1. Назначение экономико-математических моделей (ЭММ). Два принципа их спецификации. Типы уравнений в ЭММ: поведенческие уравнения и тождества (на примере макромодели).

Назначение экономико-математических моделей (ЭММ)

Экономико-математическая модель (ЭММ, эконометрическая модель) объекта – это некоторое математическое выражение (график или таблица, уравнение или система уравнений, дополненная, возможно, неравенствами, условие экстремума), связывающее воедино исходные данные и искомые неизвестные задачи.

Два принципа спецификации эконометрической модели:

1. Эконометрическая модель возникает в итоге записи математическим языком взаимосвязей исходных данных и искомых неизвестных.
2. Количество уравнений модели обязано совпадать с числом искомых неизвестных.

Типы уравнений в ЭММ: поведенческие уравнения и тождества

Рассмотрим макромодель Кейнса, экономическим объектом в которой является закрытая экономика.

Экзогенные переменные: I – объем инвестиций в экономику страны.

Эндогенные переменные: $C-$ уровень потребления в стране, $Y-$ валовой внутренний продукт (ВВП).

Применим первый метод спецификации:

1) доход состоит из потребительских расходов и инвестиционных затрат

$Y = C + I-$ уравнение представляет собой основное тождество системы национальных счетов для закрытой экономики

2) уровень потребительских затрат объясняется доходом

$C = a_0 + a_1 \cdot Y-$ с позиции математики переменная C – функция переменной Y, а именно – линейная алгебраическая функция; такое уравнение принято называть ***поведенческим***.

3) с ростом дохода увеличивается потребление, каждая доп. единица дохода потребляется не полностью, какая-то часть идет на инвестиции, поэтому $0 < a_1 < 1$

Итак:

- ***тождество*** представляет собой равенство, выполняющееся в любом случае (тождество – это уравнение без коэффициентов);
- ***поведенческое*** уравнение включает параметры (a_0, a_1), значения которых являются неизвестными и подлежат оцениванию.

2. Типы переменных в экономических моделях. Структурная и приведённая форма модели (на примере макромодели). Компактная запись.

Типы переменных в эконометрических моделях:

Экзогенные переменные – исходные данные (экономические переменные, значения которых определяются вне модели и заранее известны).

Эндогенные переменные – искомые неизвестные (экономические переменные, значения которых нужно определить внутри модели).

Лаговые переменные - экзогенные и эндогенные переменные экономических моделей, датированные предыдущими моментами времени.
Объясняемые переменные – текущие эндогенные переменные.
Предопределенные переменные – текущие и лаговые экзогенные переменные, а также лаговые эндогенные переменные, если они стоят в уравнении с текущими эндогенными переменными.

Структурная и приведенная формы модели
Рассмотрим макромодель Кейнса, экономическим объектом в которой является закрытая экономика.
Экзогенные переменные: I – объем инвестиций в экономику страны.
Эндогенные переменные: C – уровень потребления в стране, Y – валовой внутренний продукт (ВВП).
Структурная форма модели – модель, полученная в результате записи математическим языком взаимосвязей эндогенных и экзогенных переменных:

$$\begin{aligned} Y &= C + I \\ C &= a_0 + a_1 Y \\ 0 &< a_1 < 1 \end{aligned}$$

Мы можем привести модель случае методом подстановки к ***приведенной форме***, где каждая эндогенная переменная представляется в виде явной функции только экзогенных переменных:

$$C = \frac{a_0}{1-a_1} + \frac{a_1}{1-a_1} \cdot I$$

$$Y = \frac{a_0}{1-a_1} + \frac{1}{1-a_1} \cdot I$$

Компактная запись
Обозначив векторы эндогенных переменных $\overline{y} = \begin{matrix} Y \\ C \end{matrix}$ и экзогенных переменных $\overline{x} = \begin{matrix} 1 \\ I \end{matrix}$, мы можем записать макромодель Кейнса в компактном виде:

$$A\overline{y} + B\overline{x} = 0$$

Составив матрицы A и B, получим компактную запись:

$$\begin{matrix} 1 & -1 \\ -a_1 & 1 \end{matrix} \quad \begin{matrix} Y \\ C \end{matrix} + \begin{matrix} 0 & -1 \\ a_0 & 0 \end{matrix} \quad \begin{matrix} 1 \\ I \end{matrix} = \begin{matrix} 0 \\ 0 \end{matrix}$$

3. Спецификация и преобразование к приведённой форме динамических моделей. Лаговые и предопределённые переменные динамической модели. Модель Линтнера корректировки размера дивидендов. Компактная запись.
Спецификация и преобразование к приведенной форме динамических моделей
Для отражения в спецификации модели фактора времени её переменные датируются (привязываются ко времени). Модель с датированными переменными

именуется ***динамической***. Стоит отметить, что датирование переменных является третьим принципом спецификации эконометрической модели.
Датированные переменные бывают текущие (датированные текущим моментом времени) и лаговые (датированные предыдущими моментами времени).
В свою очередь, все переменные динамической модели делятся на:

1. объясняемые – текущие эндогенные переменные;
2. предопределенные (объясняющие), включающие:
   - лаговые эндогенные;
   - текущие экзогенные;
   - лаговые экзогенные.

Модель Линтнера корректировки размера дивидендов
Исходные данные: EPS – чистая прибыль на акцию;
Искомые величины: DPS – объем дивидендов на акцию;
Утверждения, на которых построена модель:

1) фирма имеет долговременную долю в чистой прибыли на акцию, которую она хотела бы выплачивать в виде дивидендов своим акционерам в текущем периоде;
2) уровень дивидендов в текущем периоде объясняется желаемым уровнем дивидендов в этом периоде и уровнем реальных дивидендов в предшествующем периоде;

Спецификация модели:

$$\begin{cases} DPS_t^w = \gamma \cdot EPS_t \\ DPS_t = \lambda \cdot DPS_t^w + (1-\lambda) \cdot DPS_{t-1} \\ 0 < \gamma < 1 \end{cases}$$

Объясняемые переменные: DPS_t^w и DPS_t – желаемый и реальный уровень дивидендов в текущем периоде;
Предопределенные переменные: DPS_{t-1} и EPS_t – реальный уровень дивидендов в предшествующем периоде и чистая прибыль на акцию в текущем периоде.

Компактная запись

Обозначив векторы эндогенных переменных $\overline{y} = \begin{pmatrix} DPS_t^w \\ DPS_t \end{pmatrix}$ и экзогенных переменных
$\overline{x} = \begin{pmatrix} DPS_{t-1} \\ EPS_t \end{pmatrix}$, мы можем записать модель Линтнера в компактном виде:

$$A\overline{y} + B\overline{x} = 0$$

Составив матрицы A и B, получим компактную запись:

$$\begin{pmatrix} 1 & 0 \\ -\lambda & 1 \end{pmatrix} \begin{pmatrix} DPS_t^w \\ DPS_t \end{pmatrix} + \begin{pmatrix} 0 & -\gamma \\ -(1-\lambda) & 0 \end{pmatrix} \begin{pmatrix} DPS_{t-1} \\ EPS_t \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

4. Спецификация и преобразование к приведённой форме эконометрических моделей. Эконометрическая модель Самуэльсона–Хикса делового цикла экономики. Компактная запись.

Спецификация и преобразование к приведённой форме эконометрических моделей

Принципы спецификации эконометрической модели:

1. Эконометрическая модель возникает в итоге записи математическим языком взаимосвязей исходных данных и искомых неизвестных.
2. Количество уравнений модели обязано совпадать с числом искомых неизвестных.
3. Переменные модели датируются, что позволяет нам получить динамическую модель, в которой текущие эндогенные переменные объясняются значениями предопределенных переменных.
4. Поведенческие уравнения модели включают в себя случайные возмущения, таким образом, мы отражаем в спецификации влияние на текущие эндогенные переменные неучтенных факторов (повышая тем самым адекватность модели).

На основании всех четырех принципов спецификации в самом общем случае структурная форма эконометрической модели имеет вид:

$$F(\overline{x}_t,\ \overline{y}_t) = \overline{u}_t$$

а приведенная форма:

$$\overline{y}_t = f(\overline{x}_t,\ \overline{u}_t)$$

Эконометрическая модель Самуэльсона–Хикса делового цикла экономики

Спецификация модели (структурная форма):

$$\begin{cases} C_t = a_0 + a_1 \cdot Y_t + u_t & 1 > a_1 > 0 \\ I_t = b \cdot (Y_{t-1} - Y_{t-2}) + v_t & b > 0 \\ G_t = g \cdot G_{t-1} + w_t & g > 1 \\ Y_t = C_t + I_t + G_t & \end{cases}$$

Приведенная форма модели:

$$\begin{cases} C_t = a_0 + a_1 \cdot Y_{t-1} + u_t \\ I_t = b \cdot (Y_{t-1} - Y_{t-2}) + v_t \\ G_t = g \cdot G_{t-1} + w_t \\ Y_t = a_0 + (a_1 + b) \cdot Y_{t-1} - b \cdot Y_{t-2} + g \cdot G_{t-1} + (u_t + v_t + w_t) \end{cases}$$

$$\text{Объясняющие переменные}: \vec{x}_t = (Y_{t-1}, Y_{t-2}, G_{t-1})$$

$$\text{Объясняемые переменные}: \vec{y}_t = (Y_t,\ C_t,\ I_t,\ G_t)$$

Компактная запись

Обозначив векторы эндогенных переменных $\overline{y} = \begin{pmatrix} Y_t \\ C_t \\ I_t \\ G_t \end{pmatrix}$ и экзогенных переменных

$\overline{x} = \begin{pmatrix} 1 \\ Y_{t-1} \\ Y_{t-2} \\ G_{t-1} \end{pmatrix}$, мы можем записать модель Линтнера в компактном виде:

$$A\overline{y} + B\overline{x} = 0$$

Составив матрицы A и B, получим компактную запись:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & -1 & -1 & 1 \end{pmatrix}\begin{pmatrix} Y_t \\ C_t \\ I_t \\ G_t \end{pmatrix} + \begin{pmatrix} -a_0 & -a_1 & 0 & 0 \\ 0 & -b & b & 0 \\ 0 & 0 & 0 & -g \\ 0 & 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ Y_{t-1} \\ Y_{t-2} \\ G_{t-1} \end{pmatrix} = \begin{pmatrix} u_t \\ v_t \\ w_t \\ 0 \end{pmatrix}$$

5. Схема построения эконометрических моделей.

Шаг 1. Спецификация модели. В частности, фрагмент модели сельско-хозяйственных государственных расходов имеет следующую спецификацию:

$$\begin{cases} G_t = g\,G_{t-1} + w_t & g > 1 \\ E(w_t) = 0;\ Var(w_t) = \sigma_w^2 \end{cases} \tag{1}$$

Спецификация эконометрической модели обязательно содержит неизвестные константы. Они называются *параметрами модели.* В (1) параметры модели g, σ_w, где g – темп роста государственных расходов, σ_w – среднеквадратичное отклонение случайного возмущения или мера влияния неучтённых факторов.

Шаг 2. Сбор и проверка статистической информации в конкретных значений переменных, входящих в модель. (Примером такой информации служит файл "Элементы использования ВВП").

Собранную статистичестическую информацию разделяют на две части: большую $\approx 80\%$ часть имеет *обучающая* вырока и используется для определения параметров модели. Остальную часть отправляют на проверку инфляции и именуют *тестовой или контрольрующей выборкой.*

Примем обучающейся информации C (2002-2017 годов). Данные за 2018 год отнесём к контролирующей выборке.

Шаг 3. Оценивание по обучающей выборке неизвестных параметров модели методами математической статистики. На этом этапе по обучающейся выборке вычислим оценку $\left(\widetilde{g}, \widetilde{\sigma}_w\right)(3)$. Оценки (3) вычислим методом наименьших квадратов.

Шаг 4. Оценённая модель проходит проверку адекватности:

$$G_t = \widetilde{g}\,(S_{\widetilde{g}})G_{t-1} + w_t\left(\widetilde{\sigma}_w\right) \tag{4}$$

$$\widetilde{G}_{t(2018)} = \widetilde{g} \cdot G_{t-1(2017)}$$

$$\delta = |\widetilde{G}_t - G_t|\,(G_t) \cdot 100 \le 15\% \tag{5}$$

Модель признаётся адекватной, если отностительная ошибка прогноза не превышает 15%.

6. Порядок оценивания линейной эконометрической модели из изолированного уравнения в Excel. Смысл выходной статистической информации функции ЛИНЕЙН.

