравнениям Юла-Уокера пеерсчитать коэффициенты частной автокорреляционной функции, и наоборот.

Если для $AR(1)$ считать четвертое, пятое и так далее значения частной автокорреляционной функции, то все он равны 0. Поэтому для процесса $AR(1)$ значения частной автокорреляционной функции, начиная со второго, равны 0 ($\varphi_{kk} = 0, \quad k > 1$). Следовательно, получен индикатор того, что исследуемый процесс точно является процессом $AR(1)$.

В общем случае процесса $AR(p)$: если $k > p$, то число столбцов в определителе Δ_k больше, чем порядок авторегрессии. В этом случае разностное уравнение $\rho_1, \ldots, \rho_k$ показывает, что начиная с $k > p$, каждое ρ выражается одной и той же линейной комбинацией предыдущих значений. Как только число столбцов больше p, то каждый столбец с номером большим k является линейной комбинацией предыдущих столбцов. Причем коэффициенты этой линейной комбинации - это коэффициенты уравнения процесса $\alpha_1, \alpha_2, \ldots, \alpha_p$. Таким образом, последний столбец всегда есть линейная комбинация p предыдущих столбцов, как только $k > p$. Поэтому определитель такой матрицы обязан обратиться в нуль.

Общий вывод: частная автокорреляционная функция авторегрессионного процесса $AR(p)$, равна 0 для $k > p$, и вообще говоря, не равна 0 при $k \leq p$.

В результате частная автокорреляционная функция для процесса AR играет точно такую же качественную роль, как автокорреляционная функция для процесса MA. Она обращается в нуль, как только $k > p$. Мы это доказали в общем случае и продемонстрировали на процессах первого и второго порядка.

Теперь делаем следующий шаг. Переходим к комбинации двух раличных типов процессов $AR(p)$ и $MA(q)$. Рассмотрим общий процесс $ARMA(p, q)$ авторегрессии-скользящего среднего:

$$x_t = \sum_{j=1}^{p} \alpha_j x_{t-j} + \sum_{j=0}^{q} \beta_j \varepsilon_{t-j}, \quad \beta_0 = 1, \text{ - в развернутом виде;}$$

$\alpha_p(L)x_t = \beta_q(L)\varepsilon_t$ - в операторном виде.

Чтобы выяснить, как ведет себя автокорреляционная и частная автокорреляционная функции процесса ARMA, начнем с простой ситуации, а именно с процесса $ARMA(1,1)$.

В операторной записи: $(1-\alpha L)x_t = (1+\beta L)\varepsilon_t$, или в обычном виде:

$$x_t = \alpha x_{t-1} + \varepsilon_t + \beta\varepsilon_{t-1}$$

Условие стационарности процесса ARMA: $|\alpha| < 1$.

Для вычисления дисперсии процесса удобно использовать $MA(\infty)$ представление, так называемое представление линейного фильтра:

$$x_t = \frac{1+\beta L}{1-\alpha L}\varepsilon_t = (1+\beta L)(1+\alpha L+\alpha^2 L^2 + alpha^3 L^3 + \ldots)\varepsilon_t =$$

$$= \left[1 + (\alpha+\beta)L + \alpha(\alpha+\beta)L^2 + \alpha^2(\alpha+\beta)L^2 + \ldots\right]\varepsilon_t.$$

$x_{t-1} = Lx_t$

$x_t = \alpha x_{t-1} + \beta\varepsilon_t + \beta\varepsilon_{t-1}$

$$\frac{\gamma(1)}{\gamma(0)} = \frac{\alpha\gamma(0)+\beta\sigma_\varepsilon^2}{\gamma(0)} = \alpha + \frac{\beta\sigma_\varepsilon^2}{\gamma(0)} = \alpha + \frac{\beta\sigma_\varepsilon^2(1-\alpha^2)}{(1+\beta^2+2\alpha\beta)\sigma_\varepsilon^2} =$$

$$= \frac{\alpha+\alpha\beta^2+2\alpha^2\beta+\beta-\alpha^2\beta}{1+\beta^2+2\alpha\beta} = \frac{\alpha+\beta+\alpha^2\beta+\beta+\alpha\beta^2}{1+\beta^2+2\alpha\beta} =$$

$$= \frac{(\alpha+\beta)+\alpha\beta(\alpha+\beta)}{1+\beta^2+2\alpha\beta} = \frac{(\alpha+\beta)+(1+\alpha\beta)}{1+\beta^2+2\alpha\beta}.$$

$$\text{Тогда } Var(x_t) = \left[1+(\alpha+\beta)^2\left[1+\alpha^2+\alpha^4+\ldots\right]\right]\sigma_\varepsilon^2 = \left[1+\frac{(\alpha+\beta)^2}{1-\alpha^2}\right]\sigma_\varepsilon^2 =$$

$$= \frac{1+\beta^2+2\alpha\beta}{1-\alpha^2}\sigma_\varepsilon^2 \quad \Rightarrow \quad \text{~~автокорреляционная~~ функция} \quad \gamma(0) = \frac{1+\beta^2+2\alpha\beta}{1-\alpha^2}\sigma_\varepsilon^2.$$

автоковариационную

Чтобы посчитать первую автокорреляцию, умножим выражение $x_t = \alpha x_{t-1} + \varepsilon_t + \beta\varepsilon_{t-1}$ на x_{t-1} и возьмем математическое ожидание от обеих частей.

Получим $\gamma(1) = \alpha\gamma(0) + \beta Cov(\varepsilon_{t-1}, x_{t-1})$.

Умножив выражение $x_{t-1} = \alpha x_{t-2} + \varepsilon_{t-1} + \beta\varepsilon_{t-2}$ на ε_{t-1} и взяв математическое ожидание, получаем $Cov(\varepsilon_{t-1}, x_{t-1}) = \sigma_\varepsilon^2$. После подстановки получаем $\gamma(1) = \alpha\gamma(0) + \beta\sigma_\varepsilon^2$.

Окончательно:

$$\rho_1 = \frac{\gamma(1)}{\gamma(0)} = \frac{(\alpha+\beta)(1+\alpha\beta)}{1+\beta^2+2\alpha\beta}.$$

Последующие значения автокорреляционной функции получить проще. Если умножить x_t на x_{t-2} и взять математическое ожидание, получим: $\gamma(2) = \alpha\gamma(1)$. Для всех значений γ с индексом большим, чем ~~ma~~ порядок MA-части, получаем, что из равенства $\gamma(k+1) = \alpha\gamma(k)$ следует равенство $\rho_{k+1} = \alpha\rho_k$. То есть мы получили то же самое соотношение, которое имели для "чистой"модели $AR(1)$. Мы назвали это уравнениями Юла-Уокера для автокорреляционной функции. Другими словами, мы установили, что, начиная со второй, автокорреляции $ARMA(1,1)$ ведут себя так же, как автокорреляции $AR(1)$, но первые автокорреляции этих процессов различаются.

Для $AR(1)$, автокорреляции имели вид $\rho_i = a_1^i$. Для процесса $ARMA(1,1)\rho_1 \neq \alpha_1, \rho_1 = \frac{(\alpha+\beta)(1+\alpha\beta)}{1+\beta^2+2\alpha\beta}, \rho_i = \alpha\rho_{i-1}, i = 2, \ldots$.

Однако, начиная со второго номера, значения автокорреляционной функции все равно убывают экспоненциально.

Для процесса $ARMA(p,q)$ справедливо аналогичное утверждение, если, конечно, процесс стационарен. Первые p значений автокорреляционной функции определяются через коэффициенты AR и MA-частей, а потом значения автокорреляционной функции выражаются в виде суммы экспоненциально затухающих слагаемых.

Для процесса $ARMA(p,q)$, применяя тот же метод умножения уравнения процесса на значения x_{t-i} и последующего взятия математического ожидания, получим, что для $k > \max(p, q+1)$, автокорреляции ρ_k определяются разностным уравнением, соответствующим AR-части. А все предыдущие значения ρ_k могут быть выражены через коэффициенты $\alpha_1, \ldots, \alpha_p$, $\beta_1, \ldots, \beta_q$ как решения системы линейных уравнений, полученной способом аналогичным процессу $ARMA(1,1)$. Следовательно, начиная с номера $\max(p, q+1)$, автокорреляционная функция "затухает"том же смысле, что и для процесса AR. Таким же качественным поведением характеризуется и частная автокорреляционная функция процесса $ARMA(p,q)$. Интуитивно такое свойство понятно. Мы говорили, что для $MA(q)$ автокорреляционная функция для номеров, больших q, просто равна нулю. Поэтому влияние MA-части при $k > q$ как бы прекращается, и дальше "работает"только $AR(p)$. Наоборот, частная автокорреляционная функция для процесса $AR(p)$ для k, больших чем p, равна нулю. То есть $AR(p)$ перестает влиять на частную автокорреляционную функцию, и остается только влияние $MA(q)$.

Вывод: Если процесс относится к типу $ARMA(p,q)$, то, начиная с некоторого номера (причем этот номер важен, он нам говорит о величине p и q), и автокорреляционная, и частная автокорреляционная функции ведут себя как сумма затухающих экспонент, если ряд стационарен.

Пример 1.

Рассмотрим нестационарный ряд случайного блуждания. Его уравнение имеет вид: $x_t = x_{t-1} + \varepsilon_t$. Если взять первую разность, равную $\Delta x_t = x_t - x_{t-1} = (1-L)x_t$, то уравнение сведется к следующему: $\Delta x_t = y_t = \varepsilon_t$. То есть в первых разностях ряд станет стационарным.

Такой подход приводит к стационарности не только случайное блуждание.

Пример 2.

Рассмотрим ряд вида: $x_t = \alpha + \beta t + \gamma t^2 + \varepsilon_t$.

Его плотностью детерминированная часть (тренд) является параболической функцией времени. Очевидно, что этот ряд нестационарный. Математическое ожидание этого процесса зависит от времени.

Проверим: Приводится ли такой ряд взятием последовательных разностей к стационарному?

Если мы возьмем первую последовательную разность, то получим: $\Delta x_t = \alpha + \beta t + \gamma t^2 + \varepsilon_t - \alpha - \beta(t-1) - \gamma(t-1)^2\varepsilon_{t-1} = \beta + 2\gamma t - \gamma + (\varepsilon_1 - \varepsilon_{t-1}$.

Степень полинома, описывающего тренд, понизилась на единицу. Если провести взятие второй разности, то останется $\Delta^2 x_t = \Delta x_t - \Delta x_{t-1} = 2\gamma + (\varepsilon_t - 2\varepsilon_{t-1} + \varepsilon_{t-2})$, то есть получим стационарный процесс.

Правда, видно, как в уравнение начинает "проникать"скользящее среднее. Полученный двукратным взятием разностей стационарный процесс является процессом $MA(2)$. Но, по крайне мере, взятием последовательных разностей исходный ряд с квадратичным трендом приводится к стационарному виду.

Бокс и Дженкинс на основании этого свойства предложили выделить класс нестационарных рядов, которые взятием последовательных разностей можно привести к стационарному виду, а именно к виду $ARMA$.

Если ряд после взятия d последовательных разностей приводится к стационарному, то этот ряд носит название $ARIMA(p,d,q)$.

