29. Свойства МНК-оценок параметров линейной модели множественной регрессии (ЛММР) при нормальном векторе случайных остатков: закон распределения случайного вектора $\tilde{\vec{a}}$.

Теорема: пусть в линейной модели множественной регрессии справедливы все предпосылки теоремы Гаусса-Маркова-Эйткена, и кроме того, случайный остаток имеет нормальный закон распределения, тогда случайные векторы $\tilde{\vec{a}}$ и $\tilde{\vec{u}}$ являются нормально распределенными и независимыми.

Доказательство:

$$\tilde{\vec{a}} = (X^T X)^{-1} X^T \vec{y}, \quad \tilde{\vec{u}} = \vec{y} - X\tilde{\vec{a}}, \quad \vec{y} = X\vec{a} + \vec{u}$$

(1)

$$Cov(\vec{u}, \vec{u}) = \sigma_u^2 E$$

(2)

$$E(\vec{u}) = 0$$

(3)

$$\Delta\tilde{\vec{a}} = \tilde{\vec{a}} - \vec{a} = QX^T(X\vec{a} + \vec{u}) - \vec{a} = QX^T X\vec{a} + QX^T\vec{u} - \vec{a} = QX^T\vec{u} \ \ (4), \text{т.к.}\, QX^T X\vec{a} = E$$

Видим, что $\Delta\tilde{\vec{a}}$ – линейное преобразование нормального распределенного $\vec{u}$ значит, является нормально распределенной и $E\left(\Delta\tilde{\vec{a}}\right) = 0$.

30. Свойства МНК-оценок параметров линейной модели множественной регрессии (ЛММР) при нормальном векторе случайных остатков: закон распределение оценки $\tilde{\sigma}_0^2$.

Теорема: пусть в линейной модели множественной регрессии справедливы все предпосылки теоремы Гаусса-Маркова-Эйткена, и кроме того, случайный остаток имеет нормальный закон распределения, тогда оценки $\tilde{\sigma}_0^2$ имеют следующее распределение:

$$\tilde{\sigma}_0^2 = \frac{\vec{u}^T \Omega^{-1} \vec{u}}{n - (k+1)}$$

Доказательство:

Рассмотрим величину Ψ ($\Psi(\vec{a})$ – это эффективная линейная несмещенная оценка, обладающая свойством наименьших квадратов). Она зависит от выборки $(\vec{y}, X)$, а значит, является случайной переменной.

Начнем с оценки вектора случайных остатков $\tilde{\vec{u}} = \vec{y} - X\tilde{\vec{a}}$ (1). Представим этот вектор как выход линейного преобразования вектора $\vec{u}$. Для этого подставим в правую часть (1) правую часть от $\tilde{\vec{a}} = f^*(X, \vec{y}) = M(X)\vec{y} = (X^T\Omega^{-1}X)^{-1}X^T\Omega^{-1}\vec{y}$ и приведем подобные члены:

$$\tilde{\vec{u}} = \vec{y} - X(X^T\Omega^{-1}X)^{-1}X^T\Omega^{-1}\vec{y} = (E - XQ \cdot X^T\Omega^{-1})\vec{y}$$

Здесь приняли обозначение $Q = (X^T\Omega^{-1}X)^{-1}$. Теперь в правую часть предпоследнего равенства подставляем правую часть $\vec{y} = X\vec{a} + \vec{u}$ и, раскрывая скобки, получаем искомое преобразование:

$$\tilde{\vec{u}} = (E - XQ \cdot X^T\Omega^{-1})(X\vec{a} + \vec{u}) = (E - XQ \cdot X^T\Omega^{-1})\vec{u} \ \ (2)$$

В компактном виде получаем $\tilde{\vec{u}} = H\vec{u}$ $(2')$. С учетом (2) находим значение квадратичной формы $\tilde{\psi}$:

$$\tilde{\psi} = \tilde{\vec{u}}^T\Omega^{-1}\tilde{\vec{u}} = \tilde{\vec{u}}^T\Omega^{-1}\tilde{\vec{u}} - \vec{u}^T(\Omega^{-1}XQX^T\Omega^{-1})\vec{u} \ \ (3)$$

Возьмем за данное, что $E(\vec{u}^T(\Omega^{-1}XQX^T\Omega^{-1})\vec{u}) = \sigma_0^2(k+1)$ (4)