Пусть у нас построена линейная эконометрическая модель с изолированными переменными:

$$\begin{cases} y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + \ldots + a_k x_{kt} + u_t \\ E(u_t) = 0; \; E\left(u_t^{\,2}\right) = \sigma_u^2; \end{cases}$$

$x_{1t},\ldots,x_{nt}$ – объясняющие переменные, y_t – эндогенная переменная $(t = 1, 2, \ldots, n)$

Порядок оценивания модели состоит в следующем:

- Вводим исходные данные или открываем из существующего файла, содержащего анализируемые данные;
- На панели инструментов "Стандартная" щелкаем на кнопке "Вставка" функции;
- В окне "Категория" выбираем "Статистические", в окне "Функция" – ЛИНЕЙН, щелкаем ОК;
- Заполняем аргументы функции:
  - Известные значения y – диапазон, содержащий данные результативного признака;
  - Известные значения x – диапазон, содержащий данные факторов независимого признака;
  - Константа – логическое значение, которое указывает на наличие или на отсутствие свободного члена в уравнении (если Константа = 1, то свободный член рассчитывается обычным образом, если Константа = 0, то свободный член равен 0);
  - Статистика – логическое значение, которое указывает, выводить дополнительную информацию или нет (если статистика = 1, то дополнительная информация выводится, если Статистика = 0, то выводятся только оценки параметров уравнения).
- Щелкаем ОК.
- Выделяем область пустых ячеек $5 \times (k+1)$, т.е. (5 строк, $k+1$ столбцов) для вывода результатов регрессионной статистики.
- Ставим курсор на конец формулы в Строке формул
- Нажимаем комбинацию клавиш <CTRL>+<SHIFT>+<ENTER>.

Дополнительная регрессионная статистика будет выводиться в порядке, как на следующей схеме:

$\tilde{a}_k$	$\tilde{a}_{k-1}$	$\ldots$	$\tilde{a}_1$	$\tilde{a}_0$
$S_{\tilde{a}_k}$	$S_{\tilde{a}_{k-1}}$	$\ldots$	$S_{\tilde{a}_1}$	$S_{\tilde{a}_0}$
R^2	$\tilde{\sigma}_u$	#	#	#

F	v_2	#	#	#
RSS	ESS	#	#	#

$\widetilde{a}_0, \dots, \widetilde{a}_k -$ оценки коэффициентов;

$S_{\widetilde{a}_0}, \dots, S_{\widetilde{a}_k} -$ стандартные ошибки коэффициентов;

R^2 – коэффициент детерминации;

$\widetilde{\sigma}_u$ – оценка меры влияния случайного возмущения;

F – статистика Фишера;

ESS – сумма квадратов оценок случайных возмущений;

$RSS = \sum \left(\widetilde{y}_i - \overline{\widetilde{y}}\right)^2$;(но это не точно)

v_2

7. Случайная переменная и закон её распределения. Нормальный закон распределения и его параметры.

Случайной переменной u – называется переменная величина, возможные значения которой $(q_1, q_2, \dots, q_n)$ появляются в результате некоторого эксперимента (опыта) с вероятностями этих значений $(p_1, p_2, \dots, p_n)$; Вот полная запись определения случайной переменной, которая называется *законом распределения*:

$$u = \begin{Bmatrix} q_1, q_2, \dots, q_n \\ p_1, p_2, \dots, p_n \end{Bmatrix}.$$

Закон распределения случайной переменной называют *диффиренциальным законом или вероятностной функцией,* а в ситуации непрерывной случайно велечины - плотностью вероятности.

Нормальный закон распределения

Нормальный закон (Муавра-Гаусса). Имеет уравнение:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

Где μ – математическое ожидание, σ – среднее квадратическое отклонение.

8. Случайная переменная и закон её распределения. Распределение хи-квадрат.

Переменная x, с областью возможных значений X, называется случайной, если каждое ее значение суть результат случайного события $A: x = q$, где q – элемент множества X.

Законом распределения случайной величины x называется скалярная функция $P_x(q)$скалярного аргумента q, которая определена на всей числовой оси и характеризует объективную возможность (вероятность) появления в опыте события $x = q$. (Объяснение от Бывшего смотри **пункт 7**).

Распределение χ^2 **(хи-квадрат)** с n степенями свободы — это распределение суммы квадратов n независимых стандартных нормальных случайных величин.

Пусть $X_1, \dots, X_n$ — совместно независимые стандартные нормальные случайные величины, то есть: $X_i \sim N(0,1)$. Тогда случайная величина $Y = X_1^2 + \cdots + X_n^2$ имеет распределение хи-квадрат с n степенями свободы, обозначаемое $\chi^2(n)$.

9. Случайная переменная и закон её распределения. Распределение Стьюдента Квантиль t крит уровня и её расчёт в Excel.

Переменная x, с областью возможных значений X, называется случайной, если каждое ее значение суть результат случайного события $A: x = q$, где q – элемент множества X.

Законом распределения случайной величины x называется скалярная функция $P_x(q)$скалярного аргумента q, которая определена на всей числовой оси и характеризует объективную возможность (вероятность) появления в опыте события $x = q$. (Объяснение от Бывшего смотри **пункт 7**).

Пусть $Y_0, Y_1, \dots, Y_n$ – независимые стандартные нормальные случайные величины, такие что $Y_i \sim N(0,1), i = 1, \dots, n$. Тогда распределение случайной величины t, где

$$t = \frac{Y_0}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} Y_i^2}}$$

называется распределением Стьюдента с n степенями свободы. Пишут $t \sim t(n)$. Её распределение абсолютно непрерывно и имеет плотность:

$$f_t(y) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n}\,\Gamma\left(\frac{n}{2}\right)}\left(1 + \frac{y^2}{n}\right)^{-\frac{n+1}{2}}$$

, где Γ – гамма-функция Эйлера.

Пусть F_n — функция распределения Стьюдента $t(n)$ с n степенями свободы, и $\alpha \in [0,1]$. Тогда α – квантилью этого распределения называется число $t_{\alpha,n}$ такое, что $F_n(t_{\alpha,n}) = 1 - \alpha$.

Расчет в Excel: =СТЬЮДЕНТ.ОБР.2Х(вероятность,степени_свободы) (T.INV.2T())

Вероятность: $1-\alpha = 1-0,95$
Степени свободы: $n-(k+1)$.

10. Ковариация $Cov(x,y)$ и коэффициент корреляции, $Cor(x,y)$ пары случайных переменных (x,y).

Пусть x и y пара случайных переменных. Характеристика взаимосвязи расчитывается по формуле и называется ковариацией:

$$Cov(x,y) = E(x\cdot y)\; - E(x)\cdot E(y)$$

Если ковариация положительная, то с ростом x возрастает y и наоборот. Если x и y независимые, то ковариация равна 0.

Нормированная ковариация вычисляется по формуле и носит название коэффициента корреляции:

$$Cor(x,\ y)\; = \rho_{x,y}\; = \frac{\sigma_{x,y}}{\sigma_x\cdot\sigma_y}$$

11. Случайная переменная и закон её распределения. Закон распределения Фишера. Квантиль $F_{\text{крит}}$ уровня и её расчёт в Excel.

Переменная x, с областью возможных значений X, называется случайной, если каждое ее значение суть результат случайного события A: $x=q$, где q – элемент множества X.

Законом распределения случайной величины x называется скалярная функция $P_x(q)$скалярного аргумента q, которая определена на всей числовой оси и характеризует объективную возможность (вероятность) появления в опыте события $x=q$. (Объяснение от Бывшего смотри **пункт 7**).

Пусть Y_1, Y_2 – две независимые случайные величины, имеющие распределение хи-квадрат: $Y_i \sim \chi^2(d_i)$, где $d_i \in \mathbb{N}, i=1,2$. Тогда распределение случайной величины $F = \dfrac{Y_1/d_1}{Y_2/d_2}$, называется распределением Фишера со степенями свободы d_1 и d_2. Пишут $F \sim \mathrm{F}(d_1,d_2)$.

$$GQ = \frac{ESS_1}{ESS_2} \sim P_F(q) \tag{1}$$

Эта дробь является <u>статистикой</u> критерия проверяемой гипотезы о гомоскедастичности случайного возмущения. Величина GQ имеет распредение Фишера с кол-ом степеней свободы m,n.

Гипотеза о гомоскедастичности принимается как не противоречащая реальным данным, если оказываются справедливыми следующие два неравенства:

$$\begin{cases} GQ \overset{?}{\leq} F_{\text{крит}} \\ \dfrac{1}{GQ} \overset{?}{\leq} F_{\text{крит}} \end{cases}$$

Где символом $F_{\text{крит}}$ мы обозначаем квантиль распределения Фишера заданного уровня $1-\alpha$, например $1-\alpha\ = 0.95$.
Расчет в Excel: =F.ОБР(вероятность; степени_свободы1; степени_свободы2) (F.INV())
Вероятность: $1-\alpha = 1-0,05$
Степени_свободы1: m_1
Степени_свободы2: m_2

12. Случайный вектор и его основные количественные характеристики. Случайный вектор левых частей схемы Гаусса–Маркова при гомоскедастичном неавтокоррелированном случайном возмущении.
Упорядоченный набор случайных переменных принято называвать ***случайным вектором***:

$$\vec{x}^T\ =\ (x_1,\ x_2\ \dots,\ x_n) \tag{2}$$

Для практики важны следующие две случайные характеристики:
1. Математической ожидание случайного вектора

$$E\left(\vec{x}^T\right)\ =\ \vec{m}_{\vec{x}}^T\ =\ (m_1, m_2,\ \dots,\ m_n) \tag{3}$$

– это веткор из математический ожиданий случайных компонент.
Математическое ожидание – это среднее значение. Математическое ожидание – это *константа*.
2. Коварионная матрица:

$$Cov(\vec{u},\vec{u})\ =\ \Omega_{\vec{x}}\ =\ \begin{pmatrix} {\sigma_1}^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & {\sigma_2}^2 & \dots & \sigma_{2,n} \\ \dots & \dots & \ddots & \dots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_n^2 \end{pmatrix} \tag{4}$$

так принято называть квадратную симметричную матрицу, на главной диагонали которой располагаются дисперсии компонент случайного вектора, а недиагональные

элементы – это ковариации компонент. Ковариация, например σ_{1n}, это *константа* характризующая взаимосвязь компоненты x_1 и x_n. Если x_1 и x_n независимые, то $\sigma_{1n}\ =\ 0$.
Компактная запись схемы Гаусса-Маркова:

$$\vec{y} = X\vec{a} + \vec{u}$$
$$\text{, где :}$$
$$\vec{y} = (y_1,\ \dots, y_n)^T$$
$$\vec{u} = (u_1,\ \dots, u_n)^T$$
$$\vec{a} = (a_0, a_1,\ \dots, a_n)^T$$
$$X = \begin{pmatrix} 1 & x_{11} & \dots & x_{1k} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{nk} \end{pmatrix}$$

Свойство операции вычисления ожидаемого вектора: если обобщить свойство

$E(c_1x + c_2y) = c_1E(x) + c_2E(y)$ на аффинное преобразование случайного вектора $\vec{a}$ в случайный вектор $\vec{y} = X\vec{a} + \vec{u}$, то оно примет вид $E(\vec{y}) = X\vec{a}$.
Свойство операции вычисления ковариационной матрицы случайного вектора: Если же обобщить свойство $Var(c_1x + c_2y) = c_1^2Var(x) + c_2^2Var(y) + 2c_1c_2Cov(x, y) = \vec{c}^T Cov(\vec{x}, \vec{x})\vec{c}$ на аффинное преобразование случайного вектора $\vec{a}$ в случайный вектор $\vec{y} = X\vec{a} + \vec{u}$, то оно примет вид $Cov(\vec{y}, \vec{y}) = Cov(\vec{u}, \vec{u}) = \sigma_u^2 \cdot I$.

13. Основные количественные характеристики аффинного преобразования случайного вектора (на примере вектора МНК – оценок коэффициентов линейной модели при гомоскедастичном неавтокоррелированном случайном возмущении).

Афинное преобразование – это линейное неоднородное преобразование. Пусть символом $\vec{x}$ обозначен случайный вектор. Аффинным преобразованием этого вектора принято называть вектор $\vec{y}$, который вычисляется по следующему правилу:

$$\vec{y} = A \cdot \vec{x} + \vec{b};$$

Здесь символом A обозначена матрица коэффициентов, символом $\vec{b}$ обозначен вектор свободных членов.
Отметим правила расчёта основных характеристик аффинного преобразования:

$$E(\vec{y}) = A \cdot E(\vec{x}) + \vec{b};$$
$$Cov(\vec{y}, \vec{y}) = A \cdot Cov(\vec{x}, \vec{x}) \cdot A^T;$$

При сделанных предположениях (предпосылках) теоремы Гаусса–Маркова оптимальные оценки коэффициентов функции регрессии вычисляются по правилу:

$$\vec{\tilde{a}} = \left(X^T \cdot X\right)^{-1} \cdot X^T \cdot \vec{y}; \qquad (*)$$

Вышеуказанные правила расчета основных характеристик аффинного преобразования случайного вектора дают основу определения математического ожидания вектора оценок коэффициентов.
В правой части уравнения (*) матрица X является константой или фиксированной величиной, а вектор $\vec{y}$ является случайным значением в силу того, что:

$$\vec{y} = X \cdot \vec{a} + \vec{u}; \qquad (**)$$

в правой части (**) первое слагаемое $X \cdot \vec{a} = \alpha$ – это вектор констант, а второй вектор случайный $\vec{u}$ и мы можем трактовать вектор $\vec{y}$ выражения (**), как афинное преобразование ветора $\vec{u}$.
Следовательно, так как $\vec{y}$ является случайным, то и оптимальные оценки коэффициентов вектора $\vec{a}$ из уравнения коэффициентов (**), так же будут случайными оценками.
Найдём математическое ожидание от оценок коэффициентов $\vec{a}$.