$ARIMA$ - процесс авторегрессии - интегрированного скользящего среднего. При этом p - параметр AR-части, d - степень интеграции, и q - это параметр MA-части.

В операторном виде процесс $ARIMA(p,d,q)$ записываетя как: $\alpha_p(L)\Delta^d x_t = \beta_q(L)\varepsilon_t$.

Или по-другому $\underbrace{\alpha_p(L)(1-L)^d}_{\text{p+d}} x_t = \beta_q(L)\varepsilon_t$.

Этот процесс нестационарный, потому что здесь не выполняется условие, что все корни характеристического уравнения по модулю меньше единицы.

Но, если обозначить $(1-L)^d x_t = y_t$, то y_t - это стационарный процесс.

Основная заслуга Бокса и Дженкинса в том, что они сделали этот подход лет $25-30$ назад весьма популярным и ввели в практику программы для применения этого подхода в компьютерный пакет научных программ для системы $IBM-360$, распространенной в $1970-1980$гг. по всему миру.

Подход Бокса - Дженкинса

Бокс и Дженкинс предложили следующий подход, состоящий из нескольких этапов, к выбору модели типа $ARIMA$ по наблюдаемой реализации временного ряда.

1 этап "Идентификация модели"

1. Установить порядок интеграции d, то есть добиться стационарности ряда, взяв достаточное количество последовательных разностей. Другими словами, "остационарить"ряд. Используются все описанные ранее средства визульный анализ коррелограммы визуальный анализ автокорреляционной и частной автокорреляционной функций, тест на единичный корень. $d \le 2$ - обычно.

2. После этого мы получаем временной ряд y_t, которому нужно подобрать уже $ARMA(p,q)$. Исходя из поведения автокорреляционной и частной автокорреляционной функций, установить параметры p и q. Рекомендуется, чтобы $\gamma + \beta \le 3$. *p + q <=3, если нет сезонной компоненты*

2 этап "Оценивание модели"

Оценивание коэффициентов $\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q$ при условии, что мы уже знаем p и q.

3 Этап "Диагностика модели"

Стандартная для эконометрического подхода процедура. По остаткам осуществляется тестирование или диагностика построенной модели.

4 Этап "Использование модели"

Использование модели, в основном, для прогнозирования будущих значений временного ряда.

Бокс и Дженкинс применили этот подход ко многим временным рядам, которые были известны в то время, как к финансовым, так и к макроэкономическим.

Бокс и Дженкинс смогли построить модели типа $ARIMA$ для всех исследуемых рядов и установили, в частности, что практически все экономические процессы описываются моделями с относительно небольшими величинами параметров p и q, а параметр d обычно не превышает 2. К тому же оказалось, что точность прогнозирования по моделям $ARIMA$ оказалась выше, чем давали в то время экономические модели.

Если исследуемый ряд нестационарный, то его автокорреляционная функция не будет убывать.

Если ряд стационарен, то, начиная с какого-то номера, теоретические автокорреляции будут убывать. Поэтому можно рассчитать их оценки - выборочные автокорреляции, и посмотреть, убывают они или нет. Если ряд окажется стационарным, перейти к определению параметров p и q. Если нет, то надо построить ряд первых разностей и проверить на стационарность его.

Рассмотрим, например, процесс $x_t = \alpha x_{t-1} + \varepsilon_t$, $\varepsilon_t \sim WN(0,\sigma^2)$.

При $\alpha > 1$ взятие первой разности не поможет сделать ряд стационарным. Это нестационарный ряд взрывного типа, оценки его "автокорреляционной"функции растут с увеличением сдвига во времени.

Если же $\alpha = 1$, то ряд представляет собой случайное блуждание и после взятия первой разности он станет стационарным.

Переходим к оцениванию по выборке статистических характеристик исследуемого процесса в предположении его эргодичности. В качестве оценки математического ожидания применяется статистика $\bar{x} = \frac{1}{T}\sum_{t=1}^{T} x_t$, то есть обычное среднее по выборке.

В качестве оценки дисперсии процесса обычно принимается следующая величина: $S^2 = \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})^2$. Обратите внимание, делитель не $(T-1)$, как привычно для обработки независимых наблюдений, а T.

Для оценки коэффициента теоретической автокорреляции используем $\hat{\rho}_k = \frac{\sum_{t=k+1}^{T}(x_t-\bar{x})(x_{t-k}-\bar{x})}{TS^2}$.

Это означает, что для расчета выборочных ковариаций используется одинаковый делитель - T, что, в случае независимых наблюдений, дает смещенные оценки соответствующих теоретических ковариаций. Если бы мы требовали несмещенности оценок, то у нас бы были разные делители при различных k. Кроме того, при увеличении k уменьшается объем выборки, пригодный для расчета оценки соответствующего коэффициента корреляции. На практике, по рекомендации, идущей еще от Бокса и Дженкинса, не рассчитываются оценки $\hat{\rho}_k$ для $k > \frac{T}{4}$.

Одним из следствий выбора именно таких оценок значений автокорреляционной функции является гарантированная положительная определенность выборочной корреляционной матрицы. Ранее мы отмечали наличие такого свойства у теоретической автокорреляционной функции. При одинаковом делителе у оценок автокорреляционной функции при различных k положительная определенность выборочной корреляционной матрицы гарантирована. Поэтому мы предпочитаем пусть смещенную оценку, но гарантирующую это свойство, тем более, что у нас обычно T - относительно большое. Во-вторых, при дополнительном предположении о нормальности распределения значений процесса именно эта оценка совпадает с оценкой метода максимального правдоподобия. Оценка $\hat{\theta}$ параметра θ несмещенная, если $E(\hat{\theta}) = \theta$.

I этап "Идентификация модели"

Некоторые выводы, полученные ранее:

1. Если при k большем некоторого q выборочные автокорреляции $\hat{\rho_k}$ становятся близкими к нулю, то подходящей моделью может быть $MA(q)$.

2. Если при k большем некоторого p частные выборочные автокорреляции $\hat{\phi_{kk}}$ становятся близкими к нулю, то подходящей моделью может быть $AR(p)$.

3. Если и $\hat{\rho_k}$ и $\hat{\phi_{kk}}$ стремятся к нулю плавно, затухают по экспоненте, то подходящей моделью является смешанная модель $ARMA(p,q)$.

Важным является также принцип экономичности: предпочтительны модели с небольшими значениями p и q.

Для уточнения того, что понимать под словами "ϕ_{kk} близки к нулю"или "ϕ_{kk} близки к нулю"могут быть использованы следующие результаты.

Бокс и Дженкинс (1976) показали следующее.

Если x_t - это процесс $MA(q)$ и процесс белого шума ε_t распределен по нормальному закону, то для $k > q$ и при больших T выборочная корреляция r_k асимптотически распределена по нормальному закону со следующими параметрами:

$$\bar{r_k} = E(r_k) = 0,$$
$$S_r^2 = Var(r_k) \approx \frac{1}{T} \text{ при } k = 1,$$
$$Var(r_k) \approx \frac{1}{T}(1 + 2\sum_{j=1}^{k-1} r_j^k) \text{ при } k > 1.$$

(каждой реализации случайного процесса при любом k соответствует свое значение ρ_k).

Этот результат также используется для оценки дисперсии путем замены ρ_k на ϕ_{kk}.

Гипотеза $H_0 : \rho_k = 0$ для $k > q$ означает, что процесс x_t - есть процесс $MA(q)$.

Если x_t - это процесс $AR(p)$, то аналогичное утверждение справедливо для частных выборочных автокорреляций ϕ_{kk}, то есть $H_0 : \phi_{kk} = 0$ при $k > p$ ϕ_{kk} асимптотически распределена по нормальному закону со следующими параметрами:

$$E(\phi_{kk}) = 0,$$
$$Var(\phi_{kk}) \approx \frac{1}{T},$$
$$(\text{при } k > p \ \sqrt{T}\phi_{kk} \sim N(0,1)).$$

II этап "Оценивание коэффициентов моделей типа ARMA"

Рассмотрим модель $AR(p) : \alpha_1 x_{t-1} + ... + \alpha_p x_{t-p} + \varepsilon_t$. Для нее можно применить обычный метод наименьших квадратов (МНК).

Поскольку регрессоры относятся к предыдущим моментам времени ($t - j \neq t$), а ε_t - белый шум, то корреляция регрессоров x_{t-j} со случайным возмущением ε_j отсутствует.

Метод наименьших квадратов (МНК) дает 1) смещенные оценки $\mathsf{E}\widehat{\theta} \neq \theta$; 2) состоятельные оценки $\lim\limits_{n\to\infty} \widehat{\theta} \to \theta$, несмотря на то, что присутствует стохастический регрессор.

Нужно проверить по остаткам, действительно ли наши предположения о том, что ε_t - белый шум, выполнены.

Если дополнительно белый шум является гауссовым ($\varepsilon_t \mathcal{N}(0, \sigma_\varepsilon^2)$), то значения x_t распределены нормально, а оценки коэффициентов $\widehat{\alpha}_1, ..., \widehat{\alpha}_p$ произведенные МНК, состоятельны и асимптотически нормальны.

Если данные имеют ненулевое выборочное среднее $\bar{x}$, то можно либо вычесть это среднее из данных и строить регрессию без свободного члена, т.е. рассматривать $x_t - \bar{x}$, либо просто строить регрессию со свободным членом θ.

Другими словами, для оценки математического ожидания процесса $AR(p)$ можно использовать две статистики: уже упомянутую в прошлой лекции $\bar{x} = \frac{1}{T}\sum_{t=1}^{T} x_t$ и $\widehat{\mu} = \frac{\widehat{\theta}}{1 - \widehat{\alpha}_1 - ... - \widehat{\alpha}_p}$, в которой использованы оценки МНК. Для гауссова процесса ε_t обе оценки состоятельны и асимптотически нормальны. Более того, обе они асимптотически независимы от оценок параметров модели, полученных МНК.

Для моделей скользящего среднего $MA(q)\Big(x_t = \sum\limits_{j=1}^{q} \beta_j \varepsilon_{t-j}\Big)$ невозможно аналитически выразить остаточную сумму квадратов через значения реализации x_t и параметры модели, а следовательно, применить МНК. Для оценки параметров существуют два подхода, по-разному реализованные сегодня в специализированных компьютерных программах.

Применение методов максимального правдоподобия (ММП)

В предположении нормальности ошибки ε_t выражаем ковариационную матрицу ошибок (ее элементы - значения автокорреляционной функции) через параметры $\beta_1, \beta_2, ..., \beta_q$ обратимой модели $MA(q)$. Функция правдоподобия для нормального распределенного вектора $\vec{x} = [x_1, x_2, ..., x_T]'$ имеет вид

$$L(\vec{\beta}, \sigma_\varepsilon^2 \sum_x |\vec{x}) = \frac{1}{(\sqrt{2\pi\sigma_\varepsilon^2})^T}\Big(\sqrt{det \sum_x}\Big)^{-1} exp\Big(-\frac{(\vec{x} - \bar{x})\sum_x^{-1}(\vec{x} - \bar{x})}{2\sigma_\varepsilon^2}\Big)$$

$$\varepsilon\mathcal{N}(0, \sigma_\varepsilon^2 \sum_x)$$

Здесь через $\sum_x$ обозначена ковариационная матрица процесса $\vec{x}$. Элементы этой матрицы выражаются, согласно свойству обратимости, через параметры модели $\beta_1, \beta_2, ..., \beta_q$, поэтому процедура численной оптимизации (максимизации функции правдоподобия) позволяет найти оценки метода максимального правдоподобия, которые будут обладать обычными свойствами состоятельности и асимптотической нормальности. Кроме того, оценки параметров модели $MA(q)$ и оценка дисперсии случайного возмущения асимптотически независимы.

