стационарности процесса ... порядка ARC
первого ... авторегр

Билет 1. Понятие парной регрессии. Типы данных (пространственные и временные данные). Объясняемая и объясняющая переменные. Подгонка кривой. Мера отклонения.

28 – 31

Типы моделей:
- модели временных рядов
- регрессионные модели с одним ур.
- системы одновременных ур.

Типы данных:
- пространственные данные
  Пример: набор сведений (объем пр-ва, кол-во работников, доход и др.) по разным фирмам в один и тот же момент времени
- временные ~~ряды~~ данные
  Пример: ежеквартальные данные по инфляции, средней з/п, нац. доходу, денеж. эмиссии за последний год.

Отличительной чертой временных ~~рядов~~ данных явл. то, что они естественным образом упорядочены по времени, кроме того наблюдения в близкие моменты времени часто бывают зависимыми.

~~Модели парной регрессии~~:
Модели временных рядов:
к этому классу относятся модели:

тренда: $y(t) = T(t) + \varepsilon_t$

$T(t)$ — временной тренд заданного параметрич. вида

ε_t — случайная (стохаст.) компонента

сезонности: $y(t) = S(t) + \varepsilon_t$

$S(t)$ — периодическая (сезонная) компонента

тренда и сезонности: $y(t) = T(t) + S(t) + \varepsilon_t$ (аддитивная)

$y(t) = T(t)S(t) + \varepsilon_t$ (мультипликативная)

Также относятся: модели адаптивного прогноза, модели авторегрессии и скользящего среднего (ARIMA) и др.

Общей чертой явл. то, что они объясняют поведение временного ряда, исходя ~~из~~ только из его предыдущих значений.

...авторегрессии первого порядка AR(...
Условия стационарности процесса авторегр...
первого поря...

Регрессионные модели с одним ур.:

В таких моделях зависимая (объясняемая) переменная y представляется в виде функции $f(x, \beta) = f(x_1 \dots x_k, \beta_1 \dots \beta_p)$, где $x_1 \dots x_k$ - независимые (объясняющие) переменные, а $\beta_1 \dots \beta_p$ - параметры

В зависимости от вида функции $f(x, \beta)$ модели делятся на линейные и нелинейные

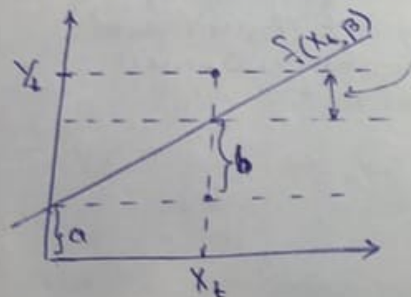
Пример: можно исследовать спрос на мороженное как ф. от времени, температуры воздуха и т.п.

Системы одновременных ур:

Эти модели описываются системами ур. Системы могут состоять из тождеств и регрессионных ур., каждое из которых может, кроме объясняющих перемен., включать в себя также объясняемые переменные из других ур. системы. Таким образом, мы имеем здесь набор объясняемых переменных связанных через ур. системы.

Модель парной регрессии:

Подгонка кривой: Пусть у нас есть набор значений двух переменных: (X_t, Y_t), $t = \overline{1,n}$. Можно отобразить пары точками на плоскости.



отклонение: $Y_t - f(X_t, \hat\beta)$

$\hat\beta = (a, b)$

Предположим, что нашей задачей явл. подобрать ф. $Y = f(X)$ из параметрич. семейства ф-ций $f(X, \beta)$, наилучшим способом описывающую зависимость Y от X. Подобрать ф. означает выбрать наилучшее значение параметра β.

В качестве меры отклонения ф-ции $f(X, \beta)$ от набора наблюдений можно можно взять:

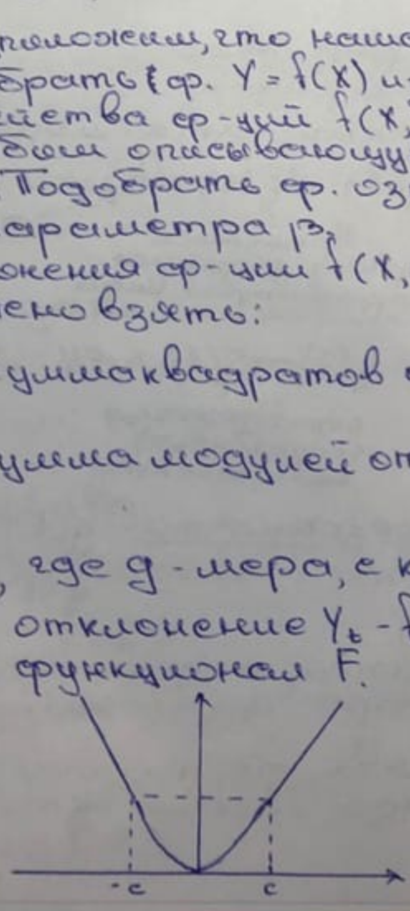
1. $F = \sum_{t=1}^{n} (Y_t - f(X_t, \beta))^2$ сумма квадратов отклонений

2. $F = \sum_{t=1}^{n} |Y_t - f(X_t, \beta)|$ сумма модулей отклонений

3. $F = \sum_{t=1}^{n} g(Y_t - f(X_t, \beta))$, где g - мера, с которой отклонение $Y_t - f(X_t, \beta)$ входит в функционал F.

Пример: ф. Хубера

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



Линейная регрессионная модель с двумя переменными.
Метод наименьших квадратов. Свойство гомоскедастичности.

$y_t = a + bx_t + \varepsilon_t \quad , t = \overline{1,n}$

x_t – неслучайная (детерминированная) величина;

y_t, ε_t – случайные величины;

x_t – незав. переменная / регрессор; y_t – зав. переменная
ε_t – ошибка.

1.) не учтены все факторы; 2) ошибки измерений;

Т.о. ε_t – случ. величина с некоторой ф-ей распределения,

которой соотв-ет ф-ия распределения случ. вел. y_t.

Основные гипотезы:

1. $y_t = a + bx_t + \varepsilon_t \quad , t = \overline{1,n}$;

2. x_t – детермин. величина; вектор $(x_1, .., x_n)^T$ неколлинеарен

вектору $i = (1, .., 1)^T$;

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

3в. $E(\varepsilon_t \varepsilon_s) = 0$, $t \neq s$ (т.е. ошибки некоррелир.);

3^+) $\varepsilon_t \sim N(0, \sigma^2)$, $t = \overline{1,n}$;

В 3а (усл-ие незав. дисперсии ошибки от номера наблюд.):

– гомоскедастичность (если выполнены 1) – 3а))
– гетероскедастичность (иначе).

Метод наимен. квадратов:

$\hat{y}_t = \hat{a} - \hat{b} x_t$ - прогноз в т. x_t

Остатки регрессии e_t

$y_t = \hat{y}_t + e_t = \hat{a} + \hat{b} x_t + e_t$, $e_t = y_t - \hat{y}_t$, e_t - случ.

МНК состоит в миним. функц-ла: $F = \sum_{t=1}^{n} (y_t - (a + b x_t + e_t))^2$

$$\hat{b} = \frac{n \sum_{t=1}^{n} x_t y_t - \sum_{t=1}^{n} x_t \sum_{t=1}^{n} y_t}{n \sum_{t=1}^{n} x_t^2 - (\sum_{t=1}^{n} x_t)^2}$$

$\hat{a} = \bar{y} - \hat{b} \bar{x}$, где $\bar{y} = \frac{1}{n} \sum_{t=1}^{n} y_t$, $\bar{x} = \frac{1}{n} \sum_{t=1}^{n} x_t$;

$\bar{y} = \hat{a} + \hat{b} \bar{x}$;

Билет 3. Линейная регрессионная модель с двумя переменными. Теорема Гаусса-Маркова. Оценка дисперсии ошибок σ^2. Оценка максимального правдоподобия коэф. регрессии

38-46 82-84, 55-57

Линейная регрессионная модель с двумя переменными:

$$Y_t = a + bX_t + \varepsilon_t, \quad t = \overline{1,n} \leftarrow \text{Регрессионное уравнение}$$

где X_t - неслучайная (детерминированная) вел-на, а Y_t, ε_t - случайные вел-ны. Y_t наз. объясняемой (зависимой) переменной, а X_t - объясняющей (независимой) переменной или регрессором

Основные гипотезы:

1. $Y_t = a + bX_t + \varepsilon_t$, $t = \overline{1,n}$ - спецификация модели
2. X_t - детермир. вел-на; вектор $(X_1 \dots X_n) = X$ не коллинеарен 1.

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

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

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

В этом случае модель называется нормальной линейной регрессионной

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

Теорема Гаусса-Маркова

В предположениях модели 1 - 3ab:

1. $Y_t = a + bX_t + \varepsilon_t$, $t = \overline{1,n}$

2. X_t - детерминированная вел-на

3a. $E\varepsilon_t = 0$, $E(\varepsilon_t^2) = Var(\varepsilon_t) = \sigma^2$

3b. $E(\varepsilon_t \varepsilon_s) = 0$, при $t \neq s$

... правдоподобия.

оценки

$$\hat{b} = \frac{n\sum X_tY_t - (\sum X_t)(\sum Y_t)}{n\sum X_t^2 - (\sum X_t)^2} = \frac{cov(X,Y)}{Var(X)}$$

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

полученные по МНК, имеют наименьшую дисперсию в классе всех линейных несмещенных оценок.

Итак, теперь у нас есть наилучшие оценки коэфф. регрессии a, b.