Для нахождения математического ожидания от оценок коэффициентов $\vec{a}$. Перепишем уравнение (*) в другую форму:

$$\tilde{\vec{a}} = \left(X^T X\right)^{-1} \cdot X^T \cdot (X \vec{a} + \vec{u}) =$$
$$= \left(X^T X\right)^{-1} \cdot \left(X^T X\right) \cdot \vec{a} + \left(X^T X\right)^{-1} X^T \vec{u} =$$
$$= \left[\text{Вспомним, что } \left(X^T X\right)^{-1} \cdot \left(X^T X\right) = I\left(\text{еденичной матрице}\right)\right] =$$
$$= \vec{a} + \left(X^T X\right)^{-1} X^T \vec{u}$$

Возьмём математическое ожидание от левой и правой части и получим, вспоминая свойство вектора $\vec{u}$, а именно, что $E(\vec{u}) = 0$:

$$E\left(\tilde{\vec{a}}\right) = E(\vec{a}) + E\left(\left(X^T X\right)^{-1} X^T \vec{u}\right) = E(\vec{a}) + \left(X^T X\right)^{-1} X^T E(\vec{u}) = E(\vec{a}) = \vec{a}$$

Коварицiонная матрица вектора $\tilde{\vec{a}}$ по определению равна:

$$Cov\left(\tilde{\vec{a}}, \tilde{\vec{a}}\right) = E\left(\left(\tilde{\vec{a}} - \vec{a}\right)\left(\tilde{\vec{a}} - \vec{a}\right)^T\right) = \begin{pmatrix} \sigma_{\tilde{a}_1}^{\ 2} & \sigma_{\tilde{a}_1, \tilde{a}_2} & \dots & \sigma_{\tilde{a}_1, \tilde{a}_n} \\ \sigma_{\tilde{a}_2, \tilde{a}_1} & \sigma_{\tilde{a}_2}^{\ 2} & \dots & \sigma_{\tilde{a}_2, \tilde{a}_n} \\ \dots & \dots & \ddots & \dots \\ \sigma_{\tilde{a}_n, \tilde{a}_1} & \sigma_{\tilde{a}_n, \tilde{a}_2} & \dots & \sigma_{\tilde{a}_n}^{\ 2} \end{pmatrix}$$

Её диагональные элементы равны $\sigma_{\tilde{a}_1} = Var\left(\tilde{a}_1\right)$ дисперсиям оценок отдельных коэффициентов. А диагональные элементы равны ковариациям оценок $\sigma_{\tilde{a}_1, \tilde{a}_2} = Cov\left(\tilde{a}_1, \tilde{a}_2\right)$. Заметим, что $\sigma_{\tilde{a}_i, \tilde{a}_j} = \sigma_{\tilde{a}_j, \tilde{a}_i}$, то есть матрица $Cov\left(\tilde{\vec{a}}, \tilde{\vec{a}}\right)$ симметричная относительно главной диагонали. Далее по выведенной нами формуле вычислим:

$$\tilde{\vec{a}} - \vec{a} = \left(X^T X\right)^{-1} X^T \vec{u}$$

И подставим в формулу ковариации выше получим:

$$E\left(\left(\left(X^T X\right)^{-1} X^T \vec{u}\right)\left(\left(X^T X\right)^{-1} X^T \vec{u}\right)^T\right) =$$
$$= E\left(\left(X^T X\right)^{-1} X^T \left(\vec{u} \cdot \vec{u}^T\right) X \left(X^T X\right)^{-1}\right) =$$
$$= \left(X^T X\right)^{-1} X^T E\left(\vec{u} \cdot \vec{u}^T\right) X \left(X^T X\right)^{-1} =$$
$$= \left[\text{вспомним, что } E\left(\vec{u} \cdot \vec{u}^T\right) = \sigma_u^2 \cdot I\left(\text{еденичная матрица}\right)\right] =$$
$$= \sigma^2 \left(X^T X\right)^{-1} X^T X \left(X^T X\right)^{-1} = \sigma^2 \left(X^T X\right)^{-1}.$$

Следовательно ковариация равна:

$$Cov\left(\tilde{\vec{a}}, \tilde{\vec{a}}\right) = \sigma^2 \left(X^T X\right)^{-1} = \sigma^2 \cdot Q$$

14. Случайный вектор, веса компонент случайного вектора и факторизация его ковариационной матрицы. Случайный вектор в схеме Гаусса – Маркова при гетероскедастичном неавтокоррелированном случайном возмущении.

Упорядоченный набор случайных переменных принято называть ***случайным вектором***:

$$\vec{x}^T = (x_1, x_2 \ldots, x_n)$$

Веса компонент случайного вектора и факторизация его коварицонной матрицы

Пусть $\vec{x}^T = (x_1, x_2 \ldots, x_n)$ случайный вектор, пусть x_i какая-то компонента вектора; вес компонент x_i – это константа, которая вычисляется по следующему правилу:

$$p_i = \frac{\sigma_0^2}{\sigma_i^2}$$

где σ_0^2 обозначена произвольная, но фиксированная положительная константа. Тогда

$$\Omega_{\vec{u}} = \sigma_0^2 \cdot \begin{pmatrix} \frac{1}{\sqrt{p_1}} & 0 & \ldots & 0 \\ 0 & \frac{1}{\sqrt{p_2}} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \frac{1}{\sqrt{p_n}} \end{pmatrix}^2$$

Ковариационная матрица вектора в этой схеме является диагональной, но диагональные элементы (дисперсии случайных остатков) этой матрицы теперь ***неодинаковы*** (следует от гетероскедастичности).

Составим уравнения наблюдений в рамках трансформированной модели:

$$\begin{cases} \sqrt{p} \cdot y = \sqrt{p} \cdot a_0 + \sqrt{p} \cdot a_1 x_1 + \sqrt{p} \cdot a_2 x_2 + \ldots + \sqrt{p} \cdot a_k x_k + \sqrt{p} \cdot u \\ E\left(\sqrt{p} \cdot u\right) = 0; \; E\left(\left(\sqrt{p} \cdot u\right)^2\right) = \sigma_0^2 \end{cases}$$

$$v = \sqrt{p} \cdot u$$

$$\begin{cases} \sqrt{p} \cdot y = \sqrt{p} \cdot a_0 + \sqrt{p} \cdot a_1 x_1 + \sqrt{p} \cdot a_2 x_2 + \ldots + \sqrt{p} \cdot a_k x_k + v \\ E(v) = 0; \; E\left(v^2\right) = \sigma_0^2 \end{cases}$$

$$\underbrace{\boxed{P^{\frac{1}{2}} \cdot \vec{y}}}_{\vec{y}'} = \underbrace{\boxed{P^{\frac{1}{2}} \cdot X}}_{X'} \cdot \vec{a} + \vec{v} \qquad (6.4.5)$$

Символом $P^{\frac{1}{2}}$ обозначена следующая квадратная матрица:

$$P^{\frac{1}{2}} = \begin{pmatrix} \frac{1}{\sqrt{p_1}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{p_2}} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \frac{1}{\sqrt{p_n}} \end{pmatrix}$$

$$\Omega_{\vec{v}} = \sigma_0^2 \cdot E - \text{скалярная}$$

$$\Omega_{\vec{u}} = \sigma_0^2 \cdot P^{-1} \tag{6.4.6}$$

Получается, что уравнения наблюдений (6.4.5) удовлетворяют всем предпосылкам и это значенит, что по всем этим уравнениям мы можем оценить параметры с помощью МНК.

Утверждение A), B), C), D) теоремы Гаусса-Маркова применительно к уравнениям наблюдения 6.4.5 превращаются в следующие утверждения:

$$A)\tilde{\vec{a}} = \left(X^T \cdot P \cdot X\right)^{-1} \cdot X^T \cdot P \cdot \vec{y} = Q \cdot X^T \cdot P \cdot \vec{y} \tag{6.4.7}$$

$$B)\tilde{\sigma}_0^2 = \frac{\sum_{i=1}^{n} \tilde{v}_i^2}{n-(k+1)} = \frac{\sum_{i=1}^{n} p_i \cdot \tilde{u}_i^2}{n-(k+1)} \tag{6.4.8}$$

$$\tilde{u}_i = y_i - \left(\tilde{a}_0 + \tilde{a}_1 \cdot x_{1,i} + \ldots + \tilde{a}_k \cdot x_{k,i}\right) \tag{6.4.9}$$

$$\tilde{v}_i = \sqrt{p_i} \cdot \tilde{u}_i \tag{6.4.10}$$

$$C) \sum_{i=1}^{n} \tilde{v}_i^2 = \sum_{i=1}^{n} p_i \cdot \tilde{u}_i^2 \to \min \tag{6.4.11}$$

$$D) \begin{cases} S\tilde{a}_j = \tilde{\sigma}_0 \cdot \sqrt{q_{j+1\ j+1}} \\ j = 0, 1, \dots, k \end{cases} \tag{6.4.12}$$

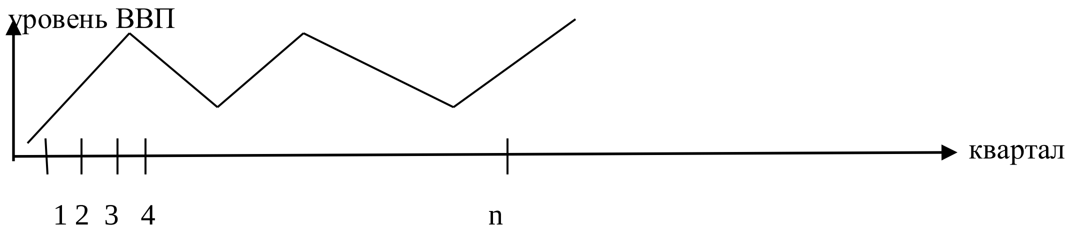
Свойство $C)$ оценок из утверждения $A)$ принято называть *взвешанными наименьшими квадратами,* что является причиной общепринятого названия формулы процедуры (6.4.7) ВМНК.

15. Временной ряд и его структура (На примере ВВП России).

Экономическая переменная $\tilde{y}$, датированная дискретными моментами времени, называется временным рядом.

Познакомимся с понятием временного ряда, зайдя на сайт Госкомстата или Росстат. Важным примером временных рядов являются квартальные уровни ВВП страны. Эти уровни экономисты называют значениями временного ряда. Аргументом временного ряда является время.

По данным Госкомстата построим график квартальных уровней ВВП. Он имеет следующую структуру:



Структура уровней временного ряда:

1) Повторяющийся из года в год закон изменения квартальных уровней. В 1 квартале ВВП самый низкий, в 4 самый высокий. Для каждого года вывод справедлив. Это означает, что в структуре квартальных уровней присутствует сезонная составляющая.
2) В уровнях ВВП можно увидеть и тенденцию (тренд), которую можно расчленить на следующие промежутки времени:

   2011-2013 ВВП восходило/высокие цены на нефть

   2014-2016 тенденция приобрела отрицательный характер/санкции западных стран

   С 2017 тенденция приобрела восходящий характер/экономика адаптируется к санкциям

3) Рассматривая график в крупном масштабе, мы можем обнаружить небольшие хаотичные изменения геометрии графика. Это означает, что в структуре уровней ряда присутствует случайная составляющая

Вывод: временный ряд- это датированная дискретными моментами времени количественная характеристика изучаемого объекта и в уровнях временного ряда можно выделить следующие составляющие: тренд T(t), сезонность S(t), U_t случайная составляющая.

Если просуммировать эти составляющие, то модель называется аддитивной, вот ее вид $y_t = \mathrm{T(t)} + S(t) + C(t) + U_t$, где $C(t)$ - циклическая составляющая, имеющая период отличный от года.

16. Модели тренда временного ряда.

Конкретизируем уравнение тренда $\mathrm{T(t)}$.

Наиболее популярные модели тренда:

1) Линейная функция времени $T(t) = a_0 + a_1 t$
2) Квадратичная парабола времени $T(t) = a_0 + a_1 t + a_2 t^2$
3) Парабола 3 порядка времени $\mathrm{T(t)} = a_0 + a_1 t + a_2 t^2 + a_3 t^3$

17. Моделирование сезонной составляющей при помощи фиктивных переменных.