Применение процедуры нелинейной оптимизации поиска на сетке (grid-search procedure).

Этот подход, позволяющий облегчить вычислительные затраты, применили Бокс и Дженкинс.

В общем случае модели $MA(q)$ 1) Фиксируем начальные значения $\beta_1^0, ..., \beta_q^0$ и определяем остатки через наблюдения: $e_1 = x_1 - \bar{x}, e_2 = x_2 - \bar{x} - \beta_1^0 e_1$ И так далее. Процедура получения остатков достаточно проста, все остатки с отрицательными и нулевым индексами в используемых выражениях заменяем нулями. 2) Затем находим $\min\limits_{\vec{\beta}} \sum e_i^2$. 3) Численным поиском находим оценки параметров.

Рассмотрим идею этого подхода на примере процесса $MA(2) : x_t = \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2}$.

1. Найдем сначала оценку математического ожидания процесса $\bar{x} = \mathsf{E}x_t$ в предположении о его стационарности и эргодичности.

2. Затем выберем некоторые значения (β_1, β_2), например $(0,5; 0,5)$.

3. Исходя из уравнения процесса $(x_k = \bar{x} + \beta_1 e_{k-1} + \beta_2 e_{k-2} + e_k)$, x_1 выражается через предыдущие ошибки, но они не известны. Поэтому будем считать, что

$$x_1 = \bar{x} + e_1,$$

$$x_2 = \bar{x} + \beta_1 e_1 + e_2,$$

$$x_3 = \bar{x} + \beta_1 e_2 + \beta_2 e_1 + e_3,$$

$$x_4 = \bar{x} + \beta_1 e_3 + \beta_2 e_2 + e_4$$

Это дает нам $e_1 = x_1 - \bar{x}$. Затем полагаем

$$e_2 = x_2 - \bar{x} - \beta_1 e_1,$$

$$e_3 = x_3 - \bar{x} - \beta_1 e_2 - \beta_2 e_1$$

$$e_4 = x_4 - \bar{x} - \beta_1 e_3 - \beta_2 e_2$$

и так далее.

4. Теперь можно вычислить сумму квадратов отклонений $\sum e^2 = S(\beta_1, \beta_2)$. Поскольку она посчитана для выбранных значений параметров (β_1, β_2), мы можем рассматривать ее как функцию этих переменных.

5. Далее какой-либо численной процедурой, например поиском на сетке, можно перебирать комбинации β_1, β_2 или численно искать минимум функции $\min\limits_{\beta_1, \beta_2} \sum e_t^2$.

Коэффициенты (β_1^*, β_2^*), которые обеспечивают минимум выражения $\sum e_t^2$ и будут оценками коэффициентов модели $MA(2)$.

Замечание: Была доказана асимптотическая эквивалентность этой процедуры оцениванию методом максимального правдоподобия (при $T \to \infty$).

Для оценки параметров модели $ARMA(p,q)$ может применяться комбинация метода наименьших квадратов с поиском на сетке.

Рассмотрим модель $ARMA(2,2)$.

Пусть уравнение модели имеет вид

$$(1 - \alpha_1 L - \alpha_2 L^2) x_t = (1 + \beta_1 L + \beta_2 L^2) \varepsilon_t$$

Его можно переписать в виде

$$x_t = \frac{(1 + \beta_1 L + \beta_2 L^2) \varepsilon_t}{1 - \alpha_1 L - \alpha_2 L^2}$$

Введем вспомогательный случайный процесс $x_t = \dfrac{\varepsilon_t}{1 - \alpha_1 L - \alpha_2 L^2}$. Тогда процессы x_t и z_t связаны соотношением $x_t = (1 + \beta_1 L + \beta_2 L^2) z_t$, которое напоминает уравнение $MA(2)$, только вместо ε_t стоит z_t.

Определим "наблюдения"z_t через наблюдения x_t, сконструировав z_t так же, как ранее остатки модели MA, т.е. значения z_t, которые еще не определены (с нулевыми и отрицательными индексами), полагаем равными нулю. Тогда:

$$z_1 = x_1 \quad (x_1 = z_1 + \beta_1 z_0 (= 0) + \beta_2 z_{-1} (= 0))$$

$$z_2 = x_2 - \beta_1 z_1 \quad (x_2 = z_2 + \beta_1 z_1 + \beta_2 z_0 (= 0))$$

$$z_3 = x_3 - \beta_1 z_2 - \beta_2 z_1$$

$$\dots$$

Далее из определения процесса z_t следует, что значения z_t связан с остатками e_t исходной модели следующими соотношениями: $z_t - \alpha_1 z_{t-1} - \alpha_2 z_{t-2} = e_t$. Относительно процесса z_t модель стала $AR(2)$.

Зная "реализацию"z_t для выбранных значений коэффициентов (β_1^0, β_2^0), можно оценить коэффициенты α_1, α_2 с помощью МНК.

В результате находим оценки коэффициентов $(\widetilde{\alpha}_1, \widetilde{\alpha}_2, \widetilde{\beta}_1, \widetilde{\beta}_2)$, но получены они по-разному. Оценки $(\widetilde{\beta}_1, \widetilde{\beta}_2)$ заданы как начальные значения, т.е. $(\widetilde{\beta}_1, \widetilde{\beta}_2) = (\beta_1^0, \beta_2^0)$. А потом, исходя из них, построены оптимальные оценки $(\widetilde{\alpha}_1, \widetilde{\alpha}_2)$.

Применяя численные методы оптимизации (например, поиск на сетке) оцениваем значения параметров, обеспечивающие $\min\limits_{\alpha_1, \alpha_1, \beta_1, \beta_2} \sum e_i^2$.

Если ε_t - белый гауссовский шум, то **для оценивания модели** $ARMA(p,q)$ можно такде применять **метод максимального правдоподобия**.

Общая схема его применения следующая.

1. Выражаем значения автокорреляционной функции процесса через параметры модели $\alpha_1, \alpha_2, ..., \alpha_p, \beta_1, \beta_2, ..., \beta_q$

2. Строим ковариационную матрицу порядка

3. Записываем функцию правдоподобия для имеющейся выборки $x_1, ..., x_T$

4. Решаем (как правило, численно) систему уравнений правдоподобия относительно оценок коэффициентов $\widehat{\alpha}_1, \widehat{\alpha}_2, ..., \widehat{\alpha}_p, \widehat{\beta}_1, \widehat{\beta}_2, ..., \widehat{\beta}_q$

Реальные алгоритмы метода максимального правдоподобия, реализованные в специализированных компьютерных пакетах, отличаются от этой схемы, но принципиально она осуществима и отражает логику получения оценок ММП.

После того, как параметры модели оценены, остается оценить качество полученной модели.

III Этап "Диагностика модели типа ARMA"

Исходной информацией для диагностики служат остатки модели. Поскольку предполагалось, что случайное возмущение ε_t является белым шумом, поэтому, прежде всего, надо проверять некоррелированность остатков.

Таким образом, проверке подлежат, в первую очередь, 2 момента.

Первый момент - это **качество модели**. Для проверки качества модели нужен индикатор типа критерия множественной детерминации R^2.

Второй момент - это **некоррелированность остатков**. Если остатки коррелированы, то оценивать $AR(p)$-часть методом наименьших квадратов нельзя, так как получим несостоятельные оценки.

Проверка качества модели

Как правило, при построении моделей временных рядов критерии качества подгонки моделей применяются для сравнения моделей между собой. Поскольку оценки коэффициентов проводятся путем оптимизации, фактически речь идет о выборе порядка модели, т.е. о сравнении моделей с различным числом параметров.

Абсолютные критерии, типа стандартного коэффициента множественной детерминации R^2, не применяются.

Очевидно, что модели с большими p и q дают лучшие (меньшие) значения суммы квадратов остатков, чем модели с меньшими p и q. Но это противоречит принципу экономности.

Наиболее распространенными в настоящее время является предложенный Акаике в 1974 г. **критерий AIC** (Akaike information criterion) и предложенный Шварцем в 1978 г. **критерий BIC** (Bayesian information criterion). Оба эти критерия построены примерно одинаковым способом.

Информационный критерий AIC для модели $ARMA(p,q)$ выглядит следующим образом:

$$AIC(p,q) = \ln \widehat{\sigma}^2 + 2\frac{p+q}{T}$$

где T - число наблюдений. **Байесовский информационный критерий BIC** имеет несколько другой вид:

$$BIC(p,q) = \ln \widehat{\sigma}^2 + \frac{p+q}{T}\ln T$$

Здесь $\widehat{\sigma}^2 = \dfrac{\sum e_t^2}{T-p-q}$ (другой вариант $\widehat{\sigma}^2 = \frac{\sum e_t^2}{T}$).

Структура этих критериев следующая: логарифмы остаточной суммы квадратов плюс штраф за уменьшение числа степенней свободы.

Нужно так подобрать значения параметров p и q, чтобы получить минимальное значение каждого из критериев: $\min\limits_{p,q} AIC(p,q), \min\limits_{p,q} BIC(p,q)$, в то время как по критерию R^2 мы отдавали предпочтение модели с большим его значением.

Замечание: В отличие от коэффициентов детерминации ничего нельзя сказать ни о диапазоне изменения, ни о знаке критериев Акакике и Шварца.

Критерий Акакике базируется на обобщении ММП. Приведенное выражение подразумевает, что случайное возмущение является гауссовым. Шибата показал, что для процессов авторегрессии критерий AIC переоценивает порядок модели, и оценка порядка модели на основании этого критерия несостоятельна.

Структура AIC и BIC похожа на R^2_{adj} для пространственных данных

\sum_(eps_t^2) = RSS - объясненная часть дисперсии, но было зачеркнуто Белянкиной

Критерий Шварца BIC основан на байесовском подходе и имеет более фундаментальное теоретическое обоснование. Показано, что оценка порядка модели по этому критерию является состоятельной.

Тем не менее, на практике чаще используется информационный критерий AIC. Существует масса работ, в которых сравнивается применение этих критериев по отношению к разным моделям, но окончательного вывода пока не сделано. На практике разные критерии могут привести к выбору различных моделей. Есть некоторый накопленный опыт по поводу того, к каким типам моделей приводит один критерий и к каким приводит другой. Если вы публикуете какие-то исследование, считается вполне нормальным просто указать, какой критерий вы применяете без специального обоснования вашего выбора.

Проверка автокорреляции остатков

В этом вопросе существует расхождение между тем, как следовало бы поступать с теоретической точки зрения и тем, что делается практически.