Обозначим через $\hat{Y}_t = \hat{a} + \hat{b}X_t$ прогноз значения Y_t в т. X_t

Остатки регрессии определяются из ур: $Y_t = \hat{Y}_t + e_t = \hat{a} + \hat{b}X_t + e_t$. e_t - остатки регрессии (сл. вел., наблюд.)

Оценка σ^2:

$$\sum e_t^2 = \sum(Y_t - \hat{a} - \hat{b}X_t)^2 = \sum(\overline{Y} + y_t - \hat{a} - \hat{b}\overline{X} - bx_t)^2 =$$

$$= \sum(y_t - \hat{b}x_t)^2 = \sum(bx_t + \varepsilon_t - \overline{\varepsilon} - \hat{b}x_t)^2 =$$

$$= \sum x_t^2(b-\hat{b})^2 + 2(b-\hat{b})\sum x_t(\varepsilon_t - \overline{\varepsilon}) + \sum(\varepsilon_t - \overline{\varepsilon})^2$$

Возьмем мат. ожидание от левой и правой части. Получим:

$$E\left(\sum x_t^2(b-\hat{b})^2\right) = \sigma^2$$

$$E\left(2(b-\hat{b})\sum x_t(\varepsilon_t - \overline{\varepsilon})\right) = -2\sigma^2$$

$$E\left(\sum(\varepsilon_t - \overline{\varepsilon})^2\right) = (n-1)\sigma^2$$

Таким образом, $E(\sum e_t^2) = (n-2)\sigma^2$. Отсюда следует, что $s^2 = \hat{\sigma}^2 = \frac{1}{n-2}\sum e_t^2$ явл. несмещенной оценкой дисперсии ошибок σ^2

На практике, дисперсия ошибок σ^2 неизвестна и оценивается по наблюдениям одновр. с коэфф. $\hat{a}, \hat{b}$.

$$Var(\hat{b}) = s^2\frac{1}{\sum x_t^2} = \frac{s^2}{\sum(X_t - \overline{X})^2}$$

$$Var(\hat{a}) = s^2\frac{\sum X_t^2}{n\sum x_t^2} = \frac{s^2\sum X_t^2}{n\sum(X_t - \overline{X})^2}$$

$$cov(\hat{a},\hat{b}) = -\frac{\overline{X}}{\sum x_t^2}s^2 = -\frac{\overline{X}s^2}{\sum(X_t - \overline{X})^2}$$

Билет 3. Оценка максимального правдоподобия.

Предположим, что мы ищем параметры нормальной линейной регрессионной модели:

$$Y_t = a + bX_t + \varepsilon_t$$

Ошибки регрессии ε_t независимы и распределены по нормальному закону: $\varepsilon_t \sim N(0, \sigma^2)$ или, что эквив. записью $Y_t \sim N(a + bX_t, \sigma^2)$

Пусть у нас есть набор наблюдений (X_t, Y_t), $t = \overline{1,n}$

Составим ф. правдоподобия:

$$L(Y_1 \dots Y_n, a, b, \sigma^2) = p(Y_1 \dots Y_n | X_1 \dots X_n, a, b, \sigma^2) =$$

$$= (2\pi)^{-n/2} (\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} \sum (Y_t - a - bX_t)^2\right\}$$

Для того, чтобы найти наиболее правдоподобные значения параметров: $L(Y_1 \dots Y_n, a, b, \sigma^2) \to max$.

Т.к. ф. L и $\ln L$ одновременно достиг. своего максимума достаточно искать максимум логарифма

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

$$-\frac{1}{2\sigma^2} \sum (Y_t - a - bX_t)^2$$

Необходимые условия экстремума ф. $\ln L$ имеют вид:

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

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

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

Решением этой системы будут оценки максимальн. правдоподобия

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

Линейная регрессионная модель с двумя переменными. Статист. св-ва МНК оценок параметров регрессии. Проверка гипотезы $H_0: B=B_0$. Доверительные интервалы для коэфф. регрессии.

Пусть $\varepsilon \sim N(0, \sigma^2 I_n)$, т.е. ε – многомерная норм. распред. случ. величина. (т.е. y_t имеют совмест. норм. распр.)

Тогда $\hat{a} \sim N\left(a, \sigma^2 \frac{\sum_{t=1}^{n} X_t^2}{n \sum_{t=1}^{n} x_t^2}\right)$, $\hat{b} \sim N\left(b, \sigma^2 \frac{1}{\sum_{t=1}^{n} x_t^2}\right)$ независ.

Это наз. гипотезой нормальности ошибок.

Если она не выполнена, то при $n \to \infty$ $\hat{a} \underset{n\to\infty}{\sim} N\{\cdot\}$, $\hat{b} \underset{n\to\infty}{\sim} N\{\cdot\}$

$x_t = X_t - \bar{X}$, $y_t = Y_t - \bar{Y}$;

$\hat{\sigma}^2 = s^2$; распределение оценки дисперсии ошибок s^2: $\frac{(n-2)s^2}{\sigma^2} \sim \chi^2(n-2)$

$\hat{\sigma}^2$ и $(\hat{a}, \hat{b})$ независимы: $\hat{\sigma}^2 = s^2$ – ф-ия от e_t (остатки регрессии)

$(\hat{a}, \hat{b})$ и e_t – лин. ф-ии ошибок ε_t => имеют совместное норм. распределение => e_t и $(\hat{a}, \hat{b})$ некоррелиров.

$\hat{b} = \frac{\sum x_t y_t}{\sum x_t^2} = \sum \omega_t y_t$, $\omega_t = \frac{x_t}{\sum_s x_s^2}$ – замена переменных

($\sum \omega_t = 0$; $\sum \omega_t x_t = \sum \omega_t X_t = 1$; $\sum \omega_t^2 = \frac{1}{\sum x_t^2}$; $\sum \omega_t y_t = \sum \omega_t Y_t$)

Пусть $\xi = \sum \omega_t \varepsilon_t$, тогда $e_t = \varepsilon_t - \bar{\varepsilon} - x_t \xi$, $\hat{b} = b + \xi$

$Cov(e_t, \hat{b}) = Cov(\varepsilon_t - \bar{\varepsilon} - x_t\xi, b + \xi) = E(\varepsilon_t \xi - \bar{\varepsilon}\xi - x_t \xi^2) =$

$= \sigma^2 \left(\omega_t - \frac{1}{n}\sum \omega_s - x_t \sum \omega_s^2\right) = \sigma^2\left(\omega_t - x_t \frac{1}{\sum x_s^2}\right) = 0$

Аналогично $\hat{a} = \sum\left(\frac{1}{n} - \bar{X}\omega_t\right) Y_t$ и $Cov(e_t, \hat{a}) = 0$.

Проверка гипотезы $H_0: b = b_0$.

$\hat{b} - b \sim N(0, \sigma_{\hat{b}}^2)$, где $\sigma_{\hat{b}}^2 = \frac{\sigma^2}{\sum x_t^2}$.

Оценка дисперсии оценки $\hat{b}$: $D(\hat{b}) = S^2_{\hat{b}} = \frac{S^2}{\sum x_i^2}$

Т.о. $\frac{\hat{b} - b}{\sigma_{\hat{b}}} \sim N(0,1)$, а значит $\frac{S}{\sigma} \sim \sqrt{\frac{1}{n-2}\chi^2(n-2)}$

т.е. по определению статистики Стьюдента:

$t = \frac{(\hat{b} - b)}{S_{\hat{b}}} \sim t(n-2)$ (Аналог. $t = \frac{\hat{a} - a}{S_{\hat{a}}} \sim t(n-2)$)

Эту статистику используют для проверки гипотезы.

Если $H_0: b = b_0$ верно, то $t = \frac{\hat{b} - b_0}{S_{\hat{b}}} \sim t(n-2)$

при $|t| \le t_\alpha(n-2)$ гипотеза H_0 принимается

Тогда $P\{-t_\alpha(n-2) < \frac{\hat{b} - b_0}{S_{\hat{b}}} < t_\alpha(n-2)\} = 1 - \alpha$

Отрезок $[\hat{b} - t_c S_{\hat{b}}; \hat{b} + t_c S_{\hat{b}}]$ наз. доверит. интервал.

При $H_0: b = 0 \Rightarrow t_c = \frac{\hat{b}}{S_{\hat{b}}}$.

первого ... стационарности процесса авторегр

Билет 5. Линейная регрессионная модель с двумя переменными. Анализ вариации зависимой переменной в регрессии (TSS, ESS, RSS). Коэффициент детерминации R^2.

51 - 53

Рассмотрим вариацию (разброс) $\sum(Y_t - \bar{Y})^2$ значений Y_t вокруг среднего значения. Разобьем вариацию на две части.

