Эконометрика, 2018-2019, 3 модуль
Семинары 1-2
14.01.19, 21.01.19
Группы Э_Б2016_Э_3
Семинарист О.А.Демидова

Прогнозирование по регрессионной модели

1. На основании 5 наблюдений получена МНК оценка уравнения регрессии $\hat{Y}_i = 1.56 + 0.21X_i$ и оценка остаточной дисперсии $\hat{\sigma}_\varepsilon^2 = 0.04$. Матрица наблюдений регрессоров имеет вид: $X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 3 & 4 & 6 & 8 \end{pmatrix}'$.
Построить 95% доверительный интервал для прогноза, если прогнозное значение *X*=2.

2. На основании наблюдений получена МНК оценка уравнения регрессии

$\hat{Y} = 0.2Z + 0.3W$ и оценка дисперсии ошибок $\hat{\sigma}_\varepsilon^2 = 0.04$.

Матрица наблюдений регрессоров имеет вид: $X' = \begin{pmatrix} 1\ 2\ 0\ 0\ 0 \\ 0\ 0\ 4\ 5\ 6 \end{pmatrix}$.

Ошибки имеют нормальное распределение. Постройте 95% доверительный интервал для индивидуального прогноза в точке Z = -2, W = 5.

Гетероскедастичность, взвешенный МНК

Учебник Демидова и Малахов, задачи 10.1-10.5, задание 10.1-10.4

1) Задача 10.3. (Учебник Демидова, Малахов)
Предположим, что для модели парной регрессии $\sigma_{ui}^2 = \sigma^2 X_i^4$. Как избавиться от проблемы гетероскедастичности ошибок?

2) **Задача 10.4.** (Gujarati, 4 изд., с.429, задача 11.6)
По данным с 1946 г. по 1975 г. E.A. Hanushek и J.E. Jackson (Statistical methods for social scientists, Academic, NY, 1977, c.160) оценили коэффициенты уравнений регрессий (в скобках указаны оценки стандартных отклонений):

$$\hat{C}_t = \underset{(2.73)}{26.19} + \underset{(0.006)}{0.6248}\,GNP_t - \underset{(0.0736)}{0.4398}\,D_t$$

$$\left(\frac{\hat{C}}{GNP}\right)_t = \underset{(2.22)}{25.92}\frac{1}{GNP_t} + \underset{(0.00597)}{0.6246} - \underset{(0.0597)}{0.4315}\left(\frac{D}{GNP}\right)_t,$$

где C – агрегированные частные потребительские расходы,
GNP (Gross National Product) – валовой национальный продукт,
D – национальные расходы на оборону.
С какой целью оценили второе уравнение? Какое при этом было сделано предположение о дисперсии ошибок?

Можно ли сравнивать R^2 в двух регрессиях? Ответ обоснуйте.
Дайте интерпретацию полученным результатам.

2) Учебник Демешев и Борзых

8.2 В модели $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ присутствует гетероскедастичность вида $\mathrm{Var}(\varepsilon_i) = \sigma^2 x_i^2$. Как надо преобразовать исходные регрессоры и зависимую переменную, чтобы устранить гетероскедастичность?

8.3 В модели $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ присутствует гетероскедастичность вида $\mathrm{Var}(\varepsilon_i) = \lambda |x_i|$. Как надо преобразовать исходные регрессоры и зависимую переменную, чтобы устранить гетероскедастичность?

8.4 Известно, что после деления каждого уравнения регрессии $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ на x_i^2 гетероскедастичность ошибок была устранена. Какой вид имела дисперсия ошибок, $\mathrm{Var}(\varepsilon_i)$?

8.5 Известно, что после деления каждого уравнения регрессии $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ на $\sqrt{x_i}$ гетероскедастичность ошибок была устранена. Какой вид имела дисперсия ошибок, $\mathrm{Var}(\varepsilon_i)$?

8.7 По наблюдениям $x = (1, 2, 3)'$, $y = (2, -1, 3)'$ оценивается модель $y = \beta_1 + \beta_2 x + \varepsilon$. Ошибки ε гетероскедастичны и известно, что $\mathrm{Var}(\varepsilon_i) = \sigma^2 \cdot x_i^2$.