Сезонная составляющая – это некоторая периодическая функция времени с периодом в один год. Обозначим символом τ количество единичных отрезков времени (неделя, месяц, квартал) образующих год. Тогда сезонная составляющая удовлетворяет равенству

$$S(t+\tau) = S(t)$$

Можно проверить, что следующая функция $S(t) = b_1 d_1(t) + b_2 d_2(t) + \cdots + b_{\tau-1} d_{\tau-1}(t)$ **(*)** является периодической с периодом τ

$b_1, \ldots, b_{\tau-1}$- константы (коэффициенты), которые у каждого временного ряда свои.

Символами $d_1(t), \ldots, d_{\tau-1}(t)$ обозначены индикаторы единичных отрезков времени, образующих год.

Поясним суть индикаторов, когда $\tau = 4$.

Так $d_1(t)$- это индикатор первого квартала, т.е. $d_1(t)$ принимает значение 1, если t соответствует 1 кварталу.

d_1 ={1 - для первого квартала, 0 - для других кварталов};
d_2 ={1 - для второго квартала, 0 - для других кварталов};
d_3 ={1 - для третьего квартала, 0 - для других кварталов}

При помощи модели (*) можно моделировать не только сезонную составляющую, но и влияние на соответствующую эндогенную переменную качественного фактора, который способен находиться в одном из τ состояний. Состояние этого фактора, при котором все фиктивные переменные равны 0 называется базовым (в нашем примере - это четвертый квартал года).

Спецификация квартальных уровней ВВП России:

$$\begin{cases} Y_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3 + b_1 d_1(t) + b_2 d_2(t) + b_3 d_3(t) + U_t \\ E(U_t) = 0;\ E\left(U_t{}^2\right) = \sigma_U^2 \end{cases}$$

18. Регрессионная зависимость случайных переменных. Функция регрессии, стандартные модели функции регрессии.

Функцией регрессии y на x (обозначается символом $E(y|x)$) называется ожидаемое значение случайной переменной y, вычисленное при заданном значении переменной x, т.е.

$$E(y|x) = \begin{cases} \sum_{1}^{n} r_i P_y(r_i|x) \text{для дискретной } y, \\ \int_a^b r P_y(r|x) dr \text{ для непрерывной } y \end{cases}$$

Величина $E(y|x)$ является функцией аргумента x. Эта функция позволяет представить случайную переменную у в виде

$$y = E(y|x) + u \quad \textbf{(1)},$$

где u – случайная переменная, такая, что $E(u|x) = 0$ **(2)**.

Разложение **(1)** случайной переменной y со свойством **(2)** именуется регрессионным анализом переменной y. Функция регрессии $E(y|x)$ интерпретируется в экономике как выраженный математическим языком экономический закон, по которому изменяется объясняемая (эндогенная) переменная y в ответ на изменения объясняющей (экзогенной) переменной x.

Простейшие модели функции регрессии:

1. Линейная функция $f(x) = a_0 + a_1x$
2. Парабола второго порядка $f(x) = a_0 + a_1x + a_2x^2$
3. Степенная функция $f(x) = a_0x^{a_1}$
4. Показательная функция $f(x) = a_0e^{a_1x}$

19. Схема Гаусса–Маркова.

Линейная модель множественной регрессии имеет спецификацию, которая включает в себя следующие параметры: $(a_0, a_1, \dots, a_k, \sigma_u)$.

Приступим к обсуждению статистической процедуры оценивания этих параметров.

Разместим обучающую выборку при построении лин. модели множественной регрессии в следующей таблице:

№	y	X1	X2	…	Xk
1	Y1	X11	X21	…	Xk,1
2	Y2	X12	X22	…	Xk,2
...	…	…	…	…	…
n	Yn	X1,n	X2,n	…	Xk,n

Подставляем каждую строку в уравнение линейной модели множественной регрессии. Получим систему уравнений наблюдений:

$$\begin{cases} y_1 = a_0 + a_1x_{1,1} + a_2x_{2,1} + \dots + a_kx_{k,1} + u_1, \\ \dots \\ y_n = a_0 + a_1x_{1,n} + a_2x_{2,n} + \dots + a_kx_{k,n} + u_n \end{cases}$$

Ее принято называть схемой Гаусса-Маркова.

Вот компактная запись этой схемы:

$$\vec{y} = X\vec{a} + \vec{u},$$

X- матрица объясняющих переменных, расширенная столбцом единиц(если есть свободный член)

$\vec{a}$-вектор коэффициентов модели

$\vec{u}$-вектор случайных возмущений

20. Понятие статистической процедуры оценивания параметров эконометрической модели. Линейные статистические процедуры. Требования к наилучшей статистической процедуре.

Рассмотрим лаконичную запись эконометрической модели (линейной модели множественной регрессии):

$F(y_t, \overrightarrow{x_t}; \vec{p}) = u_t$

взаимосвязь параметры

Пусть известна обучающая выборка. Статистической процедурой оценивания параметра принято называть некоторую функцию $\varphi(\vec{y}; X)$ выборки, значением этой функции являются оценки параметров модели.

$$\tilde{\vec{p}} = \begin{pmatrix} \tilde{\vec{a}} \\ \tilde{\sigma}_u^2 \end{pmatrix} = \varphi(\vec{y}; X)$$

Процедура φ называется оптимальной в заданном класе функций, если доставляемые ею оценки параметров обладают следующими свойствами:

$$\begin{cases} E(\tilde{\vec{p}}) = \vec{p} \\ Var(\tilde{p}_j) \to min \end{cases} \quad *$$

Первое св-во означает мат. Ожидание оценок параметров совпадает с истинными значениями параметров. В мат. Статистике такие оценки называют несмещенными.

Второе св-во означает, что разброс оценок параметров относительно истинных значений минимален.

Вывод, Статистическая процедура оценивания модели – это некоторая функция выборки, значением этой функции служат оценки параметров. Процедура оптимальна в заданном классе функций, если ее значения удовлетворяют *

21. **Теорема Гаусса-Маркова: выражение вектора оценок коэффициентов $\tilde{a}$ и доказательство их несмещённости.**

Если справедливы все предпосылки теоремы Гаусса-Маркова, тогда имеет место утверждение $\boldsymbol{A}$: наилучшая оценка коэффициентов модели рассчитывается по правилу $\tilde{\vec{a}} = (X^T X)^{-1} X^T \vec{y} = Q X^T \vec{y} = M\vec{y}$.
Докажем, что имеет место свойство несмещённости оценок коэффициентов, то есть $E(\tilde{\vec{a}}) = \vec{a}$.
Доказательство.

$E(\tilde{\vec{a}}) = E((X^T X)^{-1} X^T \vec{y}) = E(M\vec{y}) = ME(\vec{y}) = MX\vec{a} = (X^T X)^{-1} X^T X\vec{a} = \vec{a}$, ч.т.д.

$E(\vec{y}) = E(X\vec{a} + \vec{u}) = E(X\vec{a}) + E(\vec{u}) = \{E(\vec{u}) = 0\} = E(X\vec{a}) = \{\text{вектор} - \text{константа}\} = X\vec{a}$.

22. Теорема Гаусса-Маркова: выражение $Cov\left(\tilde{a}, \tilde{a}\right)$ и его обоснование.

$Cov\left(\tilde{a}, \tilde{a}\right) = \tilde{\sigma}_u^2 \cdot \left(X^T \cdot X\right)^{-1} = \tilde{\sigma}_u^2 \cdot Q$

Доказательство:

$\vec{\tilde{a}} = \left(X^T P X\right)^{-1} X^T P \vec{y} = M \cdot \vec{y}$ – оценивание вектора $\vec{a}$; $\vec{\tilde{a}}$ – линейное преобразование $\vec{y}$.

$\vec{\tilde{a}} = A * \vec{y} + \vec{b}$, где $A = M$, $\vec{b} = \vec{0}$

Тогда по теореме Фишера

$$Cov\left(\tilde{a}, \tilde{a}\right) = M \cdot Cov(\vec{y}, \vec{y}) \cdot M^T = \left(X^T P X^{-1}\right) \cdot X^T P \cdot \left(\sigma_0^2 P^{-1}\right) X \left(X^T P X\right)^{-1} =$$
$$= \sigma_0^2 \cdot \left(X^T \cdot X\right)^{-1} = \sigma_0^2 \cdot Q \ ■$$

Более подробное доказательство смотри пункт 13

23. Теорема Гаусса-Маркова: предпосылки и свойство наименьших квадратов $\tilde{u}^T \cdot \tilde{u} \rightarrow min.$

Пусть в уравнениях наблюдений $\vec{y} = X\vec{a} + \vec{u}$:

0. Столбцы X линейно независимы;
1. $E(u_1) = \ldots = E(u_n) = 0$;
2. $Var(u_1) = \ldots = Var(u_n) = \sigma_u^2$; - не зависят от объясняющих переменных
3. $Cov(u_i, u_j) \neq 0, i \neq j$; - Случайные остатки попарно некоррелированные
4. $Cov(u_i, x_{mj}) = 0$. - Значения объясняющих переменных не коррелированы со значениями случайных возмущений

Тогда выполняются необходимые утверждения (не все, только те, которые требуются в вопросе):

A) $\vec{\tilde{a}} = \left(X^T X\right)^{-1} X^T \vec{y}$ – оптимальная линейная процедура оценивания коэффициентов функции регрессии.

C) Оценки, вычисленные в A, обладают замечательным свойством наименьших квадратов, то есть $\sum_{i=1}^{n} \tilde{u}_i^2 \rightarrow min$. Именно это свойство является причиной общепринятого названия процедуры A – МНК.

24. Теорема Гаусса-Маркова: выражение $\tilde{\sigma}_0^2$

2-я предпосылка теоремы Гаусса-Маркова о гомоскедастичности случайного остатка не выполнена, то есть дисперсия зависит от объясняющих переменных, а остаток гетероскедастичен. В таком случае оценки параметров модели утрачивают свое свойство оптимальности (свойство минимальных дисперсий). Для построения оптимальной процедуры оценивания модели с гетероскедастичным остатком потребуется модель гетероскедастичности остатка, вот простейший вид такой модели:

$$Var(u) = \sigma_u^2 = \sigma_0^2 \left(\sum_{j=0}^{k} |x_j| \right)^{\lambda}$$

σ_0^2 имеет смысл дисперсии такой случайной величины, вес которой равен 1, поэтому называется дисперсией единицы веса.

λ-некоторое априорно заданное число (подбирается эксперементально)

$p = \dfrac{\sigma_0^2}{\sigma_u^2}$ – вес случайного возмущения

Величина $\tilde{\sigma}_0^2 = \dfrac{\vec{\tilde{u}}^T P \cdot \vec{\tilde{u}}}{n-(k+1)} = \dfrac{\sum_{i=1}^{n} p_i \cdot \tilde{u}_i^2}{n-(k+1)}$ - несмещенная оценка σ_0^2

25. Взвешенный метод наименьших квадратов (ВМНК). Практическая реализация ВМНК.

В данном случае предпосылка 2 теоремы Гаусса-Маркова нарушается (случайное возмущение гетероскедастично).

Алгоритм взвешенного метода наименьших квадратов (ВМНК) состоит в предварительной трансформации ЛММР с гетероскедастичным остатком к модели с гомоскедастичным остатком, далее проверке гомоскедастичности остатка в трансформированной модели и, наконец, в применении процедуры МНК.

$\vec{y} = X\vec{a} + \vec{u}$ - уравнения наблюдений

Ковариационная матрица имеет следующий вид:

$$\Omega_{\vec{u}} = \sigma_0^2 P^{-1} = \sigma_0^2 \cdot \begin{pmatrix} \frac{1}{p_1} & 0 & \dots & 0 \\ 0 & \frac{1}{p_2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \frac{1}{p_n} \end{pmatrix}$$

Составим уравнения наблюдений:

$$\begin{cases} \sqrt{p} \cdot y = \sqrt{p} \cdot a_0 + \sqrt{p} \cdot a_1 x_1 + \sqrt{p} \cdot a_2 x_2 + \ldots + \sqrt{p} \cdot a_k x_k + \sqrt{p} \cdot u \\ E\left(\sqrt{p} \cdot u\right) = 0; \;\; E\left(\left(\sqrt{p} \cdot u\right)^2\right) = \sigma_0^2 \end{cases}$$

$$\underbrace{P^{\frac{1}{2}} \cdot \vec{y}}_{(\vec{y}')} = \underbrace{P^{\frac{1}{2}} \cdot X}_{(X')} \cdot \vec{a} + \vec{v} \qquad (1)$$

$$P^{\frac{1}{2}} = \begin{pmatrix} \frac{1}{\sqrt{p_1}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{p_2}} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \frac{1}{\sqrt{p_n}} \end{pmatrix}$$

(1) удовлетворяют всем предпосылкам теоремы Гаусса-Маркова, следовательно, можем оценить параметры с помощью МНК.
A-D теоремы Г-М превращаются в следующие утверждения:

$A)\ \vec{\tilde{a}} = \left(X^T \cdot P \cdot X\right)^{-1} \cdot X^T \cdot P \cdot \vec{y} = Q \cdot X^T \cdot P \cdot \vec{y}$

$B)\ \tilde{\sigma}_0^2 = \dfrac{\sum_{i=1}^{n} \tilde{v}_i^2}{n-(k+1)} = \dfrac{\sum_{i=1}^{n} p_i \cdot \tilde{u}_i^2}{n-(k+1)};\ -\text{оценка дисперсии}$

$\tilde{u}_i = y_i - \left(\tilde{a}_0 + \tilde{a}_1 \cdot x_{1,i} + \ldots + \tilde{a}_k \cdot x_{k,i}\right);\ \tilde{v}_i = \sqrt{p_i} \cdot \tilde{u}_i$

$C)\ \sum_{i=1}^{n} \tilde{v}_i^2 = \sum_{i=1}^{n} p_i \cdot \tilde{u}_i^2 \to \min$

$D) \begin{cases} S\tilde{a}_j = \tilde{\sigma}_0 \cdot \sqrt{q_{j+1\,j+1}} \\ j = 0, 1, \ldots, k \end{cases}$

Свойство C) оценок коэффициентов из A) принято называть *взвешанными наименьшими квадратами,* откуда A) - взвешенный метод наименьших квадратов.

