

Экзаменационные вопросы по курсу «Эконометрика» для студентов ПМЗ-1

1. Назначение экономико-математических моделей (ЭММ). Два принципа их спецификации. Типы уравнений в ЭММ: поведенческие уравнения и тождества (на примере макромодели).
2. Типы переменных в экономических моделях. Структурная и приведённая форма модели (на примере макромодели). Компактная запись.
3. Спецификация и преобразование к приведённой форме динамических моделей. Лаговые и предопределённые переменные динамической модели. Модель Линтнера корректировки размера дивидендов. Компактная запись.
4. Спецификация и преобразование к приведённой форме эконометрических моделей. Эконометрическая модель Самуэльсона–Хикса делового цикла экономики. Компактная запись.
5. Схема построения эконометрических моделей.
6. Порядок оценивания линейной эконометрической модели из изолированного уравнения в Excel. Смысл выходной статистической информации функции ЛИНЕЙН.
7. Случайная переменная и закон её распределения. Нормальный закон распределения и его параметры.
8. Случайная переменная и закон её распределения. Распределение хиквадрат.
9. Случайная переменная и закон её распределения. Распределение Стьюдента Квантиль $t_{\text{крит}}$ уровня $1 - \alpha$ и её расчёт в Excel.
10. Ковариация $\text{Cov}(x, y)$, и коэффициент корреляции, $\text{Cor}(x, y)$ пары случайных переменных (x, y) .
11. Случайная переменная и закон её распределения. Закон распределения Фишера. Квантиль $F_{\text{крит}}$ уровня $1 - \alpha$ и её расчёт в Excel.
12. Случайный вектор и его основные количественные характеристики. Случайный вектор \vec{y} левых частей схемы Гаусса – Маркова при гомоскедастичном неавтокоррелированном случайном возмущении.
13. Основные количественные характеристики аффинного преобразования случайного вектора (на примере вектора $\vec{\theta}_{\text{МНК}}$ – оценок коэффициентов линейной модели при гомоскедастичном неавтокоррелированном случайном возмущении).
14. Случайный вектор, веса компонент случайного вектора и факторизация его ковариационной матрицы. Случайный вектор \vec{y} в схеме Гаусса – Маркова при гетероскедастичном неавтокоррелированном случайном возмущении.
15. Временной ряд и его структура (На примере ВВП России).
16. Модели тренда временного ряда.
17. Моделирование сезонной составляющей при помощи фиктивных переменных.
18. Регрессионная зависимость случайных переменных. Функция регрессии, стандартные модели функции регрессии.
19. Схема Гаусса–Маркова.
20. Понятие статистической процедуры оценивания параметров эконометрической модели. Линейные статистические процедуры. Требования к наилучшей статистической процедуре.
21. Теорема Гаусса–Маркова: выражение вектора оценок коэффициентов $\vec{\theta}$ и доказательство их несмешённости.