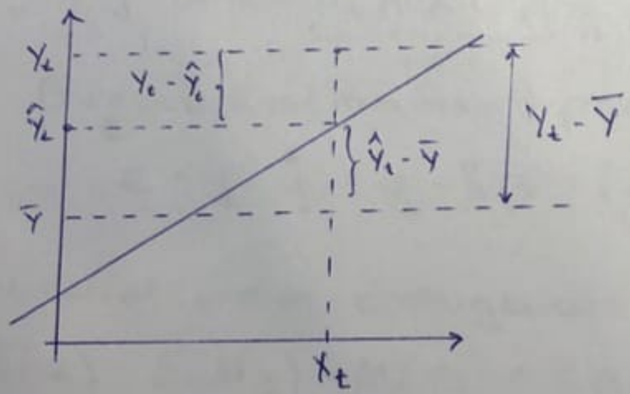
Обозначим через $\hat{Y}_t = \hat{a} + \hat{b}X_t$ предсказанное значение Y_t, тогда $Y_t - \bar{Y} = (Y_t - \hat{Y}_t) + (\hat{Y}_t - \bar{Y})$ и вариация Y_t представляется в виде:

$$\underbrace{\sum(Y_t - \bar{Y})^2}_{TSS} = \underbrace{\sum(Y_t - \hat{Y}_t)^2}_{ESS} + \underbrace{\sum(\hat{Y}_t - \bar{Y})^2}_{RSS} + 2\underbrace{\sum(Y_t - \hat{Y}_t)(\hat{Y}_t - \bar{Y})}_{=0}$$

Третье слагаемое $=0$, т.к. $y - \hat{y} = e$ - вектор остатков регрессии ортогонален конст e и вектору x.

В самом деле,

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



Статистика R^2 - коэф. детерминации

<u>Опр.</u> Коэффициентом детерминации, или долей объясненной дисперсии, называется

$$R^2 = 1 - \frac{ESS}{TSS} = \frac{RSS}{TSS}, \quad R^2 \in [0, 1]$$

$R^2 = 0 \Rightarrow$ регрессия ничего не дает

$R^2 = 1 \Rightarrow$ точная подгонка

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

Модель множественной регрессии. Метод наименьших квадратов. Теорема Гаусса-Маркова (без доказательств).

$$y_t = \beta_1 + \beta_2 x_{t2} + \ldots + \beta_k x_{tk} + \varepsilon_t,\ t = \overline{1,n} \quad \text{или}$$

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \ldots + \beta_k x_{tk} + \varepsilon_t,\ t = \overline{1,n} \quad \text{(без свободн. члена)}$$

x_{tp} – значение регрессора x_p в наблюдении t, $x_{t1} = 1$, $t = \overline{1,n}$.

Основные гипотезы:

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

2. $x_{t1}, \ldots, x_{tk}$ – детерминир. величины. Векторы $x_s = \begin{bmatrix} x_{1s} \\ \vdots \\ x_{ns} \end{bmatrix}$, $s = \overline{1,k}$ линейно независимы в $\mathbb{R}^n$;

3. $E\varepsilon_t = 0$; $E(\varepsilon_t^2) = V(\varepsilon_t) = \sigma^2$ – не зав. от t; $E(\varepsilon_t \varepsilon_s) = 0$ при $t \neq s$ (некоррел. ошибок);

3'. $\varepsilon_t \sim N(0, \sigma^2)$, $t = \overline{1,n}$;

Матричная форма:

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix},\ \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix},\ \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix},\ X = \begin{bmatrix} x_{11} & \ldots & x_{1k} \\ \ldots & \ldots & \ldots \\ x_{n1} & \ldots & x_{nk} \end{bmatrix}$$

Тогда: 1) $y = X\beta + \varepsilon$;

2) X – детерминир. матрица, $\operatorname{rank}(X) = k$;

3) $E(\varepsilon) = 0$; $V(\varepsilon) = E(\varepsilon\varepsilon^T) = \sigma^2 I_n$;

3') $\varepsilon \sim (0, \sigma^2 I_n)$, т.е. ε – норм распред. случ. вектор. с матр. ковариаций $\sigma^2 I_n$.

МНК: вектор оценок $\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}$ миним. $\sum e_t^2$ (т.е. квадрат длины вектора остатков $e = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}$):

$e = y - \hat{y} = y - X\hat{\beta}$, $ESS = \sum e_t^2 = e^T e \to \min$

Выразим $e^T e$ и/з X и β:

$e^T e = (y - X\hat{\beta})^T (y - X\hat{\beta}) = \underline{y^T y} - \underline{\underline{y^T X \hat{\beta}}} - \underline{\underline{\hat{\beta}^T X^T y}} + \underline{\hat{\beta}^T X^T X \hat{\beta}} =$

$= y^T y - 2\hat{\beta}^T X^T y + \hat{\beta}^T X^T X \hat{\beta}$

$\frac{\partial ESS}{\partial \hat{\beta}} = -2X^T y + 2X^T X \hat{\beta} = 0$ – необх. усл-не мин. ESS.

В силу обратимости матрицы $X^T X$, $2X^T X \hat{\beta} = 2X^T y$:

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

Необх усл-не мин. ESS означает, что вектор e ортогон. $x_1, \dots, x_k$, т.е.

$x_1^T e = \dots = x_k^T e = 0$ или $X^T e = 0$.

$X^T e = X^T (y - X\hat{\beta}) = X^T y - X^T X \hat{\beta} = X^T y - X^T X (X^T X)^{-1} X^T y = 0$

$e^T e = y^T y - 2\hat{\beta}^T X^T y + \hat{\beta}^T X^T X \hat{\beta} = y^T y - \hat{\beta}^T (2X^T y - X^T X (X^T X)^{-1} \cdot$

$\cdot X^T y) = y^T y - \hat{\beta}^T X^T y$.

Теорема Гаусса – Маркова.

Пусть выполнены 1) – 3), тогда оценка по МНК $\hat{\beta} = (X^T X)^{-1} X^T y$ явл. эффективной (т.е. наименьш. дисперсия и она не смещённая).

Билет 7. Модель множественной регрессии. статистические св-ва МНК-оценок ~~параметров~~ оценка дисперсии ошибок σ^2. Распределение s^2. Независимость оценок $\hat\beta$ и s^2. 67, 72-74

Модель множественной регрессии:

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

или

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

где x_{tp} - значения регрессора x_p в наблюдении t

Матричный вид: $y = X\beta + \varepsilon$

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}, \quad X = \begin{bmatrix} x_{11} & \dots & x_{1k} \\ \vdots & & \vdots \\ x_{n1} & \dots & x_{nk} \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Введем обозначения:

Вектор прогнозных значений:

$$\hat y = X\hat\beta = \underbrace{X(X^TX)^{-1}X^T}_{=N} y = Ny$$

Вектор остатков регрессии:

$$e = y - \hat y = y - X\beta = \underbrace{(I - X(X^TX)^{-1}X^T)}_{=M} y$$

Вычислим мат. ожидание и матрицу коварицаций e:

$$E(e) = E(My) = MX\beta = X\beta - X\beta = 0$$

$$Var(e) = Var(My) = M Var(y) M^T = \sigma^2 M$$

Сумма квадратов остатков $\sum e_i^2 = e^Te$ явл. естественным кандидатом на оценку дисперсии ошибок σ^2:

$$E(e^Te) = tr(Var(e)) = \sigma^2 tr(M) = (n-k)\sigma^2$$

$$s^2 = \hat\sigma^2 = \frac{e^Te}{n-k} = \frac{\sum e_t^2}{n-k}$$ - несмещенная оценка дисперсии ошибок σ^2, (т.е. $Es^2 = \sigma^2$)

Т.к. $NX = X$, $MX = 0$, то $e = My = M(X\beta + \varepsilon) = M\varepsilon$ (1)

и $\operatorname{rank}(M) = \operatorname{rank}(I-N) = \operatorname{tr}(I-N) = n-k$, то распределение

$$\frac{e^Te}{\sigma^2} \sim \chi^2(n-k) \quad \text{или} \quad (n-k)\frac{s^2}{\sigma^2} \sim \chi^2(n-k)$$

В предположении нормальной линейной множественной регрессионной модели (т.е $\varepsilon \sim N(0,\sigma^2 I)$) удается док-ть независимость оценок $\hat\beta$ и s^2.

В самом деле, из оценки МНК: $\hat\beta_{OLS} = (X^TX)^{-1}X^Ty$ получаем

$$\hat\beta_{OLS} = (X^TX)^{-1}X^T(X\beta+\varepsilon) = \beta + (X^TX)^{-1}X^T\varepsilon = \beta + A\varepsilon \quad (2)$$

Из (1) и (2) видно, что случайные векторы $\hat\beta$ и e имеют совместное ~~нор~~ многомерное нормальное распределение. Поэтому для того, чтобы показать их независимость, достаточно показать их некоррелированность.

$$AM = 0 \Rightarrow \operatorname{cov}(\hat\beta, e) = E((\hat\beta - \beta)e') = \sigma^2 AM = 0$$

Т.к. s^2 явл. функцией от e, то оценки $\hat\beta$ и s^2 также не зависимы

Модель множественной регрессии. Анализ вариации зависимой переменной в регрессии (TSS, ESS, RSS). Скорректированный коэффициент детерминации.

Посчитаем вариацию $\sum(y_t-\bar{y})^2$:

$$\sum(y_t-\bar{y})^2=\sum(y_t-\hat{y}_t)^2+\sum(\hat{y}_t-\bar{y})^2+2\sum(y_t-\hat{y}_t)(\hat{y}_t-\bar{y})$$

или в векторной форме:

$$(y-\bar{y}i)^T(y-\bar{y}i)=(y-\hat{y})^T(y-\hat{y})+(\hat{y}-\bar{y}i)^T(\hat{y}-\bar{y}i)+$$

$$+2(y-\hat{y})^T(\hat{y}-\bar{y}i)$$

0, если константа, т.е. $i \in$ лин. оболочке векторов $x_1,\dots,x_k$:

$$(y-\hat{y})^T(\hat{y}-\bar{y}i)=e^T(X\hat{\beta}-\bar{y}i)=e^TX\hat{\beta}-\bar{y}e^Ti=0,$$ т.к.