В 1970 г. Бокс и Пирс предложили **статистический критерий для проверки автокорреляции временного ряда**, который сегодня принято называть **Q-статистикой**.

Тест Бокса-Пирса проверяет гипотезу о совместном равенстве нулю всех автокорреляций временного ряда до порядка m включительно, т.е. гипотезу $H_0 : \rho_1 = \rho_2 = ... = \rho_m = 0$ против альтернативной гипотезы $H_1 : \sum\limits_{i=1}^{m} \rho_i^2 > 0$.

Бокс и Пирс показали, что при увеличении длины выборки, т.е. при больших T, статистика $Q(m) = T\sum\limits_{k=1}^{m} r_k^2$ (r_k - выборочные автокорреляции) имеет асимптотическое распределение χ^2 с m степенями свободы.

Этот тест было предложено применять для проверки наличия автокорреляции остатков типа $ARMA(p,q)$. В этом случае число степеней свободы уменьшается на $(p+q)$, т.е. статистика $Q(m)$ имеет асимптотическое распределение χ^2 с $m-p-q$ степенями свободы, а если при этом ε_t является процессом нормального белого шума, то $Q(m)$ имеет распределение χ^2 с $m-p-q$ степенями свободы.

Работа Бокса и Пирса была опубликована в 1970г. В том же году Дарбин показал, что статистика Дарбина-Уотсона не применима для проверки автокорреляции остатков при наличии в качестве регрессора объясняемой переменной с лагом. Одновременно Дарбин предложил так называемую h-статистику и альтернативную процедуру Дарбина для проверки наличия автокорреляции в таких моделях.

Позднее было замечено, что в моделях с лаговой объясняемой переменной в качестве регрессора статистика Бокса-Пирса не применима по тем же причинам, что и статистика Дарбина-Уотсона. Однако, поскольку в моделях временных рядов проверка наличия автокорреляции критически важна, статистика Бокса-Пирса нашла широкое применение и до сих пор входит в состав большинства специализированных эконометрических пакетов.

Так как статистика Бокса-Пирса имеет малую мощность, Бокс и Льюнг в 1978 г. предложили использовать для тех же целей **улучшенную Q-статистику**:

$$Q = (T+2)T\sum_{i=1}^{m}(T-i)^{-1}r_j^2$$

По сравнению со статистикой Бокса-Пирса различным слагаемым приданы разные Веса. Авторы доказали, что эта статистика имеет то же асимптотическое распределение χ^2_m, но лучше им аппроксимируется при конечном числе наблюдений.

Статистика Бокса-Льюнга теоретически не применима для тестирования автокорреляции остатков в моделях ARMA по тем же причинам, что и статистики Дарбина-Уотсона и Бокса-Пирса. Тем не менее, статистика Бокса-Льюнга входит во все специализированные пакеты, множество исследователей ею пользуются, хотя она теоретически не состоятельна.

Q-статистику Бокса-Пирса и улучшенную Q-статистику Бокса-Льюнга часто называют **портманто-статистикой** (portmanteau-statistics).

Пример. При помощи 100 нормальных случайных чисел $\{\varepsilon_t\}$ было построено 100 значений $\{y_t\}$:

$$y_t = -0,7y_{t-1} + \varepsilon_t - 0,7\varepsilon_{t-1}$$

y_0 и ε_0, были приняты равными нулю.

Происхождение данных считается неизвестным, и сравниваются три следующие модели:

Модель 1 ($AR(1)$): $y_t = \alpha_1 y_{t-1} + \varepsilon_t$
Модель 2 ($ARMA(1,1)$): $y_t = \alpha_1 y_{t-1} + \varepsilon_t + \beta_1\varepsilon_{t-1}$
Модель 3 ($AR(2)$): $y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \varepsilon_t$
Проверка автокорреляции остатков. Результаты расчетов таковы:
$Q_{0,95}(8) = 15,51 \quad Q_{0,95}(24) = 36,42$
Модель 1: $\alpha_1 = -0,835$
$Q(8) = 26,19(0,000) \quad AIC = 507,3$
$Q(24) = 41,10(0,001) \quad BIC = 509,9$
Модель 2: $\alpha_1 = -0,679 \quad \beta_1 = -0,676$
$Q(8) = 3,86(0,695) \quad AIC = 481,4$
$Q(24) = 14,23(0,892) \quad BIC = 486,6$
Модель 3: $\alpha_1 = -1,16 \quad \alpha_2 = -0,378$
$Q(8) = 11,44(0,057) \quad AIC = 492,5$
$Q(24) = 22,59(0,424) \quad BIC = 497,7$

Модель 1 должна быть отброшена по значениям статистик $Q(8)$ и $Q(24)$. Предпочтение должно быть отдано модели 2 перед моделью 3 по значениям обоих информационных критериев, а также потому, что в модели 3 значение статистики $Q(8)$ указывает на наличие определенной корреляции между остатками.

Q-статистика (Бокса-Пирса) и улучшенная Q-статистика (Бокса-Льюнга) не применимы для временных рядов с лаговой переменной в авторегрессионной части, но практически они используются для проверки автокорреляционных остатков.

Для проверки наличия автокорреляции в моделях ARMA лучше воспользоваться более мощным и универсальным способом, а именно **методом множителей Лагранжа** (Lagrange multiplier - LM), применительно к проверке автокорреляции остатков его еще называют тестом Бройша-Годфри (Breusch-Godfrey).

Он входит «триаду» классических асимптотических тестов: отношения правдоподобий, Вальда, множителей Лагранжа, и применим для широкого класса задач проверки ограничений на коэффициенты модели.

Пусть рассматривается модель множественной регрессии

$y_t = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + \varepsilon_t$

где $x_1, ..., x_k$ разные регрессоры, в том числе, возможно, и лаговые значения как объясняющих, так и объясняемой переменных.

Проверим предположение, что ε_t подчиняется авторегрессионной схеме порядка p, т.е. задается уравнением $\varepsilon_t = \gamma_1\varepsilon_{t-1} + \gamma_2\varepsilon_{t-2} + ... + \gamma_p\varepsilon_{t-p} + u_t$, где u_t - белый шум.

В терминах коэффициентов модели основная и альтернативная гипотезы принимают вид $H_0 : \gamma_1 = \gamma_2 = ... = \gamma_p = 0, \quad H_1 : \sum_{i=1}^{p} \gamma_i^2 > 0$

Метод множителей Лагранжа для проверки этой гипотезы заключается в следующем. Методом МНК строится обычная регрессия вида $y_t = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + \varepsilon_t$. Обозначим ее остатки через e_t.

1. Строится регрессия либо той де объясняемой переменной y_t, либо остатков e_t на старые регрессоры и остатки с лагом до p включительно (т.е. в качестве дополнительных объясняющих переменных используем $e_{t-1}, e_{t-2}, ..., e_{t-p}$).
$\widehat{y}_t = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + ... + \widehat{\beta}_k x_k + \widehat{\gamma}_1 e_{t-1} + \widehat{\gamma}_2 e_{t-2} + ... + \widehat{\gamma}_p e_{t-p}$

2. Проверяется гипотеза о том, что группа дополнительных объясняющих переменных является излишней. Если в качестве объясняемой переменной используются остатки, то статистика TR^2 имеет асимптотическое (при увеличении числа наблюдений) распределение χ^2 с p степенями свободы (T - число наблюдений, R^2 - коэффициент множественной детерминации регрессии, построенной в пункте 1). Если $TR^2\chi^2_p$ при $T \to \infty$, то принимаем гипотезу.

Для проверки гипотезы о равенстве нулю группы переменных обычно мы привыкли использовать F-статистику, проверяющую, фактически, статистическую значимость уменьшения остаточной суммы квадратов от включения дополнительных переменных. Разумеется, F-статистика применима и в этом случае. Но, она применима только при нормальном распределении случайного члена ε_t. Применение же теста множителей Лагранжа не требует нормальности распределения, но «работает» только асимптотически.

Современные эконометрические пакеты, в частности Eviews, сообщают пользователю при применении LM-теста значения и критические вероятности (p-values) для обеих статистик (F и TR^2), так что можно выбирать: использовать ли F-отношение, верное для конечных выборок, но в предположении нормальности, либо не требовать нормальности и использовать статистику TR^2, верную лишь асимптотически.

Для проверки нормальности существует множество тестов. Рассмотрим тест **Харке-Бера** (Jarque-Bera).

Этот тест вычисляет выборочные значения для коэффициентов асимметрии $S = \frac{1}{T\sigma^3}\sum(e_t - \bar{x})^3$ и эксцесса $K = \frac{1}{T\sigma^4}\sum(e_t - \bar{x})^4$, где $\bar{x}$ - выборочное среднее, а σ - выборочное среднеквадратичное отклонение остатков модели.

При условии нормальности остатков, статистика Харке-Бера $\frac{T-p-q-1}{6}\left[S^2 + \frac{1}{4}(K-3)^2\right]$ имеет χ^2 распределение с двумя степенями свободы.

IV этап "Прогнозирование с помощью моделей ARMA"

Прогнозирование будущих значений экономической величины является одним из основных способов применения моделей временных рядов. Использование моделей типа

ARMA обладает некоторыми особенностями по сравнению с прогнозированием по модели множественной регрессии.

При прогнозировании по правильно специфированной модели существует 2 источника ошибок прогноза:

- неопределнность будущих значений случайной величины ε
- отсутствие точных значений коэффициентов модели $\alpha_1, \alpha_2, ..., \alpha_p, \beta_1, \beta_2, ..., \beta_q$ (у нас есть только их оценки $\widehat{\alpha}_1, \widehat{\alpha}_2, ..., \widehat{\alpha}_p, \widehat{\beta}_1, \widehat{\beta}_2, ..., \widehat{\beta}_q$, оцененные по имеющейся выборке).

При прогнозировании по модели множественной регрессии мы оцениваем значение зависимой переменной y_t при заданных значениях независимых переменных - регрессоров $x_{t1}, x_{t2}, ..., x_{tk}$. Это приводит к тому, что характеристики прогноза, как случайной величины, являются, по сути, условными характеристиками при условии имеющейся выборки независимых переменных.

Иная ситуация в моделях типа ARIMA. Здесь значение переменной x_t прогнозируется для некоторого будущего момента времени $t + \tau$, при этом лаговые значения этой переменной $x_t, x_{t-1}, x_{t-2}, ..., x_{t-p}$, служащие регрессорами модели, можно рассматривать фиксированными на выборочных значениях, или случайными. Первая возможность, когда $x_t, x_{t-1}, x_{t-2}, ..., x_{t-p}$ фиксированы, приводит к **условному прогнозу**, как и для модели мноожественной регресии, а вторая, когда $x_t, x_{t-1}, x_{t-2}, ..., x_{t-p}$ случайные, - к **безусловному прогнозу**. Таким ообразом, при прогнозировании по модели типа ARIMA можно рассматривать как условный, так и безусловный прогнозы. Из курса теории вероятностей известно, что условная дисперсия случайной величины не превышает ее безусловную дисперсию, поэтому точность условного прогноза всегда выше.