Теперь согласно тому, что $E(\Psi) = E\left(\tilde{\vec{u}}^T\Omega^{-1}\tilde{\vec{u}}\right) = \sigma_0^2 n$ и равенству (4) и свойств операции с $E(\cdot)$ получаем $E(\vec{w}) = 0; Cov(\vec{w}, \vec{w}) = \sigma_0^2 E$, где $\vec{w} = Q^{\frac{1}{2}}X^T\Omega^{-1}\vec{u}$.

$$Var(w_i) = E\left(w_i^2\right) = \sigma_0^2$$

Таким образом, из равенства (3) следует требуемое доказательство равенства

$$\tilde{\sigma}_0^2 = \frac{\vec{u}^T \Omega^{-1} \vec{u}}{n - (k+1)}$$

31. Свойства МНК-оценок параметров линейной модели множественной регрессии (ЛММР) при нормальном векторе случайных остатков: закон распределения дроби $\dfrac{\tilde{a}_j - a_j}{S\tilde{a}_j}$.

$S\tilde{a}_j$ - стандартная ошибка (оценка среднеквадратического отклонения) компоненты $\tilde{a}_j$.
Докажем, что случайная переменная
$\frac{\Delta\tilde{a}_j}{S\tilde{a}_j} = \frac{\tilde{a}_j - a_j}{\tilde{\sigma}_0\sqrt{Q_{j+1,j+1}}}$ (1) имеет закон распределения Стьюдента с количеством степеней свободы $n-(k+1)$, т.е. $\frac{\Delta\tilde{a}_j}{S\tilde{a}_j} \propto t_{n-(k+1)}$ (2)
Доказательство.
Разделим числитель и знаменатель дроби (1) на константу
$\sigma(\tilde{a}_j) = \sigma_0\sqrt{Q_{j+1,j+1}}$.
Учитывая, что

$$\tilde{\sigma}_0^2 = \frac{\tilde{\tilde{u}}^T\Omega^{-1}\tilde{\tilde{u}}}{n-(k+1)} \propto \frac{\sigma_0^2}{n-(k+1)}\chi^2_{n-(k+1)}$$

получим:

$$\begin{cases} \frac{\Delta\tilde{a}_j}{\sigma(\tilde{a}_j)} = \frac{\tilde{a}_j - a_j}{\sigma_0\sqrt{Q_{j+1,j+1}}} \\ \frac{\tilde{\sigma}_0}{\sigma_0} \propto \sqrt{\frac{\chi^2_{n-(k+1)}}{n-(k+1)}} \end{cases} \quad (3)$$

Здесь символом $\dot{u}$ обозначена стандартная нормально распределённая случайная переменная.
$t_n = \frac{\dot{u}_{n+1}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\dot{u}_i^2}}$ (4) - дробь Стьюдента с n степенями свободы.
С учётом (3) и (4) получим представление (2).

32. Оценивание параметров линейной модели множественной регрессии (ЛММР) при автокоррелированном случайном остатке алгоритмом Хилдрета-Лу.

Частным случаем линейной регрессионной модели является модель динамического ряда

$$\begin{cases} y_t = a_0 + a_1x_{1,t} + \cdots + a_kx_{k,t} + u_t \\ E(u_t) = 0,\ \ u_t \in STS \end{cases} \quad (1)$$

с автокоррелированным случайным остатком u_t, который интерпретируется как некоторый стационарный временной ряд $u_t \in STS$. Простейшей, естественной и достаточно гибкой моделью стационарного временного ряда служит модель $u_t \in AR(1)$ - модель авторегрессии первого порядка. Эту модель пример в модели (1):

$$\begin{cases} y_t = a_0 + a_0x_{1,t} + \cdots + a_kx_{k,t} + u_t \\ u_t = \rho u_{t-1} + \xi_t \\ E(u_t) = 0, \quad E(u^2{}_t) = \sigma_u^2 = \frac{\sigma_\xi^2}{(1-\rho^2)} \end{cases} \quad (2)$$

Здесь ξ_t – белый шум (стационарный временной ряд с некоррелированными уровнями, имеющими нулевое математическое ожидание и одинаковую дисперсию). В этой модели оцениванию подлежат следующий набор параметров: $(a_0, a_0, \dots, a_k, \sigma_\xi^2, \rho)$.
Для начала модель (2) запишем в компактном виде:
$y_t = \vec{a}^T\vec{x}_t + u_t$ (3).
При $t-1$:
$y_{t-1} = \vec{a}^T\vec{x}_{t-1} + u_{t-1}$ (4).
Умножим (4) на константу ρ и вычтем из уравнения (3). Учитывая второе уравнение модели (2), получим трансформированную модель:

$$\begin{cases} y_t = \rho y_{t-1} + \vec{a}^T\vec{x}_t - \vec{a}^T\rho\vec{x}_{t-1} + \xi_t \\ E(\xi_t) = 0, \quad E(\xi^2{}_t) = \sigma_\xi^2 \end{cases} \quad (5)$$