$e^TX=0$ и $\bar{e}=\frac{e^Ti}{n}=0$. Поэтому верно равенство:

$$\underbrace{\|y-\bar{y}i\|^2}_{TSS}=\underbrace{\|y-\hat{y}\|^2}_{ESS}+\underbrace{\|\hat{y}-\bar{y}i\|^2}_{RSS}$$

В отклонениях $y_*=y-\bar{y}i$, $\hat{y}_*=\hat{y}-\bar{y}i$: $y_*^Ty_*=e^Te+\hat{y}_*^T\hat{y}_*$

Коэфф. детерминации $R^2=1-\frac{ESS}{TSS}=1-\frac{e^Te}{y_*^Ty_*}=\frac{\hat{y}_*^T\hat{y}_*}{y_*^Ty_*}=\frac{RSS}{TSS}$

показывает качество подгонки регрессион. модели к наблюд. значен. y_t.

Свойства:

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

2). $R^2\uparrow$ при росте кол-ва регрессоров;

3). чувствителен при изменении данных;

Чтобы устранить 2) R^2 корректируют на число регрессоров.

$$R^2_{adj}=1-\frac{e^Te/(n-k)}{y_*^Ty_*/(n-1)}$$

Свойства:

1). $R^2_{adj}=1-(1-R^2)\frac{n-1}{n-k}$

2). $R^2 \geqslant R^2_{adj}$, $k>1$

3). $R^2_{adj} \leq 1$, но может быть <0.

Модель 78-85

Билет 9. Модель множественной регрессии
Проверка статистических гипотез.

Провека гипотезы $H_0: \beta_i = \beta_{i0}$

Доказано, что:

1. Вектор оценок $\hat{\beta}_{OLS}$ имеет нормальное распред. со средним β и матрицей ковариаций $Var(\hat{\beta}_{OLS}) = \sigma^2 (X^T X)^{-1}$. т.е. $\hat{\beta}_{OLS} - \beta \sim N(0, \sigma^2 (X^T X)^{-1})$ или $\hat{\beta}_{OLS,i} - \beta_i \sim N(0, \sigma^2_{\hat{\beta}_i})$, где $\sigma^2_{\hat{\beta}_i} = \sigma^2 q^{ii}$; q^{ii} - i-ый диаг. элемент матрицы $(X^T X)^{-1}$

   В качестве оценки дисперсии возьмем
   $$s^2_{\hat{\beta}_i} = \hat{\sigma}^2_{\hat{\beta}_i} = \hat{\sigma}^2 q^{ii} = s^2 q^{ii}$$
2. Сл. вел. $(n-k)\frac{s^2}{\sigma^2}$ распред. по закону хи-квадрат с $n-k$ степенями свободы $\chi^2(n-k)$
3. Оценки $\hat{\beta}_{OLS}$ и s^2 независимы

Отсюда получаем,
$$t = \frac{\hat{\beta}_{OLS,i} - \beta_i}{s_{\hat{\beta}_i}} = \frac{(\hat{\beta}_{OLS,i} - \beta_i)/\sigma_{\hat{\beta}_i}}{s_{\hat{\beta}_i}/\sigma_{\hat{\beta}_i}} \sim t(n-k) \qquad (*)$$

Из $(*)$ $\Rightarrow$ $[\hat{\beta}_{OLS,i} - t_c s_{\hat{\beta}_i}; \hat{\beta}_{OLS,i} + t_c s_{\hat{\beta}_i}]$ явл. $100(1-\alpha)\%$-ным доверительным интервалом для истинного значения коэф. β_i, где $t_c = t_{\alpha/2}(n-k)$ - $100(\alpha/2)\%$-ная точка распределения Стьюдента с $n-k$ степенями свободы.

Для тестирования гипотезы $H_0: \beta_i = \beta_{i0}$, также можно применить статистику $(*)$, а именно, нулевая гипотеза отклоняется на уровне значимости α, если

$$|t| = \left| \frac{\hat{\beta}_{OLS,i} - \beta_{i0}}{s_{\hat{\beta}_i}} \right| > t_{\alpha/2}(n-k)$$

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

Предполагаем, что в число регрессоров включена константа: $y_t = \beta_1 + \beta_2 x_{t2} + \ldots + \beta_k x_{tk} + e_t$.

Рассмотрим статистику:

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-k}{k-1} = \frac{RSS}{ESS} \cdot \frac{n-k}{k-1} = \frac{\sum(\hat{y}_t - \bar{y})^2/(k-1)}{\sum e_t^2/(n-k)} =$$

$$= \frac{\frac{\hat{y}_*^T \hat{y}_*}{\sigma^2} \cdot \frac{1}{k-1}}{\frac{e^T e}{\sigma^2} \cdot \frac{1}{n-k}}$$

Знаменатель имеет распред: $\frac{1}{n-k}\chi^2(n-k)$

Числитель имеет распред: $\frac{1}{k-1}\chi^2(k-1)$

Т.к. $\hat{\beta}_{OLS}$ и e независимы, то статистика F имеет распред. ~~фун~~ Фишера.

$$F \sim F(k-1, n-k).$$

Проверка гипотезы $H_0: H\beta = r$

Пусть H - $q \times k$ матрица, β - $k \times 1$ вектор, r - $q \times 1$ вектор. $\operatorname{rank}(H) = q$ и $q \leq k$

Вектор $\hat{\beta}_{OLS}$ имеет нормальное распределение со средним β и матрицей ковариаций $\sigma^2(X^T X)^{-1}$.

Отсюда получаем, что $H\hat{\beta} - r \sim N(H\beta - r, \Sigma)$, где Σ - $q \times q$ матрица и $\Sigma = Var(H\hat{\beta} - r) = \sigma^2 H (X^T X)^{-1} H^T$

Получаем, что

$$\frac{1}{\sigma^2}(H\hat{\beta} - r)^T (H(X^T X)^{-1} H^T)^{-1} (H\hat{\beta} - r) \sim \chi^2(q)$$

Используя независимость $\hat{\beta}$ и e получаем:

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

$$F = \frac{(\hat{\beta} - \beta)^T H^T (H(X^T X)^{-1} H^T)^{-1} H(\hat{\beta} - \beta)/q}{e^T e/(n-k)} \sim F(q, n-k) \quad (**)$$

Условие $F < F_\alpha(q, n-k)$ задает $100(1-\alpha)\%$-ную доверит. область для коэффициентов β.

...стационарности процесса авторегр... первого порядка AR(1)

Билет 9.

В случае $H = I$:

$$F = \frac{(\hat\beta - \beta)^T (X^T X)(\hat\beta - \beta)/k}{e^T e/(n-k)} \sim F(k, n-k)$$

Проверка гипотезы: $H_0: \beta_{k-q+1} = \beta_{k-q+2} = \ldots = \beta_k = 0$.

Разобьем все матрицы с размером по одной из сторон k на блоки со сторонами $(k-q)$ и q:

$$H = [0 \; I_q], \quad \hat\beta = \begin{bmatrix} \hat\beta_1 \\ \hat\beta_2 \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \quad X = [X_1 \; X_2]$$

$$H\hat\beta = [0 \; I_q]\begin{bmatrix} \hat\beta_1 \\ \hat\beta_2 \end{bmatrix} = \hat\beta_2, \quad X\hat\beta = X_1\hat\beta_1 + X_2\hat\beta_2$$

здесь X_1 - $n\times(k-q)$, X_2 - $n\times q$, $\beta_1, \hat\beta_1$ - $(k-q)\times 1$, $\beta_2, \hat\beta_2$ - $q\times 1$

Введем обозначения:

$$Q = X^T X = \begin{bmatrix} X_1^T \\ X_2^T \end{bmatrix}[X_1 \; X_2] = \begin{bmatrix} X_1^T X_1 & X_1^T X_2 \\ X_2^T X_1 & X_2^T X_2 \end{bmatrix} = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix}$$

$$Q^{-1} = \begin{bmatrix} Q^{11} & Q^{12} \\ Q^{21} & Q^{22} \end{bmatrix}$$

В этих обозначениях числитель дроби в (**) для F при условии, что верна нулевая гипотеза $H_0: \beta_2 = 0$ имеет вид (с точностью до множителя $1/q$):

$$\hat\beta_2^T X_2^T M_1 X_2 \hat\beta_2,$$

где M_1 - матрица ортогонального проектирования на $\Pi_1^\perp$ - ортогональное дополнение к подпр-ву Π_1 в $\mathbb{R}^n$

Т.к. $(e_*^T e_* - e^T e) = \hat\beta_2^T X_2^T M_1 X_2 \hat\beta_2$, то

$$F = \frac{(e_*^T e_* - e^T e)/q}{e^T e/(n-k)} = \frac{(ESS_R - ESS_{UR})/q}{ESS_{UR}/(n-k)} \sim F(q, n-k)$$

ESS_R - сумма квадратов остатков "короткой" регрессии

ESS_{UR} - сумма квадратов остатков "длинной" регрессии

Различные аспекты множественной регрессии. Тест Чоу. Мультиколлинеарность.

Тест Чоу – проверка гипотезы H_0.

$H_0: \beta' = \beta''$; $\sigma' = \sigma''$.

т.е. проверка на совпадение двух регрессионных моделей.