22. Теорема Гаусса-Маркова: выражение $\text{Cov}(\hat{a}, \hat{a})$ и его обоснование.
23. Теорема Гаусса-Маркова: предпосылки и свойство наименьших квадратов $\hat{u}^T \cdot \hat{u} \rightarrow \min$.
24. Теорема Гаусса-Маркова: выражение $\tilde{\sigma}_0^2$.
25. Взвешенный метод наименьших квадратов (ВМНК). Практическая реализация ВМНК.
26. Обобщённый метод наименьших квадратов (ОМНК).
27. Система нормальных уравнений и явный вид её решения при оценивании методом наименьших квадратов (МНК) линейной модели парной регрессии.
28. Ковариационная матрица оценок коэффициентов линейной модели парной регрессии: явные выражения $\text{Var}(\tilde{a}_0)$, $\text{Var}(\tilde{a}_1)$ и $\text{Cov}(\tilde{a}_0, \tilde{a}_1)$.
29. Свойства МНК-оценок параметров линейной модели множественной регрессии (ЛММР) при нормальном векторе случайных остатков: закон распределения случайного вектора \hat{u} .
30. Свойства МНК-оценок параметров линейной модели множественной регрессии (ЛММР) при нормальном векторе случайных остатков: закон распределение оценки $\tilde{\sigma}_0^2$.
31. Свойства МНК-оценок параметров линейной модели множественной регрессии (ЛММР) при нормальном векторе случайных остатков: закон распределения дроби $\frac{\tilde{a}_j - a_j}{S\tilde{a}_j}$.
32. Оценивание параметров линейной модели множественной регрессии (ЛММР) при автокоррелированном случайном остатке алгоритмом Хилдрета-Лу.
33. Порядок проверки статистических гипотез (на примере гипотезы об адекватности ЛММР).
34. Спецификация и оценивание нелинейных по коэффициентам моделей множественной регрессии со специальными функциями регрессии (на примере производственной модели с функцией Кобба-Дугласа).
35. Оптимальное точечное прогнозирование значений эндогенной переменной по линейной модели (случай гомоскедастичного и неавтокоррелированного случайного возмущения).
36. Тест Голдфелда-Квандта гомоскедастичности случайного возмущения в ЛММР.
37. Тест Дарбина-Уотсона отсутствия автокорреляции у случайного возмущения в ЛММР.
38. Коэффициент детерминации как мерило качества спецификации эконометрической модели. Скорректированный коэффициент детерминации и его использование для модификации ЛММР.
39. Связь коэффициента детерминации с коэффициентом корреляции эндогенной переменной и её оценки.
40. F-тест качества спецификации эконометрической модели.
41. Процедура интервального прогнозирования значений эндогенной переменной по оценённой линейной эконометрической модели с гомоскедастичным неавтокоррелированным случайным возмущением.

42. Процедура проверки адекватности оценённой линейной эконометрической модели.
43. Последствия, симптомы и методика устранения ошибки спецификации эконометрической модели, состоящей в неверном выборе функции регрессии.
44. Последствия и симптомы ошибки спецификации линейной эконометрической модели, состоящей во включении незначимой объясняющей переменной.
45. Последствия и симптомы ошибки спецификации линейной эконометрической модели, состоящей в пропуске значимой объясняющей переменной.
46. Последствия и симптомы ошибки спецификации линейной эконометрической модели, состоящей в игнорировании гетероскедастичности случайного возмущения.
47. Оценивание линейной модели с автокоррелированным остатком AR(1) алгоритмом Хилдрета – Лу.
48. Проблема совершенной мультиколлинеарности и её выявление методом дополнительной регрессии.
49. Вложенные модели. Тест Вальда вложенной модели.
50. Простейшая модель автокорреляции случайного возмущения AR(1) в ЛММР.
51. Простейшая модель гетероскедастичности случайного возмущения в ЛММР. Запись оценённой эконометрической модели с гетероскедастичным случайным возмущением.
52. Оптимальное точечное прогнозирование значений эндогенной переменной по оценённой эконометрической модели с гетероскедастичным неавтокоррелированным случайным возмущением. Характеристика точности оптимального точечного прогноза по ЛММР с гетероскедастичным случайным возмущением.
53. Процедура интервального прогнозирования значений эндогенной переменной по оценённой линейной эконометрической модели с гетероскедастичным неавтокоррелированным случайным возмущением.
54. RESET - тест предпосылки теоремы Гаусса - Маркова $H_0: E(u|X) = 0$.
55. Тест Jarque-Bera гипотезы $H_0: u \sim N(0, \sigma^2)$ о нормальном законе распределения случайного возмущения в эконометрической модели.

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 случаю методом подстановки к **приведенной форме**, где каждая эндогенная переменная представляется в виде явной функции только экзогенных переменных:

$$\begin{aligned} C &= \frac{a_0}{1-a_1} + \frac{a_1}{1-a_1} \cdot I \\ Y &= \frac{a_0}{1-a_1} + \frac{1}{1-a_1} \cdot I \end{aligned}$$

Компактная запись

Обозначив векторы эндогенных переменных $\bar{y} = \begin{pmatrix} Y \\ C \end{pmatrix}$ и экзогенных переменных

$\bar{x} = \begin{pmatrix} 1 \\ I \end{pmatrix}$, мы можем записать макромодель Кейнса в компактном виде:

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

Составив матрицы A и B , получим компактную запись:

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

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 $\bar{y} = \begin{pmatrix} DPS_t^w \\ DPS_t \end{pmatrix}$ и экзогенных переменных

$\bar{x} = \begin{pmatrix} DPS_{t-1} \\ EPS_t \end{pmatrix}$, мы можем записать модель Линтнера в компактном виде:

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

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:

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

а приведенная форма:

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

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:

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

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

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

Компактная запись

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

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

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

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 сельскохозяйственных государственных расходов имеет следующую спецификацию:

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

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 значениях переменных, входящих в модель. (Примером такой информации служит файл "Элементы использования ВВП").

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

Примем обучающейся информации С (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 $(\tilde{g}, \tilde{\sigma}_w)$ (3). Оценки (3) вычислим методом наименьших квадратов.

Шаг 4. Оценённая модель проходит проверку адекватности:

$$G_t = \tilde{g} (S_{\tilde{g}}) G_{t-1} + w_t (\tilde{\sigma}_w) \quad (4)$$

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

$$\delta = |\tilde{G}_t - G_t| / (G_t) \cdot 100 \leq 15\% \quad (5)$$

Модель признаётся адекватной, если относительная 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:

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

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

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

| | | | | |
|-------|-------|---|---|---|
| F | v_2 | # | # | # |
| RSS | ESS | # | # | # |

$\tilde{a}_0, \dots, \tilde{a}_k$ – оценки коэффициентов;

$S_{\tilde{a}_0}, \dots, S_{\tilde{a}_k}$ – стандартные ошибки коэффициентов;

R^2 – коэффициент детерминации;

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

F – статистика Фишера;

ESS – сумма квадратов оценок случайных возмущений;

$$RSS = \sum (\tilde{y}_i - \bar{\tilde{y}})^2 ; \text{(но это не точно)}$$

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 *дифференциальным 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 $Y = X_1^2 + \dots + X_n^2$ 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 рассчитывается 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 хиквадрат: $Y_i \sim \chi^2(d_i)$, где $d_i \in \mathbb{N}, i = 1, 2$. Тогда распределение случайной величины $F = \frac{Y_1/d_1}{Y_2/d_2}$, называется распределением Фишера со степенями свободы d_1 и d_2 .

Пишут $F \sim F(d_1, d_2)$.

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

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

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

$$\begin{cases} GQ \stackrel{?}{\leq} F_{\text{крит}} \\ \frac{1}{GQ} \stackrel{?}{\leq} F_{\text{крит}} \end{cases}$$

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

Расчет в Excel: =F.OBR(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 называть **случайным вектором**:

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

Для практики важны следующие две случайные характеристики:

1. Математической ожидание случайного вектора

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

– это вектор из математических ожиданий случайных компонент.

Математическое ожидание – это среднее значение. Математическое ожидание – это константа.

2. Ковариационная матрица:

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

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 характеризующая взаимосвязь компоненты x_1 и x_n . Если x_1 и x_n независимые, то $\sigma_{1n} = 0$.

Компактная запись схемы Гаусса–Маркова:

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

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 случайному возмущении).

Аффинное 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:

$$\vec{a} = (X^T \cdot X)^{-1} \cdot X^T \cdot \vec{y}; \quad (*)$$

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:

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

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 аффинное преобразование вектора \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:

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

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

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

Ковариационная матрица вектора \vec{a} по определению равна:

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

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

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

И подставим в формулу ковариации выше получим:

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

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**:

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

Веса компонент случайного вектора и факторизация его ковариационной матрицы

Пусть $\vec{x}^T = (x_1, x_2 \dots, x_n)$ случайный вектор, пусть x_i какая-то компонента вектора; вес компоненты 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

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

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:

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

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

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

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

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

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

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

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 А), В), С), Д) теоремы Гаусса-Маркова применительно к уравнениям наблюдения 6.4.5 превращаются в следующие утверждения:

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

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

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

$$\tilde{v}_i = \sqrt{p_i} \cdot \tilde{u}_i \quad (6.4.10)$$

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

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

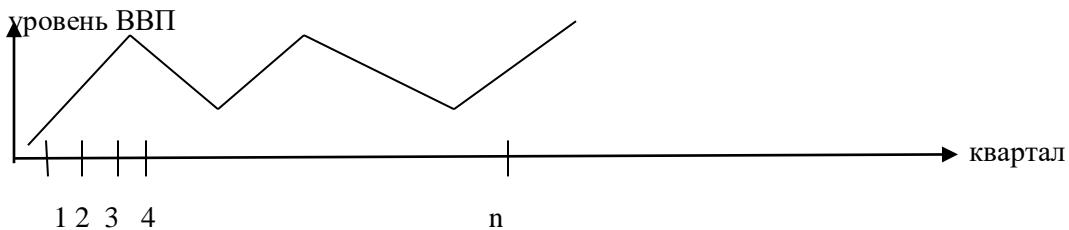
Свойство С) оценок из утверждения А) принято называть *взвешанными наименьшими квадратами*, что является причиной общепринятого названия формулы процедуры (6.4.7) ВМНК.

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

Экономическая переменная \hat{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 санctionам

- 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 $y_t = T(t) + S(t) + C(t) + U_t$, где $C(t)$ - циклическая составляющая, имеющая период отличный от года.

16. Модели тренда временного ряда.

Конкретизируем уравнение тренда $T(t)$.

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 $T(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3$

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 $S(t) = b_1 d_1(t) + b_2 d_2(t) + \dots + b_{\tau-1} d_{\tau-1}(t)$ является периодической с периодом τ

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

Символами $d_1(t), \dots, d_{\tau-1}(t)$ 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.

$$\begin{aligned} d_1 &= \{1 - \text{для первого квартала}, 0 - \text{для других кварталов}\}; \\ d_2 &= \{1 - \text{для второго квартала}, 0 - \text{для других кварталов}\}; \\ d_3 &= \{1 - \text{для третьего квартала}, 0 - \text{для других кварталов}\} \end{aligned}$$

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:

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

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.

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

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

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

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):

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 классе функций, если доставляемые ею оценки параметров обладают следующими свойствами:

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

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

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 несмешённости.

Если справедливы все предпосылки теоремы Гаусса-Маркова, тогда имеет место утверждение A : наилучшая оценка коэффициентов модели рассчитывается по правилу $\vec{a} = (X^T X)^{-1} X^T \vec{y} = Q X^T \vec{y} = M \vec{y}$.

Докажем, что имеет место свойство несмешённости оценок коэффициентов, то есть $E(\vec{a}) = \vec{a}$.
 Доказательство.

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

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

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

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

Доказательство:

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

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

Тогда по теореме Фишера

$$\begin{aligned} Cov(\tilde{a}, \tilde{a}) &= M \cdot Cov(\vec{y}, \vec{y}) \cdot M^T = (X^T P X^{-1}) \cdot X^T P \cdot (\sigma_0^2 P^{-1}) X (X^T P X)^{-1} = \\ &= \sigma_0^2 \cdot (X^T \cdot X)^{-1} = \sigma_0^2 \cdot Q \blacksquare \end{aligned}$$

23. Теорема Гаусса-Маркова: предпосылки и свойство наименьших квадратов

$$\tilde{u}^T \tilde{u} \rightarrow \min.$$

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

0. Столбцы X линейно независимы;
1. $E(u_1) = \dots = E(u_n) = 0$;
2. $Var(u_1) = \dots = Var(u_n) = \sigma_u^2$; - 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):

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

С) Оценки, вычисленные в А, 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 А – 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 экспериментально)

$$p = \frac{\sigma_0^2}{\sigma_u^2} - \text{вес случайного возмущения}$$

$$\text{Величина } \tilde{\sigma}_0^2 = \frac{\tilde{u}^T P^{-1} \tilde{u}}{n - (k + 1)} = \frac{\sum_{i=1}^n p_i \cdot \tilde{u}_i^2}{n - (k + 1)} - \text{несмещенная оценка } \sigma_0^2$$

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.

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

Ковариационная матрица имеет следующий вид:

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

Составим уравнения наблюдений:

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

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

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

(1) удовлетворяют всем предпосылкам теоремы Гаусса-Маркова, следовательно, можем оценить параметры с помощью МНК.