Это есть авторегрессионная моделью в ней предопределенная переменная y_{t-1} коррелирует со значением случайного остатка ξ_{t-1} в предшествующий период времени.
Алгоритм Хилдрета – Лу. Перегруппируем члены в модели (5):

$$\begin{cases} y_t - \rho y_{t-1} = \vec{a}^T(\vec{x}_t - \rho\vec{x}_{t-1}) + \xi_t \\ E(\xi_t) = 0, \quad E(\xi^2{}_t) = \sigma_\xi^2 \end{cases} \quad (6)$$

Если бы константа ρ была известна, то модель (4) была бы ЛММР, в рамках которой все предпосылки теоремы Гаусса – Маркова справедливы. Так что можно было бы оптимально оценить МНК ее параметры $\vec{a}$, σ_{ξ}^2.

Но ρ – неизвестная константа, значение которой принадлежит промежутку [0, 1). Алгоритм ее определения нелинейным МНК, принадлежащий Хильдрету и Лу, состоит из следующих шагов.

Шаг 1. Задаться в промежутке [0, 1) набором пробных значений $\rho = \rho_i$ по правилу

$\rho = \rho_i = \frac{i-1}{N}, \quad i = 1, 2, \dots, N.$ (7),

где N – некоторое натуральное число.

Шаг 2. При каждом значении (7) составить в рамках модели (6) систему уравнений наблюдений:

$$\begin{cases} y_2 - \rho y_1 = \vec{a}^T(\vec{x}_2 - \rho\vec{x}_1) + \xi_2 \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ y_n - \rho y_{n-1} = \vec{a}^T(\vec{x}_n - \rho\vec{x}_{n-1}) + \xi_n \end{cases} \quad (8)$$

и вычислить на основании этой системы МНК – оценки

$\vec{a}, \quad \sigma_{\xi}^2, \quad \sum_{j=2}^{n} \widetilde{\xi_j^2}$ (9)

Шаг 3. Выбрать из множества пробных значений (7) такую величину $\rho_i = \tilde{\rho}$ и соответствующие ей оценки (9), при которой имеет место экстремум:

$\sum_{j=2}^{n} \widetilde{\xi_j^2} \to min.$ (10).

Выбранные величины и будут искомыми оценками параметров.

33. Порядок проверки статистических гипотез (на примере гипотезы об адекватности ЛММР).

Шаги:

Гипотеза проверки адекватности: H_0: значение эндогенной переменной равно сумме значения функции регрессии и остатка u_0, где $E(u_0) = 0, E(u_0^2) = \sigma_u^2$. Альтернативная гипотеза $H_1 = \overline{H_0}$

Строится статистика критерия проверяемой гипотезы: $t = \frac{\widetilde{y_0} - y_0}{S_{\widetilde{y_0}}}$

Принимается уровень значимости α и определяется критерий данной гипотезы и множество принятия данной гипотезы.

Вычисляется значение статистики, проверяется попадание данной статистики в область принятия гипотезы.

34. Спецификация и оценивание нелинейных по коэффициентам моделей множественной регрессии со специальными функциями регрессии (на примере производственной модели с функцией Кобба-Дугласа).

Возьмем вид уравнений наблюдений объекта в рамках создаваемой модели: $\vec{y} = X\vec{a} + \vec{u}$ (1)

Эти уравнения, именуемые схемой Гаусса-Маркова, являются линейными относительно вектора $\vec{a}$ оцениваемых коэффициентов функции регрессии. Уравнения наблюдений (1) возникают тогда и только тогда, когда функция регрессии в модели

$$\begin{cases} y = f(\vec{x}; \vec{a}) + u; \\ E(u|\vec{x}) = 0, \quad Var(u^2|\vec{x}) = \sigma_u^2 \end{cases} \quad (2)$$

линейна относительно своих коэффициентов. Однако нередко оцененная линейная модель оказывается неадекватной по той причине, что функция регрессии (поведенческое уравнение) необоснованно принята линейной по коэффициентам, и тогда экономист вынужден на стадии спецификации эконометрической модели привлекать именно нелинейные по коэффициентам функции регрессии. Примерами такой модели служат модель линейной парной регрессии и линейная модель множественной регрессии. Классическим примером модели, в которой поведенческое уравнение нелинейно по коэффициентам, служит производственная функция Кобба-Дугласа:

$$\begin{cases} Y = a_0 K^{\alpha} L^{\beta} \\ a_0 > 0, 0 < \alpha, \beta < 1 \end{cases} \quad (3)$$