$y_t = \beta'_1 x_{t1} + \beta'_2 x_{t2} + \ldots + \beta'_k x_{tk} + \varepsilon'_t$, $t = \overline{1, n}$ (n наблюдений)

$y_t = \beta''_1 x_{t1} + \beta''_2 x_{t2} + \ldots + \beta''_k x_{tk} + \varepsilon''_t$, $t = \overline{n+1, n+m}$ (m наблюдений)

Длинная (без ограничений) регрессия (объединение двух):

$ESS_{UR} = ESS_1 + ESS_2$

число степеней свободы: $(n-k) + (m-k) = n+m-2k$

Пусть H_0 верна, тогда регрессия с ограничениями:

$y_t = \beta_1 x_{t1} + \ldots + \beta_k x_{tk} + \varepsilon_t$, $t = \overline{1, n+m}$

$\updownarrow ESS_R$

Учитывая, что наложено k ограничений:

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

Если эта F-статистика больше критического значения $F_c = F_\alpha(k, n+m-2k)$, то H_0 отвергается на уровне значимости α (т.е. не смогли две выборки объединить в одну).

Мультиколлинеарность.

Одно из условий: лин. незав. столбцов X или $\mathrm{rank}(X^T X)^{-1} = k$.

т.е. не можем строить МНК оценки.
если вырождена => исключаем один столбец.

если невырождена:

Пример $C=\beta_1+\beta_2 S+\beta_3 N+\beta_4 T+\varepsilon$, где C — потребление,

S — зарплата, N — доход вне работы, T — полный доход $(S+N)$.

Тогда $\forall h: C=\beta_1+\beta_2' S+\beta_3' N+\beta_4' T+\varepsilon$, где

$\beta_2'=\beta_2+h,\ \beta_3'=\beta_3+h,\ \beta_4'=\beta_4-h$ $\quad(\beta_1,\beta_2',\beta_3',\beta_4')$

Т.о. одни и те же наблюдения можно объяснить различными β.

Учитывая $T=S+N$: $C=\beta_1+(\beta_2+\beta_4)S+(\beta_3+\beta_4)N+\varepsilon$.
$(\beta_1,\beta_2+\beta_4,\beta_3+\beta_4)$

В общем случае если $\text{rank}(X^TX)^{-1}=l<k$, то только l лин. комбинаций можно оценить

Признаки:
1. Небольшое изменение данных приводит к существенному изменению коэффициентов;
2. Дисперсии оценок высоки ($\sigma^2_{\beta_k}$); t_{β_k} статистики малы. => β_k незначимы; сама модель значима (F-стат и R^2).
3. Оценки коэфф. β_k имеют неправильные значения или знаки (т.е. не соотв. ожиданиям).

Как бороться? Простые методы
1. Уменьшить дисперсию модели (искать то, что ещё не включили в модель)
2. Увеличить объём выборки n.
3. Увеличить дисперсию объясн. перемен. (регрессоров) $\sigma^2_{\beta_k}$
4. Собрать данные, которые м/у собой не очень связаны (не коррелир. => снизить корреляцию).

5. Объединить регрессоры;
6. Исключить некоторые незначимые регрессоры;
7. Использовать внешнюю информацию.

...ционарности процесса авторег... первого порядка AR(1)

Билет 11. Различные аспекты множественной регрессии. Коэф. частной корреляции. Процедура пошагового отбора переменных.

В том случае, когда имеются одна независимая и одна зависимая переменные, естественной мерой зависимости (в рамках линейного подхода) явл. (выборочный) коэффициент корреляции между ними. Использование множественной регрессии позволяет обобщить это понятие на случай, когда имеется несколько независимых переменных.

Предположим, что $y = \alpha + x_1\beta_1 + x_2\beta_2 + \varepsilon$, где

y - $n \times 1$ вектор наблюдений зависимой переменной

x_1, x_2 - $n \times 1$ векторы независимых переменных

α, β_1, β_2 - (скалярные) параметры

ε - $n \times 1$ вектор ошибок.

Наша цель: Определить корреляцию между y и первым регрессором x_1 после исключения влияния x_2.

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. — " — x_1 на x_2 — " — $\hat{x}_1 = \hat{\gamma}_1 + \hat{\gamma}_2 x_2$
3. Удалим влияние x_2, взяв остатки $e_y = y - \hat{y}$ и $e_{x_1} = x_1 - \hat{x}_1$
4. Определим (выборочный) коэф. частной корреляции между y и x_1 при искл. влияния x_2 как (выборочный) коэф. корреляции между e_y и e_{x_1}:

$$r(y, x_1 | x_2) = r(e_y, e_{x_1})$$

Из св-в МНК $\Rightarrow$ e_y и e_{x_1} не коррелированы с x_2

связь коэф. частной и обычной корреляции:

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

$$r(y, x_1 | x_2) \in [-1, 1]$$

$r(y, x_1 | x_2) = 0$, означает, говоря нестрого, отсутствие прямого (линейного) влияния переменной x_1 на y.

Также

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

или

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

Описанная выше процедура обобщается на случай когда исключается влияние не одной, а нескольких переменных: достаточно переменную x_2 заменить на набор переменных X_2.

Процедура пошагового отбора переменных.

Основными пошаговыми процедурами яв. процедура последовательного присоединения, процедура присоединения-удаления и процедура последовательного удаления

Процедура присоединения-удаления:

На первом шаге из исходного набора объясняющих переменных выбирается переменная, имеющая наибольший по модулю коэф. корреляции с зависимой переменной y.

Второй шаг состоит из двух шагов. 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 коэф. детерминации. Затем сравнивается значение F-статистики для проверки гипотезы H_0 о незначимости этого регрессора с некоторым заранее заданным пороговым значением $F_{искл}$. Если $F < F_{искл}$, 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 коэф. корреляции и сравниваем значение F-статистики для проверки гипотезы H_0 о незнач. этого регрессора с некоторым заранее заданным пороговым значением $F_{вкл}$. Если $F > F_{вкл}$, то x_p включается в список регрессоров. Обычно выбирают $F_{искл} < F_{вкл}$. Второй шаг повторяется до тех пор пока происходит измен. списка регрессоров.

Стационарные ряды. Модели ARMA. Понятие временного ряда. Понятие строго стационарного процесса. Три условия для стационарности в широком смысле. Процесс белого шума.

Временной ряд — послед-ть наблюдений некоторой переменной, которое производится ч/з равные промежутки времени.

Выборка $X_1, .., X_n$ — случ. вел., $x_1, .., x_n$ — наблюдения.

Всё построено на ф-ии распред. $F(\vartheta_1, .., \vartheta_n) = P\{X_1 < \vartheta_1, X_2 < \vartheta_2, .., X_n < \vartheta_n\}$

Распредел. случ. вел. $X_1, .., X_n$ имеет совместную плот-ть распр. $p(x_1, .., x_n)$.

• Ряд x_t, $t = \overline{1,n}$ наз. строго стационарным (в узком смысле)

если $\forall m: m < n$ совместн. распред. вер-ей случ. вел. $X_{t_1}, .., X_{t_m}$

такое же как и для $X_{t_1+\tau}, .., X_{t_m+\tau}$, $\forall t_1, .., t_m, \tau: t_1 \le t_2 \le .. \le t_m \le n$,

$1 \le t_1 + \tau \le t_2 + \tau \le .. \le t_m + \tau \le n$. $p(x_{t_1}, .., x_{t_m}) = p(x_{t_1+\tau}, .., x_{t_m+\tau})$.

Св-ва строго стац. ряда не изменяются при изменении начала отсчета.

$m = 1 \Rightarrow X_{t_1}$ и $E(X_{t_1}) = \mu$, $Var(X_{t_1}) = \sigma^2$ не зав. от времени

Теснота связи наблюдений $Corr(X_t, X_{t+\tau}) = \dfrac{cov(X_t, X_{t+\tau})}{\sqrt{Var(X_t)}\sqrt{Var(X_{t+\tau})}}$,

где $cov(X_t, X_{t+\tau}) = E[(X_t - E(X_t))(X_{t+\tau} - E(X_{t+\tau}))] = \gamma(\tau)$ не зав. от t.

$cov(X_t, X_t) = Var(X_t) = \gamma(0)$

т.о. $Corr(X_t, X_{t+\tau}) = \dfrac{\gamma(\tau)}{\gamma(0)} = \rho(\tau)$, $\rho(0) = 1$.

• Ряд x_t, $t = \overline{1,n}$ наз. слабо стационарным (в широком смысле), если его числ. харак-ки не зависят от времени:

1. $E(X_t) = \mu$; 2. $Var(X_t) = \sigma^2$; 3. $Cov(X_t, X_{t+\tau}) = \gamma(\tau)$ $\forall t, \tau$

Слабая и строгая стац. не следуют друг из друга.

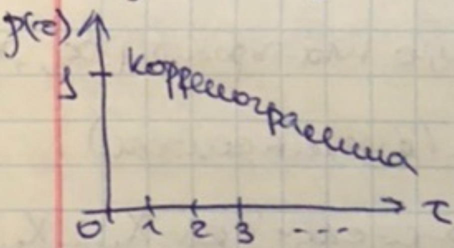
Ряд x_t, $t=\overline{1,n}$ наз. Гауссовским, если совместное распр. случ. вел. $X_1, .., X_n$ явл. n-мерным норм. распр-ем: $x_t \sim N(\mu, 6^2)$

$\rho(\tau)$ — ф-ция автокорреляции (безразмерна):

1. $[-1; 1]$ — интервал автокоррел. ф-ции

2. $\rho(0) = 1$

3. $\rho(-\tau) = \rho(\tau)$ — симметричная



статистич. диаграмма для x_t и $x_{t+\tau}$ совпадают

Полное описание совместн. распр. случ. величин $X_1, .., X_n$ это n+1 параметр: $\mu, \gamma(0), \gamma(1), .., \gamma(n-1)$ $(\mu, \rho(0), \rho(1), .., \rho(n-1))$

• Белый шум — стац. врем. ряд ε_t:

1. $E(\varepsilon_t) = 0$; 2. $Var(\varepsilon_t) = 6^2 > 0$; 3. $\rho(\tau) = 0, \forall \tau \neq 0$

т.е. X_t и X_s при $t \neq s$ некоррелир.