26. Обобщённый метод наименьших квадратов (ОМНК).

Приступаем к доказательству утверждений теоремы Гаусса-Маркова. Мы расширим предпосылки с номерами 2 и 3 этой теоремы отказавшись от них и предполагая, что вектор случайных возмущений $\vec{u}$ имеет нулевое математическое ожидание и коварицонную матрицу $Cov\,(\vec{u}, \vec{u}) = \sigma_0^2 \cdot P^{-1}$.

Докажем утверждение А:

1) Мы будем разыскивать оценку вектора $\vec{a}$ в классе всех линейных функций определённых на векторе значений эндогенной переменной $\vec{y}$, так что определению подлежит матрица M этого линейного преобразования мы собираемся разыскать M.

$$\vec{\tilde{a}} = M \cdot \vec{y}$$

2) Поиск матрицы M удобно осуществить создавая оптимальную статистическую процедуру оцениевания значения y_0 произвольной линейной функцией вектора

коэффициентов модели

$$y_0 = \vec{x}_0^T \cdot \vec{a}$$

3) Процедуру оценивания числа y_0 мы будем отыскивать в классе линейных функций $\vec{y}$, где $\vec{m}$ – это строка линейных коэффициентов.

$$\tilde{y}_0 = \vec{m}_0 \cdot \vec{y}$$

И будем отыскивать опираясь на два требования оптимальности:

$$\begin{cases} E\left(\tilde{y}_0\right) = y_0 = \vec{x}_0^T \cdot \vec{a} \\ Var\left(\tilde{y}_0\right) \to \min \end{cases}$$

Вычислим мат ожидание символа $\tilde{y}_0$. Первое требование оптимальности приводит к следующим уравнению отностительно искомых коэффициентов $\vec{m}$:

$$E\left(\tilde{y}_0\right) = \vec{m} \cdot X \cdot \vec{a}, \ \Rightarrow \vec{m}^T \cdot X = \vec{a} = \vec{x}_0^T \cdot \vec{a}$$

Теперь найдём дисперсию опираясь на следующее выражение:
$Cov\,(\vec{u}, \vec{u}) = \sigma_0^2 \cdot P^{-1}$.

Значит дисперсия: $Var\left(\tilde{y}_0\right) = \sigma_0^2 \cdot \vec{m}^T \cdot P^{-1} \cdot \vec{m}$

Следовательно искомые оценки коэффициентов нужно найти в процессе решения следующей оптимизационной задачи:

$$\begin{cases} \vec{m}^T \cdot P^{-1} \cdot \vec{m} \to \min \\ X^T \cdot \vec{m} = \vec{x}_0 \end{cases} \tag{5.11'}$$

Задачу 5.11' решаем методом Лагранжа:
Шаг 1. Составим функция Лагранжа:

$$L(\vec{m}, \vec{l}) = \vec{m}^T \cdot P^{-1} \cdot \vec{m} + \vec{l}^T \cdot \left(\vec{x}_0 - X^T \cdot \vec{m}\right)$$

символом l обозначен множитель Лагранжа.
Шаг 2. Составим необходимые условия экстремума функции Лагранжа:

$$\begin{cases} \dfrac{\partial L}{\partial \vec{m}} = 2 \cdot P^{-1} \cdot \vec{m} - X \cdot \vec{l} = 0, \\ \dfrac{\partial L}{\partial l} = \vec{x}_0 - X^T \cdot \vec{m} = 0. \end{cases} \tag{5.14}$$

Шаг 3. Система 5.14 решается аналитически или методом подстановки:

$$\vec{m} = P \cdot X \cdot \left(X^T \cdot P \cdot X\right)^{-1} \cdot \vec{x}_0 \tag{5.15}$$

Следовательно, наилучшая оценка расчитавается по правилу 5.16:

$$\tilde{y}_0 = \vec{m}^T \cdot \vec{y} = \vec{x}_0^T \cdot \left(X^T \cdot P \cdot X\right)^{-1} \cdot \vec{y} \tag{5.16}$$

И последнее действие

$$y_0 = \vec{x}_0^T \cdot \vec{a} \Rightarrow \tilde{\vec{a}} = \left(\left(X^T \cdot P \cdot X\right)^{-1} \cdot X^T\right)(= M) \cdot P \cdot \vec{y} \blacksquare$$

Это и есть ОМНК.

27. Система нормальных уравнений и явный вид её решения при оценивании методом наименьших квадратов (МНК) линейной модели парной регрессии.

Из $\tilde{\vec{a}} = \left(X^T P X\right)^{-1} X^T P \vec{y}$ видно, что $\tilde{\vec{a}}$ вычисляется в процессе решения системы из $k+1$ линейных алгебраических уравнений с $k+1$ неизвестными: $\left(X^T P X\right)\tilde{\vec{a}} = X^T P \vec{y}$.

Эта система называется системой нормальных уравнений.

В ситуации процедуры МНК, т.е. $P = E$ её подробная запись принимает следующий вид:

$$\begin{cases} n\tilde{a}_0 + [x]\tilde{a}_1 = [y] \\ [x]\tilde{a}_0 + \left[x^2\right]\tilde{a}_1 = [xy] \end{cases}$$

где $[x] = \sum_{i=1}^{n} x_i;\ [y] = \sum_{i=1}^{n} y_i;\ \left[x^2\right] = \sum_{i=1}^{n} x_i^2\ ;\ [xy] = \sum_{i=1}^{n} x_i y_i$

Тогда явный вид решения этой системы:

$$\begin{cases} \tilde{a}_1 = \dfrac{Cov(x, y)}{Var(x)};\quad \tilde{a}_0 = \overline{y} - \overline{x}\tilde{a}_1 \end{cases}$$

28. Ковариационная матрица оценок коэффициентов линейной модели парной регрессии: явные выражения $Var\left(\tilde{a}_0\right)$, $Var\left(\tilde{a}_1\right)$, $Cov\left(\tilde{a}_0,\ \tilde{a}_1\right)$.

Линейная модель парной регрессии имеет вид:

$$\begin{cases} y = a_0 + a_1 x + u \\ E(u) = 0,\ Var(u) = \sigma_u^2 \end{cases}$$

$$Cov\left(\tilde{\vec{a}},\ \tilde{\vec{a}}\right) = \begin{pmatrix} Var\left(\tilde{a}_0\right) & Cov\left(\tilde{a}_0,\ \tilde{a}_1\right) \\ Cov\left(\tilde{a}_0,\ \tilde{a}_1\right) & Var\left(\tilde{a}_1\right) \end{pmatrix}$$

$$Cov\left(\tilde{\vec{a}},\ \tilde{\vec{a}}\right) = \sigma_u^2 Q = \sigma_u^2 \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix}$$

Учитывая, что $R = X^T X = \begin{pmatrix} n & [x] \\ [x] & \left[x^2\right] \end{pmatrix}$ - матрица, образованная коэффициентами и свободными членами системы $\begin{cases} n\tilde{a}_0 + [x]\tilde{a}_1 = [y] \\ [x]\tilde{a}_0 + \left[x^2\right]\tilde{a}_1 = [xy] \end{cases}$

Получаем:

$$Var\left(\tilde{a}_0\right) = \sigma^2 Q_{11} = \sigma^2\left(\frac{1}{n} + \frac{\overline{x}^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\right)$$

$$Var\left(\tilde{a}_1\right) = \sigma^2 Q_{22} = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

$$Cov\left(\tilde{a}_0, \tilde{a}_1\right) = \sigma^2 Q_{12} = \sigma^2 Q_{21} = \frac{-\sigma^2\overline{x}}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

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 $E(\Delta\tilde{\vec{a}}) = 0$.

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

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

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:

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

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

$$\begin{cases} y_t = a_0 + a_1 x_{1,t} + \dots + a_k x_{k,t} + u_t \\ E(u_t) = 0, \; u_t \in STS \end{cases} \qquad (1)$$

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):

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

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:

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

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):

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

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:

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

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

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

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:

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

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 $b_0 = ln\, a_0$, $b_1 = a_1$, $b_2 = \beta$ (6) В силу функциональной зависимости коэффициентов b_1 и b_2 уравнение (5) можно упростить: $y = b_0 + b_1 x$, (7) где $y = lnY - ln\, L = ln\frac{Y}{L}$, $x = ln\, K - ln\, L = ln\frac{K}{L}$ (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):

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

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

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

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):

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

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 $y_0 = ln\frac{Y_0}{L_0}$:

$$\begin{cases} \tilde{y}_0 = \tilde{b}_0 + \tilde{b}_1 x_0 \\ S_{\tilde{y}_0} = \tilde{\sigma}_u \sqrt{1 + q_0} \end{cases} (15)$$

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:

$Y = \tilde{a}_0 K^{\tilde{a}_1} L^{1-\tilde{a}_1}$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:

$\tilde{a}_0 = e^{\tilde{b}_0}$, $\tilde{a}_1 = \tilde{b}_1$, $S_{\tilde{a}_0} = \tilde{a}_0 S_{\tilde{b}_0}$, $S_{\tilde{a}_1} = S_{\tilde{b}_0}$ (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:

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

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.

36. **Тест Голдфелда-Квандта гомоскедастичности случайного возмущения в ЛММР.**

Тест для проверки второй предпосылки теоремы Гаусса-Маркова. Предпосылка означает независимость дисперсий случайных возмущений от значений объясняющих переменных. (гомоскедостичность)

2) $Var\left(u_1^2\right) \ = Var\left(u_2^2\right) \ = \ \dots \ = Var\left(u_n^2\right) \ = \sigma_n^2;$

Шаги теста предпосылки №2 Голдфелда-Кванта

Шаг 1. Составляется система уравнений наблюдений объекта

$$\begin{cases} y_1 = a_0 + a_1 \cdot x_{1,1} + a_2 \cdot x_{2,1} + \dots a_k \cdot x_{k,1} + u_1 \\ y_2 = a_0 + a_1 \cdot x_{1,2} + a_2 \cdot x_{2,2} + \dots a_k \cdot x_{k,2} + u_2 \\ \dots \\ y_n = a_0 + a_1 \cdot x_{1,n} + a_2 \cdot x_{2,n} + \dots a_k \cdot x_{k,n} + u_n \end{cases}$$

Шаг 2. Уравения наблюдений упорядычеваются по возрастанию сумм абсолютных значений объясняющих перменных

$$\sum_{j=1}^{k} |x_{ji}|$$

Шаг 3. По первым n_1 упорядоченным уравениям оцениваются методом наименьших квадратов параметры модели и запоминается значения ESS_1. Количество n_1 выбирается согласно следующим двум условиям:

$$a)\, n_1 \ \approx \frac{1}{3}n, \ \ b)\, n_1 \ > \ k \ + \ 1$$

Аналогично оценивается модель по последним n_1 уравениям и запоминается значение ESS_2.

Шаг 4. Вычисляется по следующему правилу дробь:

$$GQ = \frac{ESS_1}{ESS_2} \sim P_F(q)$$

Эта дробь является статистикой критерия проверяемой гипотезы о гомоскедастичности случайного возмущения. Величина GQ имеет распредение Фишера с кол-ом степеней свободы m, n.

Шаг 5. Гипотеза о гомоскедастичности принимается как не противоречащая реальным данным, если оказываеются справедливыми следующие два неравенства:

$$\begin{cases} GQ \overset{?}{\leq} F_{\text{крит}} \\ \dfrac{1}{GQ} \overset{?}{\leq} F_{\text{крит}} \end{cases}$$

Где символом $F_{\text{крит}}$ мы обозначаем квантиль распределения Фишера заданного уровня $1 - \alpha$, например $1 - \alpha \ = 0.95$.