А-Д теоремы Г-М превращаются в следующие утверждения:

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

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

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

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

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

Свойство С) оценок коэффициентов из А) принято называть *взвешенными наименьшими квадратами*, откуда А) - взвешенный метод наименьших квадратов.

26. Обобщённый метод наименьших квадратов (ОМНК).

$$\vec{a} = (X^T P X)^{-1} X^T P \vec{y} = M \cdot \vec{y}$$

Перепишем схему Гаусса-Маркова следующим образом: $\vec{u} = \vec{y} - X\vec{a}$

$$\text{Образуем квадратичную функцию вектора } \vec{a}: F(\vec{a}) = \vec{u}^T P \vec{u} = \sum_{i=1}^n \sum_{j=1}^n p_{ij} u_i u_j.$$

Если справедливы 2-3 предпосылки теоремы Гаусса-Маркова, то квадратичная

$$\text{форма принимает вид: } F(\vec{a}) = \sum_{i=1}^n u_i^2.$$

Найдем вектор коэффициентов \vec{a} , при котором $F(\vec{a}) \rightarrow \min$. Для этого запишем функцию $F(\vec{a})$ с учетом $\vec{u} = \vec{y} - X\vec{a}$

$$F(\vec{a}) = \vec{y}^T P \vec{y} - 2\vec{a}^T X^T P \vec{y} + \vec{a}^T X^T P X \vec{a}$$

$$F(\vec{a}) \rightarrow \min \Leftrightarrow \frac{\partial F(\vec{a})}{\partial \vec{a}} = 0$$

$$\frac{\partial F(\vec{a})}{\partial \vec{a}} = -2X^T P \vec{y} + 2X^T P X \vec{a} \Rightarrow \vec{a}^* = \arg \min(F(\vec{a})) = (X^T P X)^{-1} X^T P \vec{y} = \vec{a}$$

При отказе от 2 и 3 предпосылок ковариационная матрица вектора случайных возмущений

$Cov(\vec{u}, \vec{u})$ является недиагональной, следовательно, генерируется общая структура ковариационной матрицы. В этой ситуации наилучшая оценка коэффициентов рассчитывается следующим образом: $\vec{a} = (X^T P X)^{-1} X^T P \vec{y} = M \cdot \vec{y}$.

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

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

Эта система называется системой нормальных уравнений.

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

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

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

Тогда явный вид решения этой системы:

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

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

Линейная модель парной регрессии имеет вид:

$$\begin{aligned} &\begin{cases} y = a_0 + a_1 x + u \\ E(u) = 0, Var(u) = \sigma_u^2 \end{cases} \\ Cov(\vec{a}, \vec{a}) &= \begin{pmatrix} Var(\tilde{a}_0) & Cov(\tilde{a}_0, \tilde{a}_1) \\ Cov(\tilde{a}_0, \tilde{a}_1) & Var(\tilde{a}_1) \end{pmatrix} \\ Cov(\vec{a}, \vec{a}) &= \sigma_u^2 Q = \sigma_u^2 \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} \end{aligned}$$

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

Получаем:

$$\begin{aligned} Var(\tilde{a}_0) &= \sigma^2 Q_{11} = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \\ Var(\tilde{a}_1) &= \sigma^2 Q_{22} = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ Cov(\tilde{a}_0, \tilde{a}_1) &= \sigma^2 Q_{12} = \sigma^2 Q_{21} = \frac{-\sigma^2 \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{aligned}$$

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 А 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

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

Например, ошибочно выбранная функция (15.2) является линейной по x , в то время как диаграмма рассеивания свидетельствует, что функция регрессии $E(y|x) = f_T(x)$ суть нелинейная функция аргумента x .

Второй симптом - это длительное постоянство знака оценок случайных остатков

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

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

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

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 функцию 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:

$$H_0 : a_1 = 0 \quad (1)$$

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 модели.

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 $(\tilde{S}\tilde{a}_i)$ находится на уровне или превышает абсолютное значение оценки коэффициента при этой переменной (\tilde{a}_i) .