Условный прогноз:

прогноз на 1 шаг:

$$\tilde{x}_{T+1} = E\{\theta + \varepsilon_{T+1} + \beta_1\varepsilon_T + ... + \beta_q\varepsilon_{T-q+1}|x_1, ..., x_T\} = \theta + \beta_1 l_T + ... + \beta_q l_{T-q+1}$$

прогноз на 2 шага:

$$\tilde{x}_{T+2} = E\{\theta + \varepsilon_{T+2} + \beta_1\varepsilon_{T+1} + ... + \beta_q\varepsilon_{T-q+2}|x_1, ..., x_T\} = \theta + \beta_2 l_T + ... + \beta_q l_{T-q+2}$$

прогноз на q шагов:

$$\tilde{x}_{T+q} = E\{\theta + \varepsilon_{T+q} + \beta_1\varepsilon_{T+q-1} + ... + \beta_q\varepsilon_T|x_1, ..., x_T\} = \theta + \beta_q l_T$$

$\tilde{x}_{T+q+1} = E\{\theta + \varepsilon_{T+q+1} + \beta_1\varepsilon_{T+q} + ... + \beta_q\varepsilon_{T+1}|x_1, ..., x_T\} = \theta$ - условный прогноз совпадает с безусловным

Тест Харке — Бера

Материал из Википедии — свободной энциклопедии

Тест Харке—Бера (англ. *Jarque-Bera test*) — это статистический тест, проверяющий ошибки наблюдений на нормальность посредством сверки их третьего момента (асимметрия) и четвёртого момента (эксцесс) с моментами нормального распределения, у которого $S = 0, K = 3$.

В тесте Харке—Бера проверяется нулевая гипотеза $\mathcal{H}_0$: $S = 0,\ K = 3$ против гипотезы $\mathcal{H}_1$: $S \neq 0,\ K \neq 3$, где S — коэффициент асимметрии (Skewness), K — коэффициент эксцесса (Kurtosis)

Формулировка

Тест выглядит следующим образом:

$JB = n\left(\frac{S^2}{6} + \frac{(K-3)^2}{24}\right)$, где $S = \frac{\sum e_i^3}{n\hat{\sigma}_{ML}^3}, K = \frac{\sum e_i^4}{n\hat{\sigma}_{ML}^4}, e_i$ — остатки модели, n — количество наблюдений, $\hat{\sigma}_{ML}^2 = \frac{\sum e_i^2}{n}, ML$ — обозначение метода максимального правдоподобия (Maximal Likelihood). Данная статистика имеет распределение хи-квадрат с двумя степенями свободы (χ_2^2), поскольку коэффициенты S и K асимптотически нормальны, следовательно, их квадраты при нормировке дадут две случайные величины, распределённые как χ_1^2. Чем ближе распределение ошибок к нормальному, тем меньше статистика Харке—Бера отличается от нуля. При достаточно большом значении статистики *p-value* будет мало, и тогда будет основание отвергнуть нулевую гипотезу (статистика попала в «хвост» распределения).

Свойства теста

Тест Харке—Бера является *асимптотическим* тестом, то есть применим к большим выборкам. Если ошибки распределены нормально, то в соответствии с теоремой Гаусса—Маркова оценки метода наименьших квадратов будут лучшими (иметь наименьшую дисперсию в классе линейных несмещённых оценок), и коэффициенты регрессии будут также распределены асимптотически нормально.

Литература

Damodar N. Gujarati. Basic Econometrics. — 4. — The McGraw-Hill Companies, 2004. — С. 1002. — ISBN 978-0071123433.

Источник - "https://ru.wikipedia.org/wiki/Тест_Харке_—_Бера"

При прогнозировании по модели ARIMA от имеющейся выборки зависят как оценки коэффициентов модели $\widehat{\alpha}_1, \widehat{\alpha}_2, ..., \widehat{\alpha}_p, \widehat{\beta}_1, \widehat{\beta}_2, ..., \widehat{\beta}_q$, так и значения регрессоров $x_t, x_{t-1}, x_{t-2}, ..., x_{t-p}$, поэтому сложно аналитически выразить условную дисперсию ошибки прогноза через имеющиеся значения временного ряда. Общепринято ограничиваться не очень реалистичным предположением о том, что коэффициенты модели известны точно. Разумеется, это предположение уменьшает дисперсию ошибки прогноза, чем увеличивает кажущуюся точность как условного, так и безусловного прогнозов.

Показано, что если мы хотим достичь минимума среднеквадратической ошибки (MSE), т.е. не требовать несмещенности ($E\{\widehat{\theta}\} \neq \theta$), то надо взять условное математическое ожидание: $E\{x_{T+\tau}|x_1, ..., x_T\}$. Такое условное математическое ожидание будет гаранировать получение MSE, или иногда ее обозначают MMSE (minimum mean square error).

Рассмотрим ***модель MA(q):*** $x_t = \theta + \varepsilon_t + \beta_1\varepsilon_{t-1} + ... + \beta_q\varepsilon_{t-q}$.

Полагаем, что:

1. коэффициенты модели точно известны;
2. имеются значения x_t для $t \in [1; T]$.

Очевидно, что безусловным точечным прогнозом для любого момента времени будет математическое ожидание процесса, т.е. θ ($\widehat{x}_{T+1} = E\{\theta + \varepsilon_{T+1} + \beta_1\varepsilon_T + ... + \beta_q\varepsilon_{T-q+1}\} = \theta$). Условным прогнозом для момента времени $T + 1$ будет условное математическое ожидание:

$$\widehat{x}_{T+1} = E\{\theta + \varepsilon_{T+1} + \beta_1\varepsilon_T + ... + \beta_q\varepsilon_{T-q+1}|x_1, ..., x_T\}.$$

Среди случайных величин ε, которые стоят в левой части, есть такие, которые связаны с имеющимися наблюдениями. Ведь наблюдение складывается из "модельного"значения и ошибки, поэтому условные математические ожидания всех слагаемых, кроме ε_{T+1}, не равны нулю: $E\{\varepsilon_{T+1}|x_1, ..., x_T\} = 0, E\{\varepsilon_T|x_1, ..., x_T\} \neq 0, E\{\varepsilon_{T-q+1}|x_1, ..., x_T\} \neq$

0 ($E\{y|x\} = E\{y\}$, если x и y независимы). Рассмотрим, например, $E\{\varepsilon_T|x_1, ..., x_T\}$. Это математическое ожидание - остаток между наблюдением и прогнозом по модели, т.е. $E\{\varepsilon_T|x_1, ..., x_T\} = e_T = x_T - \widehat{x}_T$. Поэтому условные математические ожидания от всех предыдущих значений случайной составляющей надо заменить соответствующими остатками. Следовательно, условные математические ожидания для прошлых значений - остатки, для будущих значений - нули.

Точно так же конструируется прогноз не на 1, а на 2 и вообще на τ шагов вперед. Все последующие ε заменяются нулями, а предыдущие - заменяются реально наблюдаемыми остатками. Таким образом, для модели *MA(q)* прогноз зависит от того, какие ошибки были на предыдущих шагах. Начиная с шага $(q + 1)$ условный прогноз представляет собой просто математическое ожидание θ, т.е. условный прогноз совпадает с безусловным: $\forall\tau : \tau \geq q + 1$
$\tilde{x}_{T+\tau} = E\{x_{T+\tau}\} = \theta$, следовательно q - кратная память об ошибках.

Рассмотрим условную дисперсию ошибки прогноза на 1 шаг:
$Var\{x_{T+1} - \widehat{x}_{T+1}|x_1, ..., x_T\} = E\{(\theta + \varepsilon_{T+1} + \beta_1\varepsilon_T + ... + \beta_q\varepsilon_{T-q+1} - \theta - \beta_1 e_T - ... - \beta_q e_{T-q+1}|x_1, ..., x_T)^2\} = \sigma_\varepsilon^2$
Остатки определяются из уравнения $x_i = \widehat{x_i} + e_i$.

Аналогично дисперсия прогноза на 2 шага равна $Var\{x_{T+2} - \widehat{x}_{T+2}|x_1, ..., x_T\} = (1 + \beta_1^2)\sigma_\varepsilon^2$, а дисперсия прогноза на τ шагов составит $(1 + \beta_1^2 + ... + \beta_\tau^2)\sigma_\varepsilon^2$ при $\tau < q$. При $\tau \geq q$ дисперсия ошибки условного прогноза становтся равной дисперсии ошибки безусловного прогноза, т.е. просто дисперсии случайного процесса x_t.

Теперь рассмотрим ***модель стационарного процесса AR(p):***
$x_t = \theta + \alpha_1 x_{t-1} + ... + \alpha_p x_{t-p} + \varepsilon_t$.

Для прогноза на 1 шаг вперед можно записать:
$Var\{x_{T+1} - \tilde{x}_{T+1}|x_1, ..., x_T\} = E\{a_{T+1}^2|x_1, ..., x_T\} - E^2\{a_{T+1}|x_1, ..., x_T\}$
$a = x_{T+1} - \tilde{x}_{T+1}$
$E\{a_{T+1}^2|x_1, ..., x_T\} = E\{(\theta+\varepsilon_{T+1}+\beta_1\varepsilon_T+...+\beta_q\varepsilon_{T-q+1}-\theta-\beta_1 e_T-...-\beta_q e_{T-q+1}|x_1, ..., x_T)^2\} = E\{(\varepsilon_{T+1}|x_1, ..., x_T)^2\} = \sigma_\varepsilon^2$
$E^2\{a_{T+1}|x_1, ..., x_T\} = E^2\{\varepsilon_{T+1}|x_1, ..., x_T\} = 0$

прогноз на 2 шага:
$Var\{x_{T+2} - \tilde{x}_{T+2}|x_1, ..., x_T\} =$
$E\{(x_{T+2}-\tilde{x}_{T+2}|x_1, ..., x_T)^2\}-E^2\{x_{T+2}-\tilde{x}_{T+2}|x_1, ..., x_T\} = \{$второе слагаемое равно нулю$\}$
$= E\{(\theta + \varepsilon_{T+2} + \beta_1\varepsilon_{T+1} + ... + \beta_q\varepsilon_{T-q+2} - \theta - \beta_1 e_T - ... - \beta_q e_{T-q+2}|x_1, ..., x_T)^2\} =$
$E\{(\varepsilon_{T+2} + \beta_1\varepsilon_{T+1}|x_1, ..., x_T)^2\} = (1 + \beta_1^2)\sigma_\varepsilon^2$

прогноз на τ шагов:
$Var\{x_{T+\tau} - \tilde{x}_{T+\tau}|x_1, ..., x_T\} = (1 + \beta_1^2 + ... + \beta_\tau^2)\sigma_\varepsilon^2$, если $\tau < q$
$Var\{x_{T+\tau} - \tilde{x}_{T+\tau}|x_1, ..., x_T\} = Var\{x_{T+\tau} - \tilde{x}_{T+\tau}\}$, если $\tau \geq q$.

$$Var\{x_{T+1} - \widehat{x}_{T+1}|x_1, ..., x_T\} = Var\{\theta + \varepsilon_{T+1} + \beta_1\varepsilon_T + ... + \beta_q\varepsilon_{T-q+1} - \theta - \widehat{b_1}e\}$$

$$x_{T+1} - \widehat{x}_{T+1} = x_{T+1} - (x_T - e_T)$$

$$\widehat{x}_{T+1} = E\{\theta + \varepsilon_{T+1} + \beta_1\varepsilon_T + ... + \beta_q\varepsilon_{T-q+1}|x_1, ..., x_T\}$$

$$Var\{x|z\} := Cov(x, x|z) = E\{(x-E(x|z))^2|z\} = E\{(x_{T+1}-\widehat{x}_{T+1})-E(x_{T+1}-\widehat{x}_{T+1}|x_1, ..., x_T)\}$$

$E\{x_{T+2}|x_1, ..., x_T\} = E\{\theta+a_1x_{T+1}+...+a_px_{T-p}+\varepsilon_{T+2}|x_1, ..., x_T\} = \theta+a_1\widehat{x}_{T+1}+...+a_px_{T-p+1}$

$$\lim_{\tau\to\infty} E\{x_{T+\tau}|x_1, ..., x_T\} = \frac{\theta}{\theta - \alpha_1 - ... - \alpha_p}$$

$$x_t = a_1x_{t-1} + a_2x_{t-2} + \varepsilon_t$$

Условную дисперсию ошибки прогноза можно рассчитать аналогично случаю модели скользящего среднего, но выкладки становятся весьма громоздкими, даже для моделей малого порядка.