Итог: экономисты тестируют все предпосылки в частности предпосылка №2 тестируется тестом Голдфелда-Кванта.

37. **Тест Дарбина–Уотсона отсутствия автокорреляции у случайного возмущения в ЛММР.**

Проверяемая гипотеза в этом тесте имеет вид:

$$H_0 : Cov(u_{t+1}, u_t) = 0$$

Альтернативная гипотеза заключается в положительном значении ковариации u_{t+1}, u_t:

$$H_1 : Cov(u_{t+1}, u_t) > 0$$

Альтернатива имеет наиболее важное для практики значение. Если справедлива данная альтернатива, то причина этого обстоятельства чаще всего заключается в ошибочной спецификации модели.

Тест $DW(\text{Дарбина} - \text{Уотсона})$ проводится в итоге следующих шагов:

Шаг 1. По уравнениям наблюдений оценивается модель и вычисляется по правилу

$$DW = \frac{\sum_{i=1}^{n-1} \left(\widetilde{u}_{i+1} - u_i\right)^2}{\sum_{i=1}^{n} \left(\widetilde{u}_i\right)^2}$$

статистика критерия гипотезы H_0.

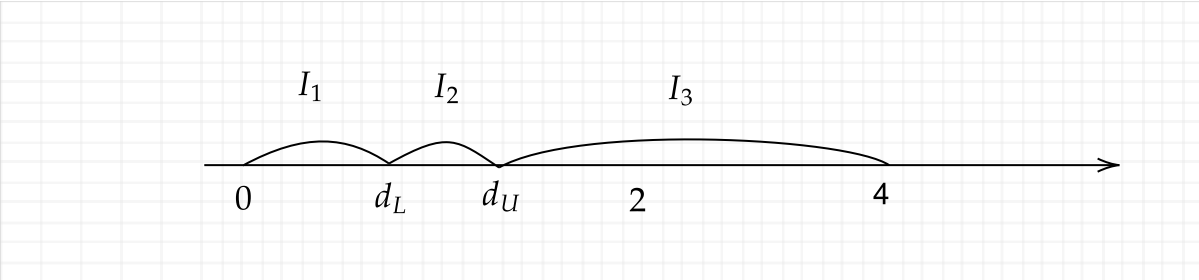
Шаг 2. По таблицам Дарбина-Уотсена

n	$k^1 = 1$		$k^1 = 2$		$k^1 = 3$		$k^1 = 4$		$k^1 = 5$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
6	0,61	1,40	–	–	–	–				
7	0,70	1,36	0,47	1,90	–	–				
8	0,76	1,33	0,56	1,78	0,37	2,29				
9	0,82	1,32	0,63	1,70	0,46	2,13				
10	0,88	1,32	0,70	1,64	0,53	2,02				
11	0,93	1,32	0,66	1,60	0,60	1,93				
12	0,97	1,33	0,81	1,58	0,66	1,86				
13	1,01	1,34	0,86	1,56	0,72	1,82				
14	1.05	1.35	0.91	1.55	0.77	1.78				
16	1,10	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,67	2,10
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,40	1,08	1,53	0,97	1,68	0,86	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,90
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
26	1,30	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,88
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83

Figure 1: Значения статистики Дарбина - Уотсона

Выбираются две величины $d_L,\ d_U$ используя два входа $n,\ k$.

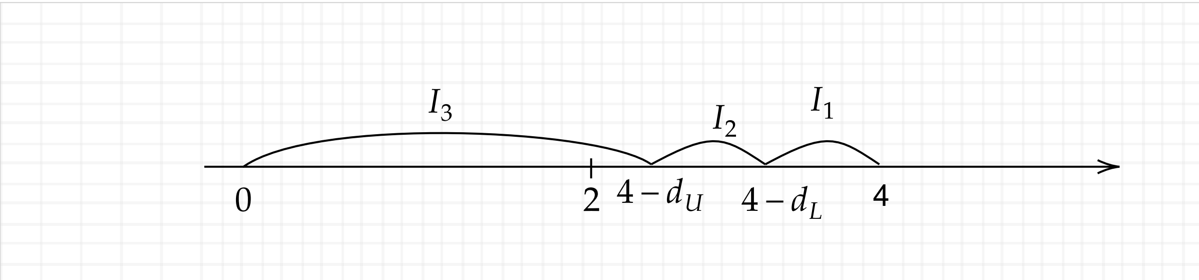
Шаг 3. Определяется один из трёх интервалов в который попадает статистика DW.



Если DW попало в I_3, то H_0 принимается, если в I_1, то гипотеза H_0 отвергается в пользу гипотезы H_1; если в интервал I_2, то ничего сказать нельзя - это *интервал неопределённости*.

Проверка гипотезы при альтернативе $Cov(u_{t+1}, u_t) < 0$

Первые два **шага** остаются без изменений, а чертёж с интервалами выглядит так:



Если статистика DW попадает I_1, то гипотеза H_0 отклонятся в пользу гипотезы H_1 (очень редкий случай), если в I_2 то ничего сказать нельзя - это *интервал неопределённости;* Если DW попадает в I_3, то H_0 принимается.

38. **Коэффициент детерминации как мерило качества спецификации эконометрической модели. Скорректированный коэффициент детерминации и его использование для модификации ЛММР.**

Оценивание эконометрической модели осуществляется на 3-ем этапе схемы её построения. Простешей характеристикой качества служит коэффициент детерминации модели R^2. R^2 – это доля эндогенной переменной модели, которая объясняется предопределёнными переменными модели.

Выведем формулу для величины R^2

Шаг 1. По уравнениям наблюдений рассчитываем оценки случайных возмущений

$$\tilde{u}_i = y_i - \left(\tilde{a}_0 + \tilde{a}_1 \cdot x_i\right) = y_i - \tilde{y}_i$$

Перепишем следующим образом уравнения наблюдений

$$\begin{cases} y_1 = \tilde{y}_1 + \tilde{u}_1 \\ y_2 = \tilde{y}_2 + \tilde{u}_2 \\ \dots \\ y_n = \tilde{y}_n + \tilde{u}_n \end{cases} \tag{4.14}$$

В уравнении (4.14) первое слагаемое в правой части объясняются перменной x, а вторые слагаемые необъясняются x-ами. Справедливо, следующая теорема:

$$\sum (y_i - \overline{y})^2 = \sum \left(\widetilde{y}_i - \overline{\widetilde{y}}\right)^2 + \sum \widetilde{u}_i^2 \tag{4.15}$$

В левой части тождества размещается характеристика изменчивости эндогенной переменной - *волатильность.* Первое слагаемое в правой части обозначим его символом RSS объясняется изменчивостью предопределённых значений и полность объясняется x. А второе слагаемое пораждено неучтёнными факторами $ESS = \sum \widetilde{u}_i^2$, левую часть $TSS = \sum \left(\widetilde{y}_i - \overline{\widetilde{y}}\right)^2$. И разделим обе часть тождества (4.15) на велечину TSS в итоге придём к формуле (4.16).

$$R^2 = 1 - \frac{ESS}{TSS} \tag{4.16}$$

Рассматривая (4.16) мы констатируем, что R^2 – это доля эндогенной переменной модели, которая объясняется предопределёнными перменными R^2.
Скоррективонная коэффицент детерминации рассчитывается по следующему правилу:

$$\overline{R}^2 = 1 - \frac{\frac{ESS}{n-(k+1)}}{\frac{TSS}{n-1}}$$

$$\overline{R}^2 \leqslant R^2$$

Числитель и позволяет отбирать в модель объясняющие перменные. Если при включении в модель новой объясняющей переменной велечина $\overline{R}^2$ возрастает, то включение этой переменной в модель полезно, если убывает, то бесмысленно.

Замечание. В числителе вычитаемого $\dfrac{ESS}{n-(k+1)}$ размещается оценка дисперсии случайного возмущения. В знаменателе находится дисперсия эндогенной переменной. При добавлении в модель новой объясняющей переменной меняется только числитель (знаменатель остаётся всегда неизменным) и поэтому увеличение $\overline{R}^2$ равносильно снижению дисперсии случайного возмущения и значит уменьшению экзогенных переменных. Добавим, что всегда имеет место следующее неравенство, при чём $\overline{R}^2$ может быть меньше 0.

39. **Связь коэффициента детерминации с коэффициентом корреляции эндогенной переменной и её оценки.**

Коэффициент детерминации равен квадрату модуля коэффициента корреляции прогноза $\tilde{y}$ и y. Он показывает, какая доля дисперсии результативного признака объясняется влиянием независимых переменных.

Доказательство:

$$Cor(y,\tilde{y}) = \frac{Cov(y,\tilde{y})}{\sqrt{Var(y)}\sqrt{Var(\tilde{y})}} = \frac{Cov(\tilde{y}+u,\tilde{y})}{\sqrt{Var(y)}\sqrt{Var(\tilde{y})}} = \frac{Cov(\tilde{y},\tilde{y}) + Cov(u,\tilde{y})}{\sqrt{Var(y)}\sqrt{Var(\tilde{y})}} =$$

$$= \frac{Var(\tilde{y}) + Cov(u,\tilde{y})}{\sqrt{Var(y)}\sqrt{Var(\tilde{y})}} = \frac{Var(\tilde{y})}{\sqrt{Var(y)}\sqrt{Var(\tilde{y})}} = \frac{\sqrt{Var(\tilde{y})}}{\sqrt{Var(y)}} = \sqrt{\frac{Var(\tilde{y})}{Var(y)}} = \sqrt{R^2}$$

***При условии, что** (для парной регрессии)*

$\tilde{y} = a_0 + a_1x;\ \ u = y - a_0 - a_1x;\ b = \frac{Cov(x,y)}{Var(x)}$ *и с использованием ковариационных правил, можно доказать, что*

$$Cov(u,\tilde{y}) = Cov(y - a_0 - a_1x, a_0 + a_1x) = Cov(y - a_1x, a_1x) = Cov(y, a_1x) - Cov(a_1x, a_1x) =$$

$$= a_1Cov(y,x) - a_1{}^2Cov(x,x) = a_1\big(Cov(y,x) - a_1Cov(x,x)\big) = a_1(a_1Var(x) - a_1Var(x)) = 0$$

(для множественной регрессии аналогично).

40. **F-тест качества спецификации эконометрической модели.**

F – тест - процедура проверки гипотезы о неудовлетворительной спецификации эконометрической модели:

$$H_0 : a_1 = a_2 = \ \dots\ = a_k = 0$$

То есть гипотеза о том, что ни одна объясняющая переменная не несёт в себе информацию об эндогенной переменной y. Альтернативой для H_0 служит гипотеза:

$$H_1 = \overline{H}_0$$

Означающая, что хотя бы один из коэффициентов отличны от нуля.

Порядок F – теста

Шаг 1. Модель оценивается методом наименьших квадратов и рассчитывается статистика F критерия гипотезы H_0:

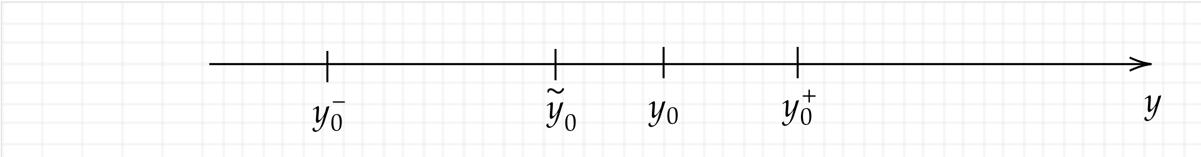
$$F = \frac{R^2 / k}{\left(1 - R^2\right) / (n - (k+1))} \qquad \text{(енто дробь)}$$

Если эта гипотеза верна, то случайная переменная F имеет закон распределения Фишера с кол-ми степеней свободы $k,\ n - k + 1$. Если велина F превосходит квантиль распределения Фишера уровня $1 - \alpha$, где $\alpha = 0.01\ - 0.05$, то гипотеза H_0 отвергается. Эта квантиль обозначена $F_{\text{крит}}$.

Вывод: F – тест позволяет объективно объяснить качество перменных модели.

41. **Процедура интервального прогнозирования значений эндогенной переменной по оценённой линейной эконометрической модели с гомоскедастичным неавтокоррелированным случайным возмущением.**

оптимальный прогноз $\widetilde{y}_0$.



Помимо точечного прогноза в финансово-экономической сфере прогноз искомой величины y_0 часто строится в виде интервала с левой границей y_0^- и правой границей y_0^+. Этот интервал накрывает неизвестное значение y_0 с заданной доверительной вероятностью. В основании лежит следующая теорема:

Пусть в моделе выполнены все предпосылки теоремы Гаусса-Маркова и случайные возмущения имеют нормальный закон распределения

Тогда следующая дробь:

$$t = \frac{\widetilde{y}_0 - y_0}{\widetilde{\sigma}\left(\triangle \widetilde{y}_0\right)} \sim t(m), \text{ где } m = m - k + 1$$

имеет закон распределения Стьюдента с количеством степеней свододы $n - k + 1$.