Для определения незначимости 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 тогда следующая дробь:

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

является случайной переменной распределённой 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 модели таких объясняющих переменных в которых справедливо следующее неравенство:

$$|\tilde{a}_j| \leq S\tilde{a}_j |t| \stackrel{?}{\leq} t_{\text{крит}} \quad (2)$$

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 :

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

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 "СТЪЮДЕНТ.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:

$$|t| \leq t_{\text{крит}}$$

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

Замечание. Из курса математической статистики известно, что процедура проверки 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 модели.

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:

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

в ситуации, когда истинной является спецификация:

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

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 коррелированы друг с другом.

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

Последствия гетероскедастичности 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

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

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. Уравнения наблюдений упорядычиваются по возрастанию сумм абсолютных значений объясняющих переменных

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

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

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

Аналогично оценивается модель по последним n_1 уравнениям и запоминается

значение ESS_2 .

Шаг 4. Вычисляется по следующему правилу дробь:

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

Эта дробь является статистикой критерия проверяемой гипотезы о гомоскедастичности случайного возмущения. Величина GQ имеет распределение 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:

$$\begin{cases} GQ \stackrel{?}{\leq} F_{\text{крит}} \\ \frac{1}{GQ} \stackrel{?}{\leq} F_{\text{крит}} \end{cases}$$

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-Квантана.

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

Спецификация AR(1) выглядит следующим образом:

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

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 мультиколлинеарности:

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

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

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 мультиколлинеарности 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 \bar{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 $\bar{R}_0^2 > \bar{R}_1^2$, то удаление x_j нецелесобразно.

Шаг №3. Шаг №1 повторяется пока условие $\bar{R}_j^2 \geq \bar{R}_{j+1}^2$ 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) может быть осуществлено с помошь 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 M1 называется вложенной в модель M2:

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

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

если для модели M2 справедлива гипотеза:

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

Модель M1 иногда также называется моделью M2 в которой справедливо ограничение (7).

Модель M2 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 :

$$W = (ESS_1 - ESS_2) / (ESS_2 / (n - k_2)) \quad (9)$$

$$W \sim \chi_r^2 \text{ при } n \rightarrow \infty; r - \text{число}$$

Пусть справедливы все предпосылки теоремы Гаусса-Маркова при производальном 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 $1 - \alpha$:

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

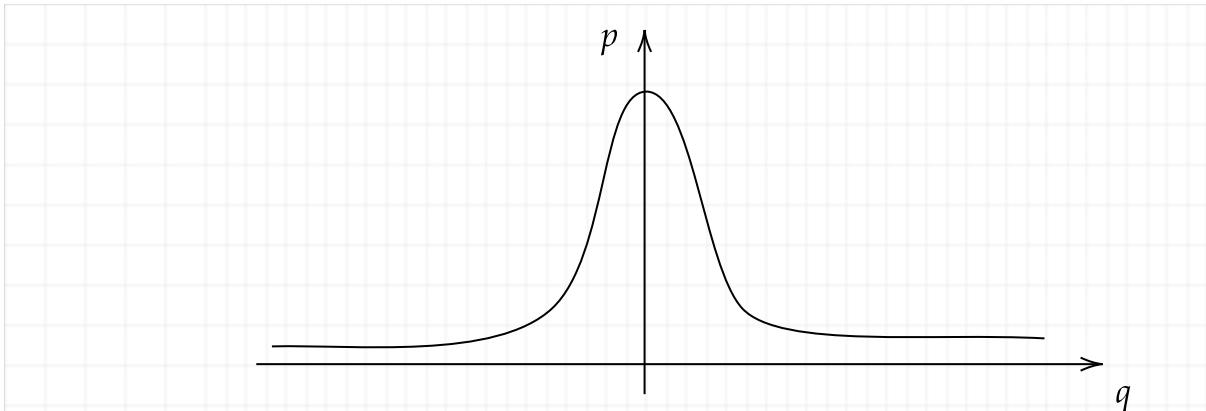
Следствие. Если случайное возмущение в модели M2 имеет нормальный закон распределения, то:

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

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 расширенным вариантом t -теста, причём последний t -тест совпадает с тестом Вальда при $r = 1$.