Y – уровень выпуска продукции за принятый отрезок времени;

K и L – уровни соответственно основного капитала и живого труда, использованные в процессе выпуска величины Y. Из (3) видно, что функция, образующая поведенческое уравнение данной системы, нелинейна по коэффициентам, то есть не относится к баз мод эконометрики. Однако (3) можно трансформ к линейной ункции при помощи операции логарифмирования. $lnY = ln\, a_0 + \alpha lnK + \beta lnL$ (5) Получили бозовою мод эконометрики Здесь функция lnY аргументов lnK и lnL линейна по коэффициентам $b_0 = ln\, a_0, \quad b_1 = a_1, \; b_2 = \beta$ (6) В силу функциональной зависимости коэффициентов b_1 и b_2 уравнение (5) можно упростить: $y = b_0 + b_1 x,$ (7) где $y = lnY - ln\, L = ln\frac{Y}{L}, \quad x = ln\, K - ln\, L = ln\frac{K}{L}$ (8) Чтобы трансформировать модель (3) в эконометрическую, необходимо случайнst возмущения u включить в поведенческое уравнение не в качестве слагаемого, а виде сомножителя, подходящего для последующей операции логарифмирования. Вот одна из подходящих спецификаций производственной эконометрической модели с функцией регрессии (3):

$$\begin{cases} Y = a_0 K^{a_1} L^{1-a_1} e^u \\ a_0 > 0, 0 < a_1 < 1 \end{cases} \quad (9)$$

$E(u|K,L) = 0, \quad Var(u \;\; |K,L) = \sigma_u^2$ (11) Логарифмирование нелинейного уравнения эконометрической модели (9) приводит с учетом обозначений (7) и (8), а также предпосылок (11) к линейной модели парной регрессии:

$$\begin{cases} y = b_0 + b_1 x + u; \\ E(u|K,L) = 0, \quad Var(u \;\; |K,L) = \sigma_u^2 \end{cases} \quad (13)$$

Таким образом, мы свели задачу построения нелинейной по коэффициентам регрессионной модели к задаче построения линейной регрессионной модели. Проследим все этапы построения производственной эконометрической модели (9), которая трансформирована в линейную регрессионную модель (13).

Для определенности предположим, что случайный остаток в модели (13) удовлетворяет всем предпосылкам теоремы Гаусса-Маркова. Пусть по обучающей выборке получена МНК-оценка модели (13):

$$\begin{cases} y = \tilde{b}_0 + \tilde{b}_1 x + u; \\ R^2 = \cdots \end{cases} \quad (14)$$

прошедшая проверку адекватности. Пусть далее требуется вычислить прогноз значений Y_0 эндогенной переменной модели (9), соответствующего значениям экзогенных переменных K_0, L_0. Для расчета этого прогноза осуществим следующие действия.

1. Найдем $x_0 = ln\frac{K_0}{L_0}$ и по модели (14) получим линейный оптимальный прогноз величины $y_0 = ln\frac{Y_0}{L_0}$:
$$\begin{cases} \tilde{y}_0 = \tilde{b}_0 + \tilde{b}_1 x_0 \\ S_{\tilde{y}_0} = \tilde{\sigma}_u \sqrt{1 + q_0} \end{cases} \quad (15)$$

2. Используя результаты (15) рассчитаем прогноз $\tilde{Y}_0$ величины Y_0 и стандартную ошибку $S_{\tilde{Y}_0}$ этого прогноза: $\tilde{Y}_0 = L_0 e^{\tilde{y}_0}$ (16)

$S_{\tilde{Y}_0} = \tilde{Y}_0 S_{\tilde{y}_0} = \tilde{Y}_0 \tilde{\sigma}_u \sqrt{1 + q_0}$ (17) Формулу (17) можно интерпретировать следующим образом: величина $S_{\tilde{y}_0}$, определенная по правилу (15), является относительной среднеквадратической ошибкой прогноза (16). Вернемся к спецификации модели (9) и представим ее оценку в стандартно виде:

$Y = \tilde{a}_0 K^{\tilde{a}_1} L^{1-\tilde{a}_1} \mathrm{u}$ (18) $(S_{\tilde{a}_0})$ $(S_{\tilde{a}_1})$ $\tilde{\sigma}_u$ Коэффициенты модели и их стандартные ошибки вычисляются с учетом (6) по формулам:

$\tilde{a}_0 = e^{\tilde{b}_0}, \;\; \tilde{a}_1 = \tilde{b}_1, \;\; S_{\tilde{a}_0} = \tilde{a}_0 S_{\tilde{b}_0}, \quad S_{\tilde{a}_1} = S_{\tilde{b}_0}$ (19)

Этапы построения регрессионной модели со стандартной функцией регрессии, нелинейной по коэффициентам. Шаг 1. В процессе спецификации эконометрических моделей с нелинейными по коэффициентам стандартными функциями регрессии случайные остатки следует включать в поведенческие уравнения в виде соответствующих сомножителей. Затем поведенческие уравнение операцией логарифмирования трансформируется в модель линейной регрессии.