Из теоремы выше вытекает равенство:

$$P\left(\left|\frac{\widetilde{y}_0 - y_0}{S\widetilde{y}_0}\right| \leqslant t_{\text{крит}}\right) = 1 - \alpha$$

$t_{\text{крит}}$ имеет значение двухсторонней квантили с кол-ом степеней свободы $n - k + 1$.

Освобождаясь от модуля мы перепишем формулу в следующем виде:

$$P\left(\widetilde{y}_0 - t_{\text{крит}} \cdot \widetilde{\sigma}\left(\triangle \widetilde{y}_0\right) \leqslant y_0 \leqslant \widetilde{y}_0 + t_{\text{крит}} \cdot \widetilde{\sigma}\left(\triangle \widetilde{y}_0\right)\right) = 1 - \alpha$$

$\left[y_0^-,\ y_0^+\right]$ – доверительный интервал

42. **Процедура проверки адекватности оценённой линейной эконометрической модели.**

1. Результаты наблюдений объекта следует разделить на два класса. В первый класс (обучающую выборку) включить основной объем результатов наблюдений 95% выборки. Оставшиеся результаты наблюдений (например, пара (x_0, y_0)) составят контролирующую выборку.

2. По обучающей выборке $(\vec{y}, X)$ оценить модель:

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

3. Задаться доверительной вероятностью $(1-\alpha)$ и по значениям регрессоров, входящих в контролирующую выборку, построить доверительные интервалы $[Y_0^-, Y_0^+]$ для соответствующих этим регрессорам значений эндогенной переменной модели.

$$y_0^- = \tilde{y}_0 - t_{\text{крит}} S_{\tilde{y}_0}, \quad y_0^+ = \tilde{y}_0 + t_{\text{крит}} S_{\tilde{y}_0}$$

$$\tilde{y}_0 = \tilde{a}_0 + \tilde{a}_1 x_0$$

$$S_{\tilde{y}_0} = \tilde{\sigma}_u \sqrt{1 + q_0}$$

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

$t_{\text{крит}}$ – двусторонний $(1-\alpha)$–квантиль распределения Стьюдента с количеством степеней свободы $v_2 = n - (k+1)$, где $(k+1) = 2$ – количество оцениваемых коэффициентов модели **(1)**.

Проверить, попадают ли значения эндогенной переменной из контролирующей выборки в соответствующие доверительные интервалы (в интервал $[y_0^-, y_0^+]$). Если да, то признать оцененную модель <u>адекватной</u>; если нет, то оцененная модель <u>не может быть признана адекватной</u> и подлежит доработке.

43. Последствия, симптомы и методика устранения ошибки спецификации эконометрической модели, состоящей в неверном выборе функции регрессии.
Нередко на этапе проверки адекватности модели выясняется, что оцененная модель оказывается неадекватной. Одной из причин этого может быть допущенная ошибка спецификации, например, пропуск значащей объясняющей переменной. Данная ошибка является частным случаем ошибки неверного выбора типа функции регрессии со всеми последствиями и симптомами.
Пусть экономист выбрал линейную функцию регрессии, в то время как адекватной является
нелинейная.
Последствием ошибочного выбора типа функции в уравнении регрессии оказывается нарушение основной предпосылки эконометрической модели: $E(u|x) \neq 0$. Это приводит к нарушению 1-ой предпосылки теоремы Гаусса-Маркова и смещённости оценок коэффициентов в модели, вычисленных по правилу *A* теоремы Гаусса-Маркова. Конечным следствием смещённости оценок оказывается необъективность точечных и интервальных прогнозов значений эндогенных переменных.
Симптомы данной ошибки:
Первый симптом состоит в несоответствие диаграммы рассеивания, построенной по выборке $(X,\ \vec{y})$ (обучающая выборка) графику функции

$$y = f_F(x;\vec{a}) \qquad (15.2)$$

Например, ошибочно выбранная функция (15.2) являяется линейной по x, в то время как диаграмма рассеивания свидетельствует, что функция регрессии $E(y|x) = f_T(x)$ суть нелинейнная функция аргумента x.
Второй симптом - это длительное постоянство знака оценок случайных остатков

$$\widetilde{u}_1,\ \dots,\widetilde{u}_i,\widetilde{u}_{i+1},\ \dots,\widetilde{u}_n$$

в упорядоченных (по возрастанию значений объясняющей переменной) уравнений наблюдений. Именно этот симптом улавливается статистикой DW теста Дарбина-Уотсона.
Для обнаружения *третьего симптома* необходимо разделить выборку $(X,\ \vec{y})$ на две примерно равные по количеству наблюдений части $(X_1,\ \vec{y}_1)$ и $(X_2,\ \vec{y}_2)$. Так чтобы различие в элементах x_i матриц X_1 и X_2 было по возможности существенным. Затем по каждой из выборок надлежит оценить

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

Сильное отличие одноимённых коэффициентов в двух оценённых вариантах модели и есть третий симптом неверного выбора функции регрессии.
Методика устранения
Допустим ошибка в выборе функции регрессии подтвердилась, в такой ситуации следует, используя диаграмму рассеивания подобрать другую подходящую фукнцию регрессии и повторить процедуру построения регрессионной модели.

44. Последствия и симптомы ошибки спецификации линейной эконометрической модели, состоящей во включении незначимой объясняющей переменной.

Вспомним определение незначащей объясняющей переменной, так например x_1, является незначащим, если справедлива гипотеза:

$$H_0 : a_1 = 0 \tag{1}$$

Если же она не справедлива, то есть справедлива альтернативная гипотеза:

$$H_1 : a_1 \neq 0$$

то переменная x_1 является значащей и её необходимо сохранить в моделе.

Последствия: Наличие в модели незначащих переменных увеличивает дисперсию оценок коэффициентов модели (т.е. ухудшает точность оценок).

Симптомы: Основной симптом наличия в модели незначащей переменной: стандартная ошибка оценки коэффициента при этой переменной $\left(S\tilde{a}_i\right)$ находится на уровне или превышает абсолютное значение оценки коэффициента при этой переменной $\left(\tilde{a}_i\right)$.

Для определения незаначимости переменной используется $t-$тест, который позируется на следующей **теореме:**

Пусть выполнены все предпосылки теоремы Гаусса-Маркова, а случайные возмущения имеют нормальный закон распределения. Пусть модель оценена методом наименьших квадратов, то гогда следующая дробь:

$$t = \frac{\tilde{a}_1}{S\tilde{a}_1}$$

является случайной переменной расапределённой по закону Стьюдента с количеством степеней свободы $m = (n - (k+1))$.

Порядок $t-$теста о незначимости объясняющей переменной в оценённой модели

Шаг 1. Визуальный поиск в оценённой моделе таких объясняющих переменных в которых спраделиво следующее неравенство:

$$|\tilde{a}_j| \leqslant S\tilde{a}_j |t| \overset{2}{\leqslant} t_{\text{крит}} \tag{2}$$

Если находится такая объясняющая переменная x_j для которого справедливо данное неравенство, то это означает, что значение оценки коэффициента $\tilde{a}_j$ скорее всего вызвана ошибкой оценивания коэффициента $a_j = 0$.

Именно с таких объясняющих переменных нужно приступать к $t-$ тесту.

Шаг 2. Расчёт статистики t:

$$t = \frac{\tilde{a}_1}{S\tilde{a}_1}$$

Что гипотеза $H_0\ a_j = 0$.

Шаг 3. Задаться значением $\alpha \in [0, 0.05]$ и при количестве степеней свободы $m = (n-(k+1))$ найти при помощи функции "СТЬЮДЕНТ.ОБР.2Х" найти двустороннюю квантиль уровня $1-\alpha$ распределения Стьюдента. Пример, выбираем уровень значимости 0.05 с кол-ом степеней свободы $m = 11$, тогда упомянутая выше квантиль равна ≈ 2.2. Часто такую квантиль обозначают $t_{\text{крит}}$.

Шаг 4. Проверить справедливость следующего неравенства:

$$t| \overset{?}{\leqslant} t_{\text{крит}}$$

Если оно справедли, то гипоза $H_0 : a_j = 0$ может быть принята, как не противоречащая реальным данным и переменная x_j удалена из модели, в противном случае приниматся гипотеза H_1 переменная x_j интерпритируется, как значащая и сохраняется в моделе.

Замечание. Из курса математической статистики известно, что процедура проверк статистической гипотезы, может приводить к ошибкам I или II рода. Ошибка I рода - отвергнуть гипотезу H_0, когда она верна. Ошибка II рода - когда мы приняли гипотезу, когда она не верна. В ситуации $t-$ теста гораздо опаснее принять гипотезу H_0, когда она не верна и следовательно исключить значащую переменную из модели (лучше сохранить незначащую чем сохранить значащую). После удаления из модели незначащих переменных необходимо повторить все тесты в моделе.

45. Последствия и симптомы ошибки спецификации линейной эконометрической модели, состоящей в пропуске значимой объясняющей переменной.

Эта ошибка по последствиям и симптомам эквивалентна неверному выбору типа функции регрессии.

Пусть на первом этапе экономист составил спецификацию модели:

$$\begin{cases} y \ = a_0 + a_1 x_1 + u; \\ E(u|x_1) = 0, E\left(u^2|x_1\right) \ = \sigma_u^2; \end{cases}$$

в ситуации, когда истинной является спецификация:

$$\begin{cases} y \ = a_0 + a_1 x_1 + a_2 x_2 + u; \\ E(u|x_1, x_2) = 0, E\left(u^2|x_1, x_2\right) \ = \sigma_u^2; \end{cases}$$

при условии, что гипотеза $H_0 : a_0 = 0$ неверна. Это означает, что экономист сделал ошибочный выбор типа функции регрессии, а именно: вместо функции двух аргументов выбрал функцию одного аргумента. Влияние данной ошибки на случайный остаток в модели (1): $E(u|\vec{x}_1) = (a_0 + a_1 x_1 + a_2 x_2) - (a_0 + a_1 x_1) = a_2 x_2 = \phi(x_2) \neq 0$.

Действительно пропуск значащей объясняющей переменной эквивалентен неверному выбору типа функции регрессии. Если пропущенная переменная (скажем, регрессор x_2) недоступна для наблюдений, то экономист может включить в модель ее заместителя - такую переменную x_2^{pr}, которая доступна для наблюдений и коррелирует с переменной x_2.

Последствия и симптомы смотри в пункте 44.

46. Последствия и симптомы ошибки спецификации линейной эконометрической модели, состоящей в игнорировании гетероскедастичности случайного возмущения.

Гетероскедастичность случайных возмущений – возмущения обладают различными дисперсиями, но не коррелированны друг с другом.

Симпотомы: при гетероскедастичности распределение u для каждого наблюдения имеет нормальное распределение и нулевое ожидание, но дисперсия распределений различна.

Последствия гетерескедастичности случайных возмущений**:**

1. Потеря эффективности оценок коэффициентов регрессии, т.е. можно найти другие, отличные от Метода Наименьших Квадратов и более эффективные оценки;
2. Смещенность стандартных ошибок коэффициентов в связи с некорректностью процедур их оценки;

Для определения гетероскедастичности случайных возмущений используют тест Голдфилда-Кванта.

Тест Голдфелда-Кванта гипотезы о гомоскедастичности случайного возмущения

Шаги теста предпосылки №2 Голдфелда-Кванта

Шаг 1. Составляется система уравнений наблюдений объекта

$$\begin{cases} y_1 = a_0 + a_1 \cdot x_{1,1} + a_2 \cdot x_{2,1} + \ldots a_k \cdot x_{k,1} + u_1 \\ y_2 = a_0 + a_1 \cdot x_{1,2} + a_2 \cdot x_{2,2} + \ldots a_k \cdot x_{k,2} + u_2 \\ \ldots \\ y_n = a_0 + a_1 \cdot x_{1,n} + a_2 \cdot x_{2,n} + \ldots a_k \cdot x_{k,n} + u_n \end{cases}$$

Замечание. Если справедлива предпосылка №2 теоремы Гаусса-Маркова, то при любых перестановках уравнений наблюдений дисперсии случайных возмущений остаются неизменными. Если же предпосылка №2 нарушается, то, как правило, дисперсии случайных возмущений в уравнения 1 и 2 возрастают (или убывают) в ответ (по мере) на возрастание абсолютных значений объясняющих переменных.