Рассмотрим, например, модель *AR(2)* без свободного члена. Тогда $\widehat{x}_{T+1} = \alpha_1 x_T + \alpha_2 x_{T-1}$, а $x_{T+1} = \alpha_1 x_T + \alpha_2 x_{T-1} + \varepsilon_{T+1}$. Очевидно, что дисперсия ошибки прогноза на 1 шаг равна σ_ε^2. Для прогноза на 2 шага соответственно получаем:

$$\tilde{x}_{T+2} = \alpha_1\tilde{x}_{T+1} + \alpha_2 x_T = \alpha_1(\alpha_1 x_T + \alpha_2 x_{T-1}) + \alpha_2 x_T,$$

$$x_{T+2} = \alpha_1 x_{T+1} + \alpha_2 x_T + \varepsilon_{T+2} = \alpha_1(\alpha_1 x_T + \alpha_2 x_{T-1} + \varepsilon_{T+1}) + \alpha_2 x_T + \varepsilon_{T+2},$$

Дисперсия ошибки прогноза на 2 шага равна: $(1 + \alpha_1^2)\sigma_\varepsilon^2$.

Для прогноза на 3 шага получим:

$$\tilde{x}_{T+3} = \alpha_1\tilde{x}_{T+2} + \alpha_2\tilde{x}_{T+1} = \alpha_1(\alpha_1(\alpha_1 x_T + \alpha_2 x_{T-1}) + \alpha_2 x_T) + \alpha_2(\alpha_1 x_T + \alpha_2 x_{T-1}),$$

$$x_{T+3} = \alpha_1 x_{T+2}+\alpha_2 x_{T+1}+\varepsilon_{T+3} = \alpha_1(\alpha_1(\alpha_1 x_T+\alpha_2 x_{T-1}+\varepsilon_{T+1})+\alpha_2 x_T+\varepsilon_{T+2})+\alpha_2(\alpha_1 x_T+\alpha_2 x_{T-1} +\varepsilon_{T+1}) + \varepsilon_{T+3}.$$

Дисперсия ошибки прогноза на 3 шага равна $(1 + \alpha_1^2 + \alpha_2^2 + 2\alpha_1^2\alpha_2 + \alpha_1^4)\sigma_\varepsilon^2$.

Очевидно, что дисперсия ошибки увеличивается от шага к шагу. Значительно более простые выражения для дисперсии ошибки прогноза получаются, если перейти от *AR(p)* представления модели к эквивалентному MA представлению $x_t = \theta + \varepsilon_t + \psi_1\varepsilon_{t-1} + ... + \psi_q\varepsilon_{t-q} + ...$, хотя и с бесконечным числом слагаемых. Тогда дисперсия ошибки прогноза на τ шагов выражается формулой $\sigma_\varepsilon^2\sum_{i=0}^{\tau-1}\psi_i^2(\psi_0 = 1)$. (*)

Для общей ***модели ARMA(p,q)*** нужно просто объединить полученные результаты. Если мы хотим получить прогноз с минимальной среднеквадратической ошибкой, то:
1) рассчитываем прогнозные значения $x_{T+\tau}$ по нашей модели, подставляя туда для времени $[1, T]$ - наблюденные значения $x_1, ..., x_T$ и рассчитанные значения остатков $e_1, ..., e_T$, а для всех последующих моментов времени $\tau = T + 1, T + 2, ...$ - заменяем остатки нулями ($e_{\tau+1} = 0, e_{\tau+2} = 0, ...$), а для значений $x_{T+1}, x_{T+2}, ...$ подставляем их прогнозные значения $\tilde{x}_{T+1}, \tilde{x}_{T+2}, ...$
2) Для получения дисперсии ошибки прогноза переходим от ARMA к MA представлению и пользуемся формулой (*).

$(*)Var\{x_{T+1}-\tilde{x}_{T+1}|x_1, ..., x_T\} = Var\{a_1x_T+\alpha_2x_{T-1}+\varepsilon_{T+1}-a_1x_T-\alpha_2x_{T-1}|x_1, ..., x_T\} = \sigma_\varepsilon^2$

Тогда $x_t = (1-\alpha L)^{-1}(1+\beta L)\epsilon_t = (1+\alpha L+\alpha^2 L^2+...)(1+\beta L)\epsilon_t = \epsilon_t+\eta_1\epsilon_{t-1}+\eta_2\epsilon_{t-2}+...$, $\eta_1 = \alpha+\beta, \eta_2 = \alpha(\alpha+\beta), ..., \eta_k = \alpha^{k-1}(\alpha+\beta)$. Начиная со второго коэффициенты убывают в геометрической прогрессии, следовательно, дисперсия ошибки прогноза на τ шагов:

$$\sigma_\epsilon^2 \sum_{i=0}^{\tau-1} \eta_i^2 = \sigma_\epsilon^2 \left(1 + \sum_{i=1}^{\tau-1} \alpha^{2i}(\alpha+\beta)^2\right) = \sigma_\epsilon^2 \left(1 + \frac{(\alpha+\beta)^2(1-\alpha^{2\tau})}{1-\alpha^2}\right).$$

Для этой модели дисперсия ошибки прогноза асимптотически равна дисперсии временного ряда.

Во всех рассмотренных случаях условный точечный прогноз асимптотически приближался к математическому ожиданию ряда, а дисперсия ошибки прогноза - к дисперсии ряда. Это означает, что для стационарного процесса влияние имеющейся информации о реализации на прогноз и его точность асимптотически убывает до нуля. К тому же при увеличении горизонта прогноза дисперсия ошибки не превышает дисперсии временного ряда. Но этот результат является следствием нереалистического предположения о том, что коэффициенты модели известны точно.

Так же, как и в моделях множественной линейной регрессии, хорошее качество подгонки модели ARIMA не гарантирует высокой точности прогноза, т.е. высокой прогнозной силы. Для оценки прогнозных свойств модели одним из общеупотребительных приемов является разбиение имеющейся реализации на 2 части.

По первым n наблюдениям выбирается и оценивается модель, а по последним $(T-n)$ наблюдениям проводится сравнение наблюденных и рассчитанных по модели значений.

Такая процедура иногда называется ***постпрогнозом***.

Нестационарные временные ряды

Модели типа ARMA охватывают все стационарные процессы, а с нестационарными временными рядами ситуация иная, фактически рассмотрим только частные виды нестационарных временных рядов. Один из таких видов уже начинал рассматриваться - $ARIMA(p,d,q)$. По определению модели $ARIMA(p,d,q), d$ - это степень интеграции ряда, т.е. ряд становится стационарным после применения d раз операции взятия последовательной разности.

Рассмотрим нестационарные ряды, которые могут быть приведены к стационарному виду с помощью взятия последовательных разностей. Было получено, что 2 разных по свойствам типа нестационарных рядов приводятся к стационарному виду с помощью взятия последовательных разностей:

1 тип.***Процесс с детерминированным полиномиальным трендом***

$x_t = P_k(t)$, где $P_k(t)$ - полином степени k от t, а ϵ_t - стационарный процесс, не обязательно белый шум.

Если ограничиться рассмотрением только линейного тренда $x_t = \alpha+\beta t+\epsilon_t$ то можно записать: $\Delta x_t = x_t - x_{t-1} = (1-L)x_t = \beta + (\epsilon_t - \epsilon_{t-1})$

$$x_t = a + bt + ct^2 + \epsilon_t$$

$$\Delta x_t = a + bt + ct^2 + \epsilon_t - (a + b(t-1) + c(t-1)^2 + \epsilon_{t-1}) =$$

$$= (b-c) + 2ct + (\epsilon_t - \epsilon_{t-1})$$

$$\Delta^2 x_t = (b-c) + 2ct + (\epsilon_t - \epsilon_{t-1}) - ((b-c) + 2c(t-1) + (\epsilon_{t-1} + \epsilon_{t-2})) =$$

$$= 2c + \epsilon_t - 2\epsilon_{t-1} + \epsilon_{t-2}.$$

$\epsilon_t - 2\epsilon_{t-1} + \epsilon_{t-2}$ - MA-часть.

Поскольку ϵ_t - стационарный процесс, то его первая разность также стационарный процесс, хотя если ϵ_t - белый шум, то появляется MA-часть. В случае полиномиального тренда для приведения к стационарному виду нужно взять последовательную разность несколько раз.

2 тип.***Процесс случайного блуждания***

$x_t = \mu + x_{t-1} + \epsilon_t$

В этом случае $\Delta x_t = \mu + \epsilon_t$, и процесс x_t называется случайным блужданием с дрейфом. Решение этого разностного уравнения можно записать в следующем виде:

$$x_t = \mu + x_{t-1} + \epsilon_t = \mu + (\mu + x_{t-2} + \epsilon_{t-1}) + \epsilon_t = 2\mu + (\mu + x_{t-3} +$$

$$+\epsilon_{t-2}) + \epsilon_{t-1} + \epsilon_t = 3\mu + x_{t-3} + \epsilon_t + \epsilon_{t-1} + \epsilon_{t-2}$$

$$x_t = x_0 + \mu t + \sum_{j=0}^{t-1} \epsilon_{t-j}$$

Если применить подход Бокса-Дженкинса и, имея некоторую реализацию, перейти к разностям, оценить модель $ARMA(p,q)$, то чтобы вернуться назад к процессу x_t, нужно выбрать, по какой схеме возвращаться. Как мы уже рассматривали, эти процессы ведут себя по-разному. В чем-то они схожи: у обоих есть линейный тренд, но они отличаются случайной частью. В первом случайная часть - это текущий шок, текущие возмущения, а во втором - это накопленные возмущения от всех предыдущих шоков.