Как модель не подходит, как база, которая создает случ. компоненту подходит.

Гауссовский белый шум; ε_t взаимно независ. и имеют нормал. одинак. распред. $N(0, 6^2)$.

Билет 13. Процесс авторегрессии первого порядка AR(1). Условия стационарности процесса авторегрессии первого порядка AR(1).

~~Про~~ Модель авторегрессии первого порядка AR(1):

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

где ε_t - белый шум, имеющий $E(\varepsilon_t) = 0$ и $Var(\varepsilon_t) = \sigma_\varepsilon^2$

X_0 - некоторая сл. вел.

$a \neq 0$ - коэф.

При этом $E(X_t) = aE(X_{t-1})$ так что рассматриваемый процесс может быть стационарным только если $E(X_t) = 0 \; \forall t = \overline{1,n}$.

Далее,

$X_t = aX_{t-1} + \varepsilon_t = a^t X_0 + a^{t-1}\varepsilon_1 + a^{t-2}\varepsilon_2 + \ldots + \varepsilon_t$

$X_{t-1} = aX_{t-2} + \varepsilon_{t-1} = a^{t-1}X_0 + a^{t-2}\varepsilon_2 + \ldots + \varepsilon_{t-1}$

...

$X_1 = aX_0 + \varepsilon_1$

Если X_0 не коррелирована со $\varepsilon_1 \ldots \varepsilon_n$, то

$cov(X_0, \varepsilon_1) = 0 \;\ldots\; cov(X_{t-1}, \varepsilon_t) = 0$ и

$Var(X_t) = Var(aX_{t-1} + \varepsilon_t) = a^2 Var(X_{t-1}) + Var(\varepsilon_t) =$

$= \{Var(X_0) = Var(X_t) = \sigma_X^2\} \Rightarrow \sigma_X^2 = a^2\sigma_X^2 + \sigma_\varepsilon^2 \Rightarrow \sigma_X^2 = \frac{\sigma_\varepsilon^2}{1-a^2}, \; |a| < 1$

Что касается автоковариаций и автокорреляций, то

$cov(X_t, X_{t+\tau}) = a^\tau \sigma_\varepsilon^2 \frac{1}{1-a^2} \qquad corr(X_t, X_{t+\tau}) = a^\tau$

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

__Опр.__ $X_t = aX_{t-1} + \varepsilon_t, \; t = \overline{1,n}$ порождает стац. временной ряд, если

1) $|a| < 1$
2) X_0 некоррелир. с $\varepsilon_1 \ldots \varepsilon_n$
3) $E(X_0) = 0$
4) $Var(X_0) = \frac{\sigma_\varepsilon^2}{1-a^2}$

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

Рассмотренная модель порождает стационарный ряд, имеющий нулевое мат. ожидание. Однако ее можно легко распространить и на временные ряды Y_t с ненулевым мат. ожиданием $E(Y_t)=\mu$, полагая, что указанная модель относится к центрированному ряду $X_t = Y_t - \mu$:

$$Y_t - \mu = a(Y_{t-1} - \mu) + \varepsilon_t, \quad t=\overline{1,n}$$

так, что

$$Y_t = aY_{t-1} + \delta + \varepsilon_t, \quad t=\overline{1,n}$$

где $\delta = \mu(1-a)$

Процесс авторегрессии порядка p AR(p). Оператор запаздывания.
Условие стационарности процесса авторегрессии порядка p AR(p).

AR(p): $X_t = a_1 X_{t-1} + a_2 X_{t-2} + \ldots + a_p X_{t-p} + \varepsilon_t$, $a_p \neq 0$, ε_t - белый шум

$\varepsilon_t \sim \sigma_\varepsilon^2$ и $Cov(X_{t-s}, \varepsilon_t) = 0 \ \forall s > 0$.

Оператор запаздывания (лага) L:

$L X_t = X_{t-1}$, если применяется p раз то: $L^p X_t = X_{t-p}$

Тогда модель будет $(a_1 X_{t-1} + \ldots + a_p X_{t-p} = (a_1 L + \ldots + a_p L^p) X_t$

Обозначим $a(L) = 1 - (a_1 L + \ldots + a_p L^p)$ и получим AR(p):

$a(L) X_t = \varepsilon_t$

Чтобы процесс был стационар. все корни алг. ур-ия

$a(z) = 0$ должны лежать вне ед. круга $|z| \leq 1$.

Решение ур-ия $a(L) X_t = \varepsilon_t$:

$X_t = \frac{1}{a(L)} \varepsilon_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j}$, где $\sum_{j=0}^{\infty} |b_j| < \infty$

Откуда следует $E(X_t) = E\left(\sum_{j=0}^{\infty} b_j \varepsilon_{t-j}\right) = \sum_{j=0}^{\infty} b_j E(\varepsilon_{t-j}) = 0$.

AR(p) с ненулевым мат. ожиданием μ:

$a(L)(X_t - \mu) = \varepsilon_t$ или $a(L) X_t = \delta + \varepsilon_t$, где

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

Решение ур-ия $a(L)(X_t - \mu) = \varepsilon_t$: $X_t = \mu + \frac{1}{a(L)} \varepsilon_t$.

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

Чтобы найти коэфф. a_j, $j = \overline{1,p}$ используем систему

ур-ий Юла-Уокера:

$$X_t = a_1 X_{t-1} + \dots + a_p \cdot X_{t-p} + \varepsilon_t \quad \Big| \cdot X_{t-k},\ k>0$$

$$X_t \cdot X_{t-k} = a_1 X_{t-1} \cdot X_{t-k} + \dots + a_p \cdot X_{t-p} \cdot X_{t-k} + \varepsilon_t$$

теперь возьмём от обеих частей мат. ожидание:

$$\gamma(k) = a_1 \gamma(k-1) + a_2 \gamma(k-2) + \dots + a_p \gamma(k-p),\ k>0 \ \Big| : \gamma(0)$$

и получим систему ур-ий Юла – Уокера:

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

Зная, что $\rho(k) = \rho(-k)$ и $\rho(0) = 1$, то можем найти $a_j,\ j=\overline{1,p}$.

Билет 15. Процесс скользящего среднего порядка q MA(q)

Условие стационарности процесса скользящего среднего первого порядка MA(1).

Процесс MA(q):

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

где ε_t - белый шум

У такого процесса $E(X_t) = 0$.

Модель можно обобщить до процесса с ненулевым мат. ожиданием μ, полагая

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

т.е.

$$X_t = \mu + \varepsilon_t + b_1\varepsilon_{t-1} + \ldots + b_q\varepsilon_{t-q}$$

При $q = 0$ и $\mu = 0$ $\Rightarrow$ процесс белого шума

При $q = 1$ $\Rightarrow$ $X_t - \mu = \varepsilon_t + b\,\varepsilon_{t-1}$ $\leftarrow$ MA(1)

Для MA(1): $Var(X_t) = (1 + b^2)\sigma_\varepsilon^2 = \gamma(0)$

$\cancel{E(X_t) = \mu}$

Процесс X_t явл. стационарным с $E(X_t) = 0$, $Var(X_t) = (1+b^2)\sigma_\varepsilon^2$

$$\begin{aligned} \gamma(1) &= b_1\sigma_\varepsilon^2 \\ \gamma(k) &= 0, \ k > 1 \end{aligned} \quad \Rightarrow \quad \gamma(k) = \begin{cases} (1+b^2)\sigma_\varepsilon^2, & k = 0 \\ b\sigma_\varepsilon^2, & k = 1 \\ 0, & k > 1 \end{cases}$$

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 этого 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 $\forall$ MA(1): $|\rho(1)| \le 0{,}5$, т.е. корреляционная связь между соседними наблюдениями невелика, тогда как у процесса AR(1) такая связь может быть сколь угодно сильной (при $|a|$ близких к 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 нее

$$\gamma(k) = \begin{cases} \left(\sum_{j=0}^{q-k} b_j b_{j+k}\right)\sigma_\varepsilon^2, & 0 \le k \le q \\ 0, & k>q \end{cases}$$

Так что MA(q) явл. стационарным процессом с нулевым мат. ожиданием и дисперсией

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

и автокорреляциями

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

Здесь статистическая связь между наблюдениями сохраняется в течении q единиц времени.

Обратимость AR(p) и MA(q): AR(p) $\to$ MA(∞)

Если влияние прошлых событий ослабевает с течением времени показательным образом: $b_j = a^j$, $0<a<\frac{1}{2}$, то искусственное предположение о том, что ряд ε_t начинается в "бесконечном прошлом", приводит к модели MA(∞):

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

Модели авторегрессии-скользящего среднего ARMA(p,q). Условие стационарности модели авторегрессии-скол. средн. ARMA(p,q).

ARMA(p,q): $X_t = \sum_{j=1}^{p} a_j X_{t-j} + \sum_{j=0}^{q} b_j \varepsilon_{t-j}$, $a_p \neq 0, b_q \neq 0$

ε_t – белый шум, $b_0 = 1$.