Шаг 2. Уравения наблюдений упорядычеваются по возрастанию сумм абсолютных значений объясняющих перменных

$$\sum_{j=1}^{k} |x_{ji}|$$

Шаг 3. По первым n_1 упорядоченным уравениям оцениваются методом наименьших квадратов параметры модели и запоминается значения ESS_1. Количество n_1 выбирается согласно следующим двум условиям:

$$a)\, n_1 \approx \frac{1}{3} n, \;\; b)\, n_1 > k + 1$$

Аналогично оценивается модель по последним n_1 уравениям и запоминается значение ESS_2.

Шаг 4. Вычисляется по следующему правилу дробь:

$$GQ = \frac{ESS_1}{ESS_2} \sim P_F(q) \qquad (6.1.10)$$

Эта дробь является статистикой критерия проверяемой гипотезы о гомоскедастичности случайного возмущения. Величина GQ имеет распредение Фишера с кол-ом степеней свободы m, n.

Шаг 5. Гипотеза о гомоскедастичности принимается как не противоречащая реальным данным, если оказываются справедливыми следующие два неравенства:

$$\begin{cases} GQ \overset{?}{\le} F_{\text{крит}} \\ \dfrac{1}{GQ} \overset{?}{\le} F_{\text{крит}} \end{cases}$$

Где символом $F_{\text{крит}}$ мы обозначаем квантиль распределения Фишера заданного уровня $1-\alpha$, например $1-\alpha\ = 0.95$.

Итог: экономисты тестируют все предпосылки в частности предпосылка № 2 тестируется тестом Голдфелда-Кванта.

47. Оценивание линейной модели с автокоррелированным остатком AR(1) алгоритмом Хилдрета–Лу.

Спецификация AR(1) выглядит следующим образом:

$$\begin{cases} y_t = a_0\ +\ \vec{a}^{T} \cdot \vec{x}_t + u_t \\ u_t = \rho \cdot u_{t-1} + \xi_t;\ |\rho| < 1 \\ E(u_t) = 0; E\left(u_t^2\right)\ = \sigma_u^2\ = \dfrac{\sigma_\xi^2}{1-\rho^2}; \end{cases}$$

Если бы нам был известен параметр авторегрессии ρ, то никаких проблем с оцениванием этой модели у нас бы не возникло. Однако в практике реальных исследований этот параметр

практически никогда нам не бывает известен. Следовательно, встает задача: оценить ρ.

Один из методов оценки является **алгоритм Хилдрета–Лу**:

Все значения параметра авторегрессии лежат в пределах $\rho \in (-1;1)$. Следовательно, идея состоит в том, чтобы из этого интервала с небольшим шагом выбирать различные значения и оценивать для каждого из них регрессию:

$$y_t = a_0\ +\ \vec{a}^{T} \cdot \vec{x}_t + u_t$$

При этом следим за суммой квадратов остатков RSS. В качестве оценки ρ берем то его значение, для которого RSS минимальна.

48. Проблема совершенной мультиколлинеарности и её выявление методом дополнительной регрессии.

Критерий мультиколениарности:

$$rk(X) < (k+1) \Leftrightarrow |Cor(\vec{x}, \vec{x})| = 0$$

На практике часто встречаются ситуации, когда упомянутый выше определитель не равен 0, но очень мал. Тогда говорят, что в исходной модели присутствует несовершенная мультиколениарность. Отметим симптомы несовершенной мультиколениарности.

Симптомы:

Симптом №1. F – тест отвергает гипотезу:

$$H_0 : a_1 = a_2 = \ \dots \ = a_k = 0,$$

но t – тест **не отвергает** гипотезу $H_0 : a_j = 0$ о незначимости многих x_j.

Симптом № 2. При добавлении в обучающую выборку одного уравнения (или удаления одного из уравнений), оценки коэффициентов модели резко меняют свои значения, то есть оценки неустойчивые $\tilde{a}_j$ и значит они ненадёжные.

При несовершенной мультиколлениарности часто используется процедура пошаговой регрессии отбора в модель объясняющих переменных. Процедура состоит из следующих шагов:

Шаг № 1. Модель оценивается по всем объясняющим переменным и отмечается скорректированный коэффициент детерминации (обозначим символом $\overline{R}_0^2$).

Шаг № 2. Из исходной оценённой модели выбирается коэффициент $\tilde{a}_j$ с максимальным значением $|t_j| = \max x_j$. Удаляется из модели, переоценивается и запоминается значение скорректированной детерминации. Если $\overline{R}_0^2 \geqslant \overline{R}_1^2$, то удаление x_j нецелесобразно.

Шаг №3. Шаг №1 повторяется пока условие $\overline{R}_j^2 \geqslant \overline{R}_{j+1}^2$ перестанет повторяться.

49. Вложенные модели. Тест Вальда вложенной модели.

В процессе проверки адекватности модели нередко возникает задача по удалению из модели незначащих переменных. Удаление из модели таких переменных (незначащих) модет быть осуществленно с помощь либо t-теста (лекция от 9 декабря), либо при помощи теста Вальда. И в том и в другом случае в основании этих тестов лежит понятие вложенной модели. Вот определение этого понятия: модель М1 называется вложенной в модель М2:

$$M1 \quad y = a_0 + a_1 x_1 + u \tag{5}$$

$$M2 \quad y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + v \tag{6}$$

если для модели М2 справедлива гипотеза:

$$H_0 : a_2 = 0; a_3 = 0; \tag{7}$$

Модель М1 иногда также называется *моделью М2 в которой справедливо ограничение (7).*

Модель М2 следует предпочесть, если справедлива гипотеза 8:

$$H_1 : a_2 \neq 0 \cup a_3 \neq 0.\text{(хотя бы одна является значащей)}$$

Гипотеза H_1 является отрицанием гипотезы H_0. Тест Вальда исследует гипотезу (7) против альтернативы (8). Он состоит из следующих шагов:

Шаг 1. МНК оценивается модель (5) и отмечается значение суммы квадратов оценок случайных возмущений.

Шаг 2. Оценивается модель (6) МНК и отмечается ESS_2.
И шаг 1 и шаг 2 осуществляется по одной и той же выборке.
Шаг 3. По правилу (9) вычисляется статистика критерия гипотезы H_0:

$$W = (ESS_1 - ESS_2) / (ESS_2 / (n - k_2)) \quad (9)$$
$$W \sim \chi_r^2 \text{ при } n \to \infty; r - \text{число}$$

Пусть справеливы все предпосылки теоремы Гаусса-Маркова при произовальном законе распределения случайных возмущений. Если справедлива гипотеза H_0, то статистика асимптотически (с ростом объёма выборки) имеет распределение хи-квадрат с числом степеней свободы r, где r число ограничений (7) в модели (6).
Шаг 4. Гипотеза H_0 отвергается в пользу гипотезы H_1 (т.е принимается модель (6)), если значение статистики превосходит квантиль распределения χ_r^2 уровня $1 - \alpha$:

$$W > \chi_r^2(1 - \alpha)$$

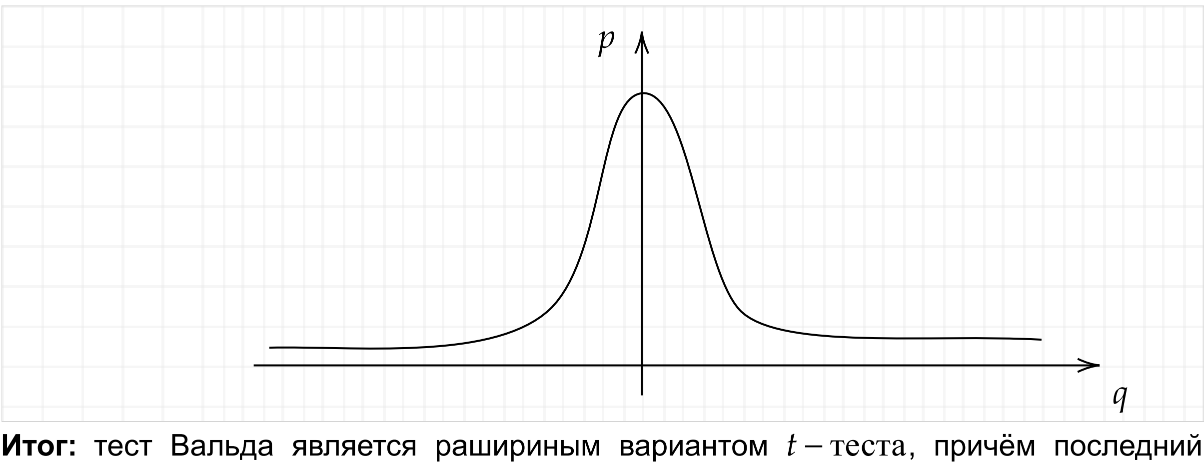
Следствие. Если случайное возмущение в модели М2 имеет нормальный закон распределения, то:

$$\frac{W}{r} \sim F_{r,n-k_2} \quad (10)$$

В такой ситуации гипотеза H_0 отвергается, если справедливо следующее неравенство:

$$\frac{W}{r} > F_{r,n-k_2}(1 - \alpha)$$

Квантиль:



Итог: тест Вальда является раширеным вариантом $t-\text{теста}$, причём последний $t-\text{тест}$ совпадает с тестом Вальда при $r = 1$.

50. Простейшая модель автокорреляции случайного возмущения AR(1).
Модель AR(1)- имеет следующую спецификацию:

$$\begin{cases} u_t = pu_{t-1} + \xi_t \\ u_1 = 0, E(u_1) = 0, Var(u_1) = \sigma_u^2 \\ \xi_t \in WN, Cov(u_{t-1}, \xi_t) = 0 \\ |\rho| < 1, \left(1 - \rho^2\right) \sigma_u^2 = \sigma_\xi^2 \end{cases}$$

Уравнение модели записывается в виде: $u_t - pu_{t-1} = \xi_t$
В случае AR(1)-процесса:

Математическое ожидание равно: $M = \dfrac{c}{1-a}$

где a – параметр модели (коэффициент авторегрессии), c – постоянная (часто для упрощения предполагается равной нулю)

Дисперсия равна: $D = \dfrac{\sigma_\xi^2}{\left(1-a^2\right)}$

Автокорреляция: $r(k) = r \cdot r^{k-1} \Rightarrow r(k) = r^k$

51. Простейшая модель гетероскедастичности случайного возмущения в ЛММР. Запись оценённой эконометрической модели с гетероскедастичным случайным возмущением.

Простейший вид такой модели: $Var(u) = \sigma_u^2 = \sigma_0^2(\sum_{j=0}^{n} |X_j|^\lambda$

В этой модели присутствуют две константы – положительная константа σ_0^2 и показатель степени λ. Параметр λ подбирается в итоге проведения теста Голдфелда-Квандта из множества значений ±0,5, ±1, ±2 так, чтобы тест просигнализировал о гомоскедастичности остатка в преобразованной ЛММР. Заметим, что если остаток $\lambda = 0$, то остаток в модели гомоскедастичен и константа σ_0^2 будет иметь смысл дисперсии случайного остатка.

54. RESET-тест первой предпосылки теоремы Гаусса-Маркова: $H_0 : E(u|X) = 0$
1. Методом наименьших квадратов оценивается базовая модель эконометрики и вычисляются прогнозные значения:
$\tilde{y}_1 = \tilde{a}_0 + \tilde{a}_1 \cdot \tilde{x}_{1,i} + ... + \tilde{a}_k \cdot \tilde{x}_{k,i}$
2. Задается целочисленное значение (p = 2,3) и оценивается вспомогательная модель (2):
$y_i = a_1 \cdot x_{1t} + a_2 \cdot x_{2t} + ... + a_k \cdot x_{kt} + b_2 \cdot y_t + ... + b_p \cdot y_i + \varepsilon i$
3. Вычисляется статистика
$RESET = RAM\left(\tilde{u}_1, \tilde{u}_2, ..., \tilde{u}_n\right) = (ESS - ESS_2) : p - 1 / ESS_2 : (n - (k + p - 1))$
4. Вычисляется критическое значение
$RESET_{\text{крит}} = F_{p-1,\,(n-(k+p-1))}(1-a)$
Если статистика больше критической точки, то гипотеза H_0 отвергается.

55. Тест Харке-Бера гипотезы о нормальном законе распределения случайного возмущения в эконометрической модели.

1. Модель оценивается методом наименьших квадратов и вычисляются оценки случайных возмущений.

2. Вычисляется статистика критерия

$$JB = JB\left(\widetilde{u}_1, \widetilde{u}_2, \dots, \widetilde{u}_n\right) = n \cdot \left\{\frac{sk^2}{6} + \frac{(kur-3)^2}{24}\right\};$$

$$sk = \frac{m_3}{m_2^{3/2}},\ kur = \frac{m_4}{(m_2)^2}, m_k = \frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{u}_1\right)^k$$

(sk – оценка асимметрии, kur – эксцесса, mk – начального момента k – ого порядка случайного остатка)

3. Вычисляется критическое значение

$$JB_{\text{крит}} = \chi_2^2(1-\alpha)$$

Если статистика больше критической точки, то гипотеза H_0 отвергается.