Ряд второго типа можно привести к стационарному виду только взятием первой разности, а ряд первого типа можно привести к стационарному виду как взятием разности, так и выделением линейного тренда. Например:

$$x_t = \theta_0 + \theta_1 t + \theta_2 t^2 + \epsilon_t$$

$$y_t = \alpha + \beta t + \gamma t^2 + z_t$$

$$z_t = \epsilon_t + 2\epsilon_{t-1} + 3\epsilon_{t-2} + ... + t\epsilon_t,\ t = \overline{1,T}$$

Проведем операцию детрендирования:

$$x_t^0 = x_t - (\theta_0 + \theta_1 t + \theta_2 t^2) = \epsilon_t - \text{стационарный ряд}$$

$$y_t^0 = y_t - (\alpha + \beta t + \gamma t^2) = z_t$$

$$D(z_t) = D(\epsilon_t + 2\epsilon_{t-1} + 3\epsilon_{t-2} + ... + t\epsilon_t) = \sigma_\epsilon^2(1 + 2 + ... + t) = \sigma_\epsilon^2 t\frac{(t+1)}{2} \Rightarrow$$

Детрендированный ряд y_t^0– не стационарный ряд.

Вопрос: Применим ли такой подход выделения линейного тренда к случайному блужданию? Что произойдет, если на самом деле процесс является случайным блужданием, пусть даже для простоты с нулевым математическим ожиданием, т.е. имеет вид $\Delta x_t = \epsilon_t$, а мы построим регрессию на время вида $x_t = \alpha + \beta t + \epsilon_t$?

Для краткости введем следующие названия для этих типов нестационарных процессов.

1 тип.***Процесс, приводимый к стационарному путем выделения линейного тренда - TSP***(trend stationary process).

Это процесс вида $x_t = \alpha + \beta t + \epsilon_t$, он приводится к стационарному процессу путем включения в регрессию линейного тренда. Это, в принципе, процесс, у которого есть детерминированный тренд.

2 тип.***Процесс, приводимый к стационарному путем взятия первой разности - DSP***(diferencing stationary process).

Это процесс вида $x_t = \mu + x_{t-1} + \epsilon_t$.

Свойства процессов TSP и DSP

Процесс TSP	Процесс DSP
Не стационарен из-за непостоянного тренда.	Не стационарен из-за непостоянной дисперсии.
Конечная память о шоках, он забывает об ошибке на предыдущем шаге сразу. Если вместо белого шума будет стоять более общий процесс $ARMA(p,q)$, то, конечно, шоки сказываются некоторое время, но их влияние со временем ослабевает, т.е. процесс с конечной памятью в случайных воздействиях.	Так как в явном решении стоит сумма всех предыдущих ϵ, то шоки помнятся все время. Это процесс с бесконечной памятью. Экономически это не очень понятно, шоки не должны сказываться постоянно.

В 1984 г. была опубликована работа Нельсона и Канга, в которой они задались вопросом: что произойдет, если процесс является DSP, а мы будем выделять линейный тренд, как для процесса типа TSP? Их результаты были получены с помощью большого количества испытаний Монте-Карло и заключались в следующем.

1. Линейная регрессия случайного блуждания без дрейфа на время дает коэффициент множественной детерминации ряда $x_t = x_{t-1} + \epsilon_t$ $R^2 \approx 0,44$ независимо от размера выборки, т.е. R^2 значим даже для малого количества точек наблюдения.

2. Если дрейф присутствует ($\mu \neq 0$), то $R^2 > 0,44$ и зависит от размера выборки. Это интуитивно можно понять, потому что как только $\mu \neq 0$, появляется свой тренд-нестационарное среднее.

3. Также оказалось, что $R^2 \to 1$ при $T \to \infty$, т.е. для процесса DSP с дрейфом $R^2 \to 1$ при увеличении объема выборки. Это явление значимого коэффициента тренда, когда в истинном уравнении тренд отсутствует, было названо **«кажущиеся тренды»** (spurious trends).

4. Оценка дисперсии остатков составляет примерно 14% от истинной дисперсии случайного возмущения, т.е. оценка дисперсии сильно занижена. Процедура оценивания указывает на значимый тренд и малую дисперсию, хотя на самом деле процесс случайного блуждания характеризуется растущей без ограничений дисперсией.

5. Остатки регрессии оказываются коррелированными с коэффициентом корреляции, примерно равным $\rho_1 = 1 - \frac{10}{T}$

6. В этом случае t-статистика не годится для проверки гипотезы о значимости коэффициента при времени, поскольку она смещена в сторону принятия гипотезы о наличии линейного тренда.

7. Независимые случайные блуждания имеют высокую корреляционную зависимость.

Поясним свойство 7. Если взять 2 независимых случайных блуждания:

$y_t = y_{t-1} + v_t$

$x_t = x_{t-1} + \epsilon_t$

где v_t, и ϵ_t - независимые белые шумы, и построить регрессию $_t = \alpha + \beta_t + _t$, то предполагается, что коэффициент β будет незначим. Однако два независимых случайных блуждания показывают высокую регрессионную зависимость, коэффициент β оказывается значимым. Как можно это пояснить? Если оба процесса имеют значимые тренды, то они оба зависят от $t \Rightarrow$ процессы зависимы межлу собой. А это означает, что если каждая из величин, между которыми мы ищем регрессионную зависимость, является DS-процессом, то регрессия между ними является «кажущейся».

Таким образом, некоторые зависимости являются кажущимися. Например представляется, что связаны динамики денежной массы и инфляции, но не учтено, что оба процесса - DSP, и сделанный экономический вывод неправомочен. Этот результат Нельсона и Канга показывает опасность прямого применения обычного регрессионного анализа в том случае, когда данные являются типа DS.

Значимость этого результата особо подчеркнута более ранней, знаменитой работой Нельсона и Плоссера, которая связана с исследованием исторических макроэкономических рядов в США. Результат этого исследования показал, что практически все ряды макроэкономических показателей США, за исключением ряда уровня безработицы, оказались рядами типа DS, т.е. типа случайного блуждания с дрейфом. Это произвело впечатление разорвавшейся бомбы. Известный экономист Саржент заметил, что все, что сделано до сих пор в области макроэкономической динамики, подлежит пересмотру. Правда, дальнейшие исследования показали, что ситуация не столь драматична.

Две эти работы, Нельсона-Канга и Нельсона-Плоссера, заставили обратить внимание на то, что надо проводить четкое разграничение между рядами двух видов: DS и TS. Если тренд в модели типа TS отсутствует, то вопрос сводился к различию между стационарными и нестационарными рядами. Раньше мы рассматривали для этого поведение выборочной автокорреляционной функции.

Визуальное исследование поведения выборочной автокорреляционной функции, т.е. применение подхода Бокса-Дженкинса, не позволяет надежно выбрать, к какому из типов относится исследуемый ряд.

Необходим метод, позволяющий формально проводить различие между рядами типа TSP и DSP.

Рассмотрим модель: $x_t = \alpha + \rho x_{t-1} + \beta t + \epsilon_t$.

Эта модель вобрала в себя черты обеих моделей: TS и DS, и гипотезы о характере ряда можно записать в виде простых гипотез о ее параметрах:

H_0: ряд является DS $\Rightarrow \begin{cases} \rho = 1 \\ \beta = 0 \end{cases}$

H_1: ряд является TS $\Rightarrow \rho < 1$, (ϵ - не просто белый шум, а некоторый стационарный ряд)

Нулевая гипотеза относится к классу общих линейных гипотез.

При традиционном подходе для ее проверки нужно оценить 2 регрессии: $x_t = \alpha + \rho x_{t-1} + \beta t + \epsilon_t$, и $x_t = \alpha + x_{t-1} + \epsilon_t$. И затем проверить значимость разности остаточных сумм квадратов, используя F-статистику.

Рассмотрим другой подход для более простой версии этой модели: $_t = \alpha + \rho x_{t-1} + \epsilon_t$, т.е. без включения линейного тренда.

В этом случае гипотезы принимают вид:

H_0: $\rho = 1 \Rightarrow$ ряд является DS.

H_1: $\rho < 1 \Rightarrow$ ряд является TS.

Проверим гипотезу о том, что $\rho = 1$ при помощи t-статистики.

После вычитания из обеих частей уравнения x_{t-1} его можно переписать в другом виде:

$\Delta x_t = \alpha + (\rho - 1)x_{t-1} + \epsilon_t$.

Пусть $\rho - 1 = \gamma$, тогда проверяемые гипотезы примут вид:

$H_0 : \gamma = 0 \Rightarrow$ ряд является DS,

$H_1 : \gamma < 0 \Rightarrow$ ряд является TS.

В классической линейной регрессии для проверки такой гипотезы применяется односторонняя t-статистика. Но в случае выполнения нулевой гипотезы, ряд x_t является случайным блужданием, его дисперсия стремится к бесконечности при увеличении t, и $\frac{\hat{\gamma}}{s.e.(\hat{\gamma})}$, где $s..(\hat{\gamma})$ - оценка дисперсии $\hat{\gamma}$, не является распределением Стьюдента. Следовательно, и асимптотическое распределение этой статистики не является нормальным. Причиной является невыполнение условий Центральной предельной теоремы в этом случае. Критические точки этого распределения приходится рассчитывать численно, используя иммитационное моделирование Монте-Карло. Впервые это распределение было выведено и затабулировано в работе Дикки и Фуллера и носит их имя. Тест, использующий для проверки типа нестационарности распределение статистики $\frac{\hat{\gamma}}{s.e.(\hat{\gamma})}$, при условии $\gamma = 0$, т.е. когда процесс принадлежит типу DS, называется **тестом Дикки- Фуллера**, и обозначается как **DF-тест**. При условии, что нулевая гипотеза о том, что $\gamma = 0$, выполнена, мы имеем процесс типа случайного блуждания (DSP). Именно для этого случая не работает t-статистика.

Позже МакКиннон расширил эти таблицы, рассмотрел некоторые другие случаи и предложил аппроксимирующие формулы для быстрого расчета критических точек в компьютерных программах. С помощью моделирования Дикки и Фуллер также рассчитали аналог F-статистики для случая, когда процесс является случайным блужданием. Это распределение также иногда называют распределением Дикки-Фуллера. Обычно из контекста ясно, о каком из распределений идет речь.

Вернемся к общей модели: $x_t = \alpha + \rho x_{t-1} + \beta t + \epsilon_t$. В этом случае мы имеем следующие гипотезы:

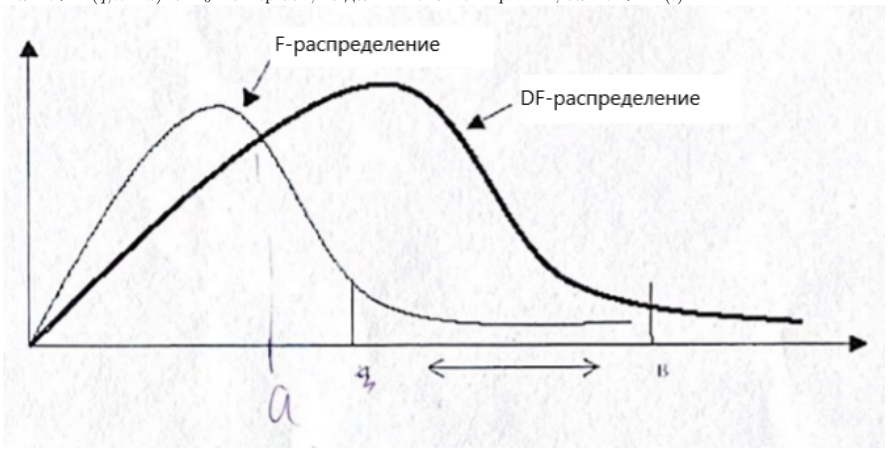
$H_0 : \rho = 1, \beta = 0$,

$H_1 : \rho < 1$,

Сравним, как ведет себя DF-статистика и F-статистика. Нас интересует критическое значение при заданом числе наблюдений и для заданного уровня значимости. Если - количество наблюдений, то в зависимости от него получим следующую таблицу для уровня значимости 5%, в которой F-статистика - это фактически $F_{q,n-k}$, где $q = 2$ - число ограничений, $n - k = T - 3$ - число регрессоров.