С ненул. мат. ожид. μ: $X_t - \mu = \sum_{j=1}^{p} a_j (X_{t-j} - \mu) + \sum_{j=0}^{q} b_j \varepsilon_{t-j}$

Оператор лага: $LX_t = X_{t-1}$, $L^p X_t = X_{t-p}$

Обозначим $a(L) = 1 - a_1 L - a_2 L^2 - \ldots - a_p L^p$

$$b(L) = 1 + b_1 L + b_2 L^2 + \ldots + b_q L^q$$

Тогда если воспользовать оператором лага ARMA(p,q)

в операторной форме: $a(L) X_t = b(L) \varepsilon_t$

Свойства ARMA(p,q):

1. Процесс стационарен, если все корни ур-ия $a(z)=0$ лежат вне един. круга $|z| \leq 1$.

2. Если процесс стационарен, то можно представить процессом MA(∞), т.е. $\exists$ MA(∞): $X_t - \mu = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}$, $c_0 = 1$, $\sum_{j=0}^{\infty} |c_j| < \infty$; в операт. форме $X_t - \mu = c(L) \varepsilon_t$, где $c(z) = \sum_{j=0}^{\infty} c_j z^j = \frac{b(z)}{a(z)}$.

3. Все корни ур-ия $b(z) = 0$ лежат вне един. круга $|z| \leq 1$ (ус-ие обратимости), то $\exists$ AR(∞), т.е. $X_t - \mu = \sum_{j=1}^{\infty} d_j (X_{t-j} - \mu) + \varepsilon_t$; в операт. форме $d(L)(X_t - \mu) = \varepsilon_t$, где $d(L) = 1 - \sum_{j=1}^{\infty} d_j z^j = \frac{a(z)}{b(z)}$

Т.о. стац. процесс ARMA(p, q) всегда можно аппроксимировать MA(∞), а при выполнении усл-ия обратимости и AR(∞).

Если $k>q$ коррелограмма ARMA(p, q) выглядит также как для $a(L)X_t=\varepsilon_t$.

Автокорреляция для ARMA(1, 1): $X_t=a_1X_{t-1}+\varepsilon_t+b_1\varepsilon_{t-1}$

$\rho(k)=a_1\rho(k-1)$, $k=2,3,\ldots$, $k>q$, $\rho(0)=1$

Совпадает с AR(1): $\rho_k=a^k$, $k\geq 2$

Если ARMA(p_1,q_1) ряд X_t и ARMA(p_2,q_2) ряд Y_t статист. независ. м/у собой, $Z_t=X_t+Y_t$, то Z_t ряд ARMA(p, q), у которого $p=p_1+p_2$, $q=\begin{cases}p_1+q_2, & \text{если } p_1+q_2>p_2+q_1\\ p_2+q_1, & \text{если } p_2+q_1>p_1+q_2.\end{cases}$

Билет 17. Подбор стационарной модели ARMA (процедура подбора в общем). Идентификация стационарной модели ARMA. Поведение автокорреляционных и частных автокорреляционных функций.

Подбор ARMA:

1. идентификация модели
2. оценивание модели
3. диагностика модели
4. прогноз.

На этапе идентификации производится выбор p и q также на этом этапе делаются предварительные грубые оценки коэф. $a_1 \ldots a_p, b_1 \ldots b_q$ идентиф. модели

На втором этапе производится уточнение оценок коэф. модели с использованием эффективных статистических методов. Для оцененных коэф. вычисляются приближенные стандартные ошибки, дающие возможность, при дополнительных предположениях о распред. сл. вел. $X_1, X_2, \ldots$ строить доверительные интервалы для этих коэф. и проверять гипотезы об их истинных значениях с целью уточнения спецификации модели.

На третьем этапе применяются различные диагност. процедуры проверки адекватности выбранной модели по имеющимся данным. Неадекватности, обнаруженные в процессе такой проверки, могут указать на необходимость ~~[illegible]~~ корректировки модели, после чего производится новый цикл подбора и т.д. до тех пор, пока не будет получена удовлетвор. модель.

Нестационарный процесс авторегрессии — интегрированного скользящего среднего ARIMA(p, d, q). Подход Бокса-Дженкинса построения модели типа ARIMA(p,d,q) по реализации временного ряда (4 этапа)

ARIMA (p, d, q) (для нестацион. рядов)

d — порядок интегрирования (сколько раз взяли последовательные разности) т.е. разности порядка d подчиняются ARMA(p, q)

$$\Delta x_t = x_t - x_{t-1}$$

В операторном виде: $a_p(L)(1-L)^d \underbrace{(\Delta^d)x_t}_{y_t\text{-стац.}} = b_q(L)\varepsilon_t$

$$\Delta^d x_t = \sum_{i=1}^{p} a_i \Delta^d x_{t-i} + \sum_{j=1}^{q} b_j \varepsilon_{t-j} + \varepsilon_t$$, ε_t — белый шум

Подход Бокса-Дженкинса:

1. Идентификация (p, q) 1а) d, 1б) p, q;
2. Оценивание $a_1, .., a_p, b_1, .., b_q$
3. Диагностика
4. Использование модели (для прогнозирования будущ. знач. ряда)

В чем смысл подхода: сначала оценивается стационарность ряда. Выявляются наличие единичных корней и порядок интегрированности. Потом ряд преобразуется взятием разности соответствующего порядка и уже для преобразованной модели строится некоторая ARMA-модель, т.к. предполагается, что полученный процесс стацион., в отличие от исходного нестацион. процесса.

Выборочное мат. ожидание: $\bar{x} = \frac{1}{T}\sum_{t=1}^{T} x_t$ (среднее по выборке).

Выборочная дисперсия: $\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})^2$ (смещ.)

Выбор. автокорр.: $\hat{\rho}(k) = \frac{1}{T\hat{\sigma}^2}\sum_{t=1}^{T}(x_t - \bar{x})(x_{t-k} - \bar{x})$

(сделано для того, чтобы была положит. опред. матрица, но жертвуется смещенность оценок).

$\hat{\rho}(k)$, $k > T/4$ (представляются не более чем $T/4$)

Идентификация:

$x_t \sim MA(q)$ и $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, то для $k > q$ и больших T выборочная коррел. асимпт. распределена по норм. закону с параметрами: $\bar{\rho}(k) = E(\rho(k)) = 0$;

$\hat{\sigma}_\rho^2 = Var(\rho(k)) \approx \frac{1}{T}$, $k = 1$

$\hat{\sigma}_\rho^2 = Var(\rho(k)) \approx \frac{1}{T}\left(1 + 2\sum_{j=1}^{k-1}\rho^k(j)\right)$, $k > 1$

$H_0: \rho(k) = 0$, $k > q \Rightarrow MA(q)$ $\quad x_t - AR(p)$

$\hat{\varphi}_{kk} = E(\varphi_{kk}) = 0$

$\hat{\sigma}_\rho^2 = Var(\varphi_{kk}) \approx \frac{1}{T}$, $k = 1$

$\hat{\sigma}_\rho^2 = Var(\varphi_{kk}) \approx \frac{1}{T}\left(1 + 2\sum_{j=1}^{k-1}\varphi_{jj}^k\right)$, $k > 1$

$H_0: \varphi_{kk} = 0$, $k > p \Rightarrow AR(p)$

$\sqrt{T}\,\varphi_{kk} \sim N\{0, 1\}$ при $k > p$

(φ_{kk} – выбор. частн. автокорр. ф-ция)

Билет 19. Оценивание коэффициентов модели ARMA. Оценивание параметров модели AR(p) методом наименьших квадратов. Оценивание параметров модели MA(q) методом максимального правдоподобия и с помощью процедуры поиска на сетке

Для оценки коэф. модели AR(p): $X_t = a_1 X_{t-1} + \ldots + a_p X_{t-p} + \varepsilon_t$ можно применить обычный метод наим. квадратов.
Поскольку регрессоры относятся к предыдущим моментам времени, а ε_t - белый шум, то корреляция регрессоров со случайным возмущением ε_t отсутствует.
МНК не дает не смещенных оценок, но дает состоятельные оценки.
Если белый шум явл. гауссовым, то X_t распред. нормально, а оценки $\hat{a}_1 \ldots \hat{a}_p$ - состоятельноны и асимптотически нормальны.

Для оценки мат. ожидания процесса AR(p) можно использовать две статистики:

$$\bar{x} = \frac{1}{T}\sum_{t=1}^{T} x_t \quad \text{или} \quad \hat{\mu} = \frac{\hat{\delta}}{1 - \hat{a}_1 - \ldots - \hat{a}_p}$$

Для оценки параметров MA(q) существует два метода:

1. Метод максимального правдоподобия

В предположение $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ ковариационную матрицу ошибок через $\beta_1 \ldots \beta_q$ обратимой модели MA(q): $x_t = \sum_{j=0}^{q} b_j \varepsilon_{t-j}$.

Функция правдоподобия для x_t:

$$L(b, \sigma_\varepsilon^2 \,|\, x_t) = \frac{1}{(2\pi\sigma_\varepsilon^2)^{T/2} (\det \Gamma)^{1/2}} \exp\left\{ - \frac{(x - \bar{x})^T \Gamma^{-1} (x - \bar{x})}{2\sigma_\varepsilon^2} \right\}$$

ковариац. матрица

Элементы матрицы Г выражаются через $\beta_1 \ldots \beta_q$ поэтому процедура численной оптимизации позволяет найти оценки ММП, которые будут обладать обычными св-вами состоятельности и асимптотич. нормальности

Кроме того, оценки параметров модели MA(q) и оценка дисперсии случайного возмущения асимптотически независимы.