Шаг 2. Построив трансформированную линейную модель, следует обратным преобразованием (потенцированием) получить оценку исходной нелинейной модели.

Шаг 3. Прогноз эндогенной переменной исходной нелинейной модели можно строить либо при помощи прогноза ее логарифма, полученного при оценке трансформированной линейной модели, либо же по оценке исходно модели.   Шаг 4. Стандартная ошибка прогноза рассчитывается по формуле (17).

35. Оптимальное точечное прогнозирование значений эндогенной переменной по линейной модели (случай гомоскедастичного и неавтокоррелированного случайного возмущения).

Рассмотрим построение оптимального (наиболее точного) прогноза искомого значения y_0 эндогенной переменной линейной модели множественной регрессии на примере модели Оукена:

$$\begin{cases} y_t = a_0 + a_1 x_t + u_t \\ E(\mathrm{u}) = 0; \quad Var(\mathrm{u}) = \sigma_u^2 \end{cases} \qquad (1)$$

y_t – темп прироста реального ВВП;
x_t – изменение уровня безработицы.
Пусть модель (1) оценена МНК по выборке:

$$(\vec{y}, X) \qquad (2)$$

В ситуации, когда все предпосылки т. Гаусса-Маркова адекватны. Таким образом, имеется оценка модели (1):

$$\begin{cases} y_t = \tilde{a}_0 + \tilde{a}_1 x_t + u_t \\ \quad (S_{\tilde{a}_0}) \quad (S_{\tilde{a}_1}) \quad \tilde{\sigma}_u \\ \qquad R^2 = \cdots \end{cases} \qquad (3)$$

Обозначим символом x_0 значение экзогенной переменной данной модели, при котором, по смыслу задачи, должен быть вычислен прогноз значения эндогенной переменной. Прогноз обозначим символом $\tilde{y}_0$; в свою очередь, для наблюденного в реальности значения переменной y в ситуации, когда $x = x_0$, будем использовать символ y_0. Заметим, что в рамках модели (1) пара (x_0, y_0) связана уравнением

$$y_0 = a_0 + a_1 x_0 + u_0 \qquad (4)$$

где случайный остаток u_0 обладает, по предположению, количественными характеристиками

$$m = E(u_0) = 0, \quad Var(u_0) = \sigma_u^2 \qquad (5)$$

Подчеркнем еще, что и величины, образующие выборку (2) связаны между собой уравнениями наблюдений, схемой Гаусса-Маркова, аналогичными уравнениями (4). Случайные остатки в этих уравнениях тоже имеют, по предположению, параметры (5). В рамках модели (1) при наличии информации об объекте-оригинале в виде выборки (2) наилучший точечный прогноз величины y_0 вычисляется по правилу

$$\tilde{y}_0 = \tilde{a}_0 + \tilde{a}_1 x_0 \qquad (6)$$

т.е. в итоге подстановки в МНК-оценку функции регрессии модели (1) значения $x = x_0$ экзогенной переменной. В свою очередь средняя квадратическая (стандартная) ошибка прогноза (6) отыскивается по формуле

$$S_{\tilde{y}_0} = \tilde{\sigma}_u \sqrt{1 + q_0} \qquad (7)$$

где

$$q_0 = \vec{x}_0^{\,T} (X^T X)^{-1} \vec{x}_0 = \vec{x}_0^{\,T} Q \vec{x}_0 \qquad (8)$$

$$\overline{x}_0^{\,T} = (1, x_0)^{\mathrm{T}} \qquad (9)$$

Для модели парной регрессии (в т.ч. модели Оукена) формула (9) может быть представлена в следующем виде:

$$q_0 = \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \qquad (10)$$

Из (10) видно, что точность прогноза (6) падает по мере удаления значения x_0 регрессора x от его выборочного среднего. Описанная выше процедура точечного прогноза в рамках линейной модели парной регрессии остается в силе и для линейной модели множественной регрессии.