T	DF	F-статистика
25	7,24	3,42
50	6,73	3,20
100	**6,49**	**3,10**
∞	6,25	3,00

Если $t > F(q, n-k) \Rightarrow H_0-$ отвергаем, но должны были бы принять, если $t < DF(t)$



На рисунке схематически показаны графики плотностей распределения Дикки-Фулера и Фишера. Видно, что DF-распределение значительно правее, чем F-распределение. Это означает, что в большом количестве случаев применение стандартной F-статистики ведет к тому, что мы считаем, что ряд относится к типу TS, в то время как он типа DS, а именно, если F-статистика попадает в область (а; Ь).

Лекция 9

Нестационарные временные ряды (продолжение)

Пример, иллюстрирующий поведение стационарных и нестационарных процессов.

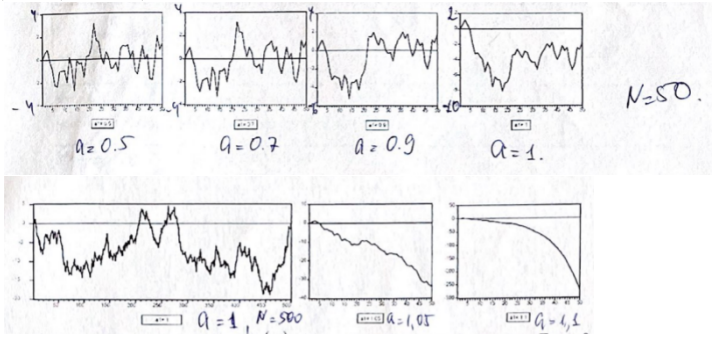
За основу возьмем наиболее простую модель - процесс AR(1):

$x_t = \alpha x_{t-1} + \epsilon_t$

Этот процесс является стационарным при выполнении условия: $-1 < \alpha_1 < 1$.

А как проявляется нестационарность ряда x_1 при нарушении этого условия?

Приведем смоделированные реализации такого ряда для следующих случаев: $\alpha_1 = 0.5, \alpha_1 = 0.7, \alpha_1 = 0.9, \alpha_1 = 1, \alpha_1 = 1.05, \alpha_1 = 1.1$..



Во всех случаях $x_1 = 1$, ϵ_t - белый гауссовский шум с $E\epsilon_t = 0$ и $Var(\epsilon_t) = 1$, математическое ожидание истинного процесса x_t равно единице, а поведение смоделированных рядов - различное.

Модель	Ко-во пересечений 0-го уровня	Среднее значение
Noise (белый шум)	25	-0,046
AR(1) $a_1 = 0,5$	14	-0,097
AR(1) $a_1 = 0,7$	8	-0,191
AR(1) $a_1 = 0,9$	8	-0,649
AR(1) $a_1 = 1,0$	1	-3,582
AR(1) $a_1 = 1,05$	1	-13,511
AR(1) $a_1 = 1,1$	1	-59,621

При возрастании значения α_1, от $\alpha_1 = 0$ (белый шум) до $\alpha_1 = 1$ количество пересечений нулевого уровня уменьшается, все более длинными становя тся периоды, в течение которых значения ряда находятся по одну сторону от нулевого уровня.

Вспомним основные свойства TSP и DSP процессов, которые сформулировали на прошлое лекции.

TSP - процессы, приводимые к стационарному взятием линейного тренда, т.е. ряды типа TSP имеют линию тренда в качестве некоторой “центральной линии”, которой следует траектория ряда, находясь то выше, то ниже этой линии, с достаточно частой сменой положений выше-ниже.

DSP - процессы, приводимые к стационарному взятием последовательных разностей, т.е. ряды типа DSP помимо детерминированного тренда (если таковой имеется) имеют еще и стохастический тренд, из-за присутствия которого траектория DSP ряда весьма долго пребывает по одну сторону от линии детерминированного тренда (выше или ниже этой линии), удаляясь от нее на значительные расстояния, так что по-существу в этом случае линия детерминированного тренда перестает играть роль “центральной” линии, вокруг которой колеблется траектория процесса.

В TSP-рядах влияние предыдущих шоковых воздействий затухает с течением времени, а в DSP-рядах такое затухание отсутствует и каждый отдельный шок влияет с одинаковой силой на все последующие значения ряда. Поэтому наличие стохастического тренда требует проведения определенной экономической политики для возвращения макроэкономической переменной к ее долговременной преспективе, тогда как при отсутствии стохастического тренда серьезных усилий для достижения такой цели не требуется - в этом случае макроэкономическая переменная "скользит" вдоль линии тренда как направляющей, пересекая ее достаточно часто и не уклоняясь от этой линии сколько-нибудь далеко, **(лекция 8)**.

Вопрос: Как правильно определить различия между TSP и DSP рядами для построения адекватной модели временного ряда, которую можно использовать для описания динамики ряда и прогнозирования его будущих значений?

Давайте займемся проблемой такой классификации.

Как оказалось, процедура Бокса-Дженкинса для отнесения ряда к одному из указанных двух классов: TSP или DSP на основании наблюдения реализации ряда на некотором интервале времени для этого не подходит.

В настоящее время существует множество процедур такой классификации, продолжают предлагаться новые процедуры, которые либо несколько превосходят старые в статистической эффективности (по крайней мере, теоретически) либо могут составить конкуренцию старым процедурам и служить дополнительным средством подтверждения классификации, произведенной другими методами.

При анализе конкретных макроэкономических временных рядов обычно применяют несколько разных статистических процедур, что позволяет несколько укрепить выводы, сделанные в пользу одной из двух (TSP или DSP) конкурирующих гипотез.

Таким образом, мы нуждаемся в методе, который позволит проводить формальные различия между TSP и DSP рядами.

Различение TSP и DSP рядов в классе моделей ARMA

Как уже отмечалось выше, для решения вопроса об отнесении исследуемого ряда $_t$ к классу TS (стационарных или стационарных относительно тренда) или DS (разностно стационарных) процессов имеется целый ряд различных процедур. Однако все эти процедуры страдают теми или иными недостатками. Процедуры, оформленные в виде формальных статистических критериев, как правило, имеют достаточно низкую мощность, что ведет к весьма частому неотвержению исходной (нулевой гипотезы), когда она в действительности не выполняется. В то же время невыполнение теоретических предпосылок, на которых основывается критерий, при применении его к реальным данным приводит к отличию реально наблюдаемого размера критерия от заявленного уровня значимости. Вследствие последнего обстоятельства теряется контроль над вероятностью ошибки первого рода, и это может приводить к слишком частому отвержению нулевой гипотезы, когда она в действительности верна. В связи с таким положением вещей исследователи обычно используют при анализе рядов на принадлежность их к классу TS или DS не один, а несколько критериев и подкрепляют выводы, полученные с использованием формальных критериев (с установленными уровнями значимости) графическими процедурами. Мы также рассмотрим несколько процедур различения TS и DS.

1. Гипотеза единичного корня

В большинстве критериев, предложенных для различения DS и TS гипотез, эта задача решается в классе моделей ARMA (стационарных и нестационарных).

Если ряд X_t имеет тип $ARIMA(p,d,q)$, то в результате его (d-кратного дифференцирования мы получаем стационарный ряд $\Delta^d x_t$, типа $ARMA(p,q)$:

$$\sum_{j=0}^{p} \alpha_j \Delta^d x_{t-1} = \sum_{j=0}^{p} \beta_j \epsilon_{t-1}$$

или

$$\alpha_p(L)\Delta^d x_t = \beta_q(L)\epsilon_t,$$

где $\alpha_p(L)$ и $\beta_q(L)$ - полиномы от оператора обратного сдвига L , имеющие степени p и q, соответственно. Заметим, что $\Delta x_t = (1-L)x_t$, так что $\Delta^d x_t = (1-L)^d x_t$ и $\alpha_p(L)(1-L)^d x_t = \beta_q(L)\epsilon_t$ или $\alpha_p(L)x_t = \beta_q(L)\epsilon_t$ где $\alpha_p(L) = \alpha_p(L)(1-L)^d$ - полином степени $(p+d)$.

Поскольку ряд $\Delta^d x_t$ стационарный, все p корней полинома $\alpha_p^*(z)$ находятся за пределами единичного круга, так что полином $\alpha_p^*(z)$ имеет корней за пределами единичного круга и d корней на границе этого круга, а точнее, корень $z=1$ кратности d.

Таким образом, ряд X_t представляется нестационарной моделью ARM$(p+d,q)$, в которой авторегрессионный полином $a(L)$ имеет ровно d корней, равных 1, а все остальные корни по модулю больше 1. Поэтому проверка нулевой гипотезы H_0 о том, что некоторый ARMA ряд X_t является DS-рядом (т.е. нестационарным рядом), может быть сведена к проверке гипотезы о том, что авторегрессионный полином $a(L)$ имеет хотя бы один корень, равный 1. Это оправданно, если исходить из предположения, что (z) не имеет корней внутри единичного круга, т.е. исключить из рассмотрения "взрывные" модели. При этом о гипотезе H_0 кратко говорят как о **гипотезе единичного корня (UR - unit root hypothesis)**, хотя точнее было бы говорить о гипотезе авторегрессионного единичного корня. В качестве альтернативной тогда выступает **TS гипотеза** о том, что рассматриваемый ARMA ряд - стационарный.

Критерии, в которых за исходную (нулевую) гипотезу берется гипотеза TS, служат скорее для подтверждения результатов проверки DS-гипотезы. В этом случае вместо проверки гипотезы единичного корня у полинома $a(z)$ проверяется гипотеза о наличии единичного корня $z=1$ у уравнения $b*(z)=0$, где $b*(L)$ - полином от оператора обратного сдвига L в представлении в виде процесса скользящего среднего. $X_t = b*(z) \in t$ ряда разностей. $X_t = X_t - X_{t-1}$ исходного процесса X_t.

$H_0 : X \sim ARMA(p+d,q) \sim$ DS-ряд $\Rightarrow$

$\Rightarrow H_0$: хотя бы 1 корень $z=1$ у полинома $\alpha_p(L) : \alpha_p(z) < 0$

H$_1 : X_t \sim ARMA(p+d,q) -$ стационарный ряд

$H_0 : X_t \sim ARMA(p+d,q) -$ TS-ряд $\Rightarrow$

H$_1 : z = 1 : b_q(z) = 0$ полином

b$_1^*(z) = 0$