2. Нелинейная оптимизация: поиск на сетке

Рассмотрим идею на примере MA(2): $x_t = b_1\varepsilon_{t-1} + b_2\varepsilon_{t-2}$

1) Найдем оценку мат. ожидания процесса $\bar{x}$ в предположении о его стационарности и эргодичности. $\bar{x} = \hat{E}(x_t)$

2) Выберем некоторые значения (b_1, b_2)

3) Выразим x_1 через предыдущие ошибки:

$$x_t = \bar{x} + b_1^0 e_{t-1} + b_2^0 e_{t-2} + e_t$$
$$x_1 = \bar{x} + b_1^0 e_0 + b_2^0 e_{-1} + e_1$$
$$\Rightarrow e_1 = x_1 - \bar{x}$$
$$x_2 = \bar{x} + b_1^0 e_1 + b_2^0 e_0 + e_2$$
$$\Rightarrow e_2 = x_2 - \bar{x} - b_1^0 e_1$$

... и т.д.

4) Вычислим $\sum_{t=1}^{T} e_t^2$. Поскольку она посчитана для выбранных значений параметров (b_1, b_2) мы можем рассматривать ее как ф-цию этих перемен.

5) Перебираем поиском на сетке комбинации b_1 и b_2 или ищем минимум ф-ции: $\min\limits_{b_1, b_2} \sum_{t=1} e_t^2$. Коэф. (b_1^*, b_2^*) которые обеспечивают минимум этого выражения будут оценками коэф. модели MA(2)

Для оценки модели ARMA(p,q) может применяться комбинация МНК и поиска на сетке.

Рассмотрим на примере ARMA(2,2):

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \varepsilon_t + b_1\varepsilon_{t-1} + b_2\varepsilon_{t-2}$$

$$X_t = \frac{(1 + b_1 L + b_2 L^2)}{(1 - a_1 L - a_2 L^2)}\varepsilon_t$$

Введем вспомогательный случайный процесс:

$z_t = \frac{1}{1 - a_1 L - a_2 L^2}\varepsilon_t$. Тогда процессы X_t и Z_t связаны соотношением: $X_t = (1 + b_1 L + b_2 L^2) Z_t$.

Билет 19.

Определим z_t через X_t, сконструировав z_t также, как ранее остатки модели MA:

$$X_1 = z_1 + b_1 z_0 + b_2 z_2 \Rightarrow z_1 = X_1$$

$$X_2 = z_2 + b_1 z_1 + b_2 z_0 \Rightarrow z_2 = X_2 - b_1 z_1$$

Значения z_t связаны с e_t исходной модели след. соотношением: $e_t = z_t - \hat{a}_1 z_{t-1} - \hat{a}_2 z_{t-2}$

Относительно z_t модель стала AR(2). Зная "реализацию" z_t для выбранных значений коэф. b_1, b_2 можно оценить a_1 и a_2 с помощью МНК.

В рез-те находим оценки (a_1, a_2, b_1, b_2), но получены они по-разному. Оценки (b_1°, b_2°) заданы как начальное значение. А потом, исходя из них, построены оптимальные оценки (a_1°, a_2°). Применяя метод оптимизации: поиск на сетке оцениваем значение параметров, обеспечивающие минимум $\sum e_t^2$

Оценивание параметров модели ARMA(p,q): 1) комбинацией метода наименьших квадратов и поиска на сетке на примере ARMA(2,2); 2 методом максимального правдоподобия.

1. Комбинация МНК + поиск на сетке. ARMA(2,2):

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \varepsilon_t + b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2}$$

В операторной форме: $a(L)X_t = b(L)\varepsilon_t$. Распишем:

$$(1 - a_1 L - a_2 L^2) X_t = (1 + b_1 L + b_2 L^2)\varepsilon_t$$

Тогда $X_t = \dfrac{(1 + b_1 L + b_2 L^2)}{(1 - a_1 L - a_2 L^2)}\varepsilon_t$.

Вводится вспомогательный процесс:

$$Z_t = \frac{1}{1 - a_1 L - a_2 L^2}\varepsilon_t^{(*)} \Rightarrow X_t = \underbrace{(1 + b_1 L + b_2 L^2) Z_t}_{\text{похож на MA(2) (поиск на сетке)}}$$

Определим Z_t через X_t:

$$X_1 = Z_1 + \underbrace{\cancel{b_1 Z_0}}_{=0} + \underbrace{\cancel{b_2 Z_2}}_{=0} \Rightarrow Z_1 = X_1$$

$$X_2 = Z_2 + b_1 Z_1 + \underbrace{\cancel{b_2 Z_0}}_{=0} \Rightarrow Z_2 = X_2 - b_1 Z_1$$

......

$(*) \Rightarrow \varepsilon_t = (1 - a_1 L - a_2 L) \cdot Z_t$ – похож на AR(2) (МНК)

Выражаются остатки: $e_t = Z_t - \hat{a}_1 Z_{t-1} - \hat{a}_2 Z_{t-2}$

Выбираются нач. значения, (b_1^0, b_2^0) и/з МНК $\hat{a}_1, \hat{a}_2$ и

потом минимизируют $\sum e_t$ и получают (b_1, b_2, a_1, a_2).

3. ММП. ARMA(p,q)

$$X_t = \sum_{j=1}^{p} a_j X_{t-j} + \sum_{j=1}^{q} b_j \varepsilon_{t-j} + \varepsilon_t \quad, \varepsilon_t \sim N(0, \sigma^2)$$

Основная цель алгоритма: максимизация логарифм. ф-ции макс. правдоподобия, а именно

$$L(a,b,\sigma) = \sum_{t=1}^{T} \ln\left(\frac{1}{\sqrt{2\pi\omega_t}} \exp\left(-\frac{\nu_t^2}{2\omega_t}\right)\right), \text{ где}$$

$$\nu_t = X_t - E[X_t \mid X_1^{t-1}], \text{ где } X_1^t = (X_1, .., X_t), X_1^0 = 0$$

$$\omega_t = E[\nu_t^2 \mid X_1^{t-1}].$$

Общая схема:

Выражаются значения автокорр. ф-ции через параметры модели, формируется ковариац. матрица порядка T, записывается ф-ция правдоподобия для выборки $X_1, .., X_T$ и решается система ур-ний правдоподобия относительно оценок коэффициентов.

Билет 21. Диагностика модели ARMA. Проверка качества модели. Информационный критерий Акаике (AIC). Байесовский информационный критерий (BIC).

Проверке подлежат два момента:

1. качество модели
2. некоррелированность остатков.

Критерии качества подгонки:

1. AIC. для модели ARMA(p, q):

$$AIC(p, q) = \ln \hat{\sigma}^2 + 2\frac{p+q}{T}$$

2 BIC для модели ARMA(p, q)

$$BIC(p, q) = \ln \hat{\sigma}^2 + \ln T \frac{p+q}{T}$$

T - число наблюдений, $\hat{\sigma}^2 = \frac{\sum e_i^2}{T-p-q}$, $\hat{\sigma}^2 = \frac{\sum e_i^2}{T}$

Проверка автокорреляции остатков модели ARMA(p,q)
Тест Бокса-Пирса (Q-статистика). Тест Бокса-Льюнга (улучш. Q-стат.)

Q-статистика Бокса-Пирса - статист. критерий для проверки автокорреляции временного ряда.

Проверяется гипотеза, что ε_t действительно белый шум:

$H_0: \rho_1 = \rho_2 = \dots = \rho_m = 0$ против альтерн. гипотезы

$H_1: \sum_{k=1}^{m} \rho^2(k) > 0.$

ARMA(p,q): $X_t = a_1 X_{t-1} + \dots + a_p X_{t-p} + \varepsilon_t + b_1 \varepsilon_{t-1} + \dots + b_q \varepsilon_{t-q}$

или в операт. форме: $a(L) X_t = b(L) \varepsilon_t$.

Пусть её оценка: $\hat{a}(L) X_t = \hat{b}(L) \varepsilon_t$, где $\hat{a}(L) = 1 - \hat{a}_1 L - \dots - \hat{a}_p L^p$

$\hat{b}(L) = 1 + \hat{b}_1 L + \dots + \hat{b}_q L^q$.

Если MA составляющая модели ARMA(p,q) обратима:

$\varepsilon_t = \frac{a(L)}{b(L)} X_t$ и оценки для ε_t можно получить заменой:

$\hat{\varepsilon}_t = \frac{\hat{a}(L)}{\hat{b}(L)} X_t$ (при большом кол-ве наблюд. $\hat{\varepsilon}_t$ должно имитировать поведение самих ε_t) => если ε_t - белый шум, то и остатки - белый шум.

Выборочная автокорреляция: $r_\varepsilon(k) = \frac{\sum \hat{\varepsilon}_t \hat{\varepsilon}_{t+k}}{\sum_{t=1}^{T} \hat{\varepsilon}_t^2}$ и сумма их квадратов $Q(m) = T \sum_{k=1}^{m} \rho_\varepsilon^2(k)$

Если $Q(m) > \chi^2_\gamma(m)$ => H_0 отвергается и нужно исследовать модель.

Тест Бокса-Льюнга или улучшенная Q-статистика

$$Q = (T+2) \cdot T \sum_{i=1}^{m} \frac{\rho_\varepsilon^2(i)}{T-i} \sim \chi^2_\gamma(m)$$

о оба не применимы.