1. Найдите оценки $\hat{\beta}_{ols}$ с помощью МНК и их ковариационную матрицу.
2. Найдите оценки $\hat{\beta}_{gls}$ с помощью обобщенного МНК и их ковариационную матрицу.

8.8 Для линейной регрессии $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$ была выполнена сортировка наблюдений по возрастанию переменной x. Исходная модель оценивалась по разным частям выборки:

Выборка	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	RSS
$i = 1, \ldots, 30$	1.21	1.89	2.74	48.69
$i = 1, \ldots, 11$	1.39	2.27	2.36	10.28
$i = 12, \ldots, 19$	0.75	2.23	3.19	5.31
$i = 20, \ldots, 30$	1.56	1.06	2.29	14.51

Известно, что ошибки в модели являются независимыми нормальными случайными величинами с нулевым математическим ожиданием. Протестируйте ошибки на гетероскедастичность на уровне значимости 5%.

8.12 Рассмотрим линейную регрессию $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$ по 50 наблюдениям. При оценивании с помощью МНК были получены результаты: $\hat{\beta}_1 = 1.21$, $\hat{\beta}_2 = 1.11$, $\hat{\beta}_3 = 3.15$, $R^2 = 0.72$.

Оценена также вспомогательная регрессия: $\hat{\varepsilon}_i^2 = \delta_1 + \delta_2 x_i + \delta_3 z_i + \delta_4 x_i^2 + \delta_5 z_i^2 + \delta_6 x_i z_i + u_i$. Результаты оценивания следующие: $\hat{\delta}_1 = 1.50$, $\hat{\delta}_2 = -2.18$, $\hat{\delta}_3 = 0.23$, $\hat{\delta}_4 = 1.87$, $\hat{\delta}_5 = -0.56$, $\hat{\delta}_6 = -0.09$, $R^2_{aux} = 0.36$

Известно, что ошибки в модели являются независимыми нормальными случайными величинами с нулевым математическим ожиданием. Протестируйте ошибки на гетероскедастичность на уровне значимости 5%.

8.15 Рассматривается модель $y_t = \beta_1 + \varepsilon_t$, где ошибки ε_t — независимые случайные величины с $\mathbb{E}(\varepsilon_t) = 0$ и $\mathrm{Var}(\varepsilon_t) = t$. Найдите наиболее эффективную оценку неизвестного параметра β_1 в классе линейных по y и несмещённых оценок.

8.16 Рассматривается модель $y_t = \beta_1 + \varepsilon_t$, где ошибки ε_t — независимые случайные величины с $\mathbb{E}(\varepsilon_t) = 0$ и $\mathrm{Var}(\varepsilon_t) = t^2$. Найдите наиболее эффективную оценку неизвестного параметра β_1 в классе линейных по y и несмещённых оценок.

8.17 Рассматривается модель $y_t = \beta_1 x_t + \varepsilon_t$, где ошибки ε_t — независимые случайные величины с $\mathbb{E}(\varepsilon_t) = 0$ и $\mathrm{Var}(\varepsilon_t) = t$. Найдите наиболее эффективную оценку неизвестного параметра β_1 в классе линейных по y и несмещённых оценок.

8.18 Рассматривается модель $y_t = \beta_1 x_t + \varepsilon_t$, где ошибки ε_t — независимые случайные величины с $\mathbb{E}(\varepsilon_t) = 0$ и $\mathrm{Var}(\varepsilon_t) = t^2$. Найдите наиболее эффективную оценку неизвестного параметра β_1 в классе линейных по y и несмещённых оценок.

8.23 Рассмотрим модель регрессии $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$, в которой ошибки ε_i независимы и имеют нормальное распределение $N(0, \sigma^2)$. Для $n = 200$ наблюдений найдите

1. вероятность того, что статистика Уайта окажется больше 10;
2. ожидаемое значение статистики Уайта;
3. дисперсию статистики Уайта.

8.24 Найдите число коэффициентов во вспомогательной регрессии, необходимой для выполнения теста Уайта, если число коэффициентов в исходной регрессии равно k, включая свободный член.

8.25 По 35 наблюдениям сотрудники НИИ оценили уравнение регрессии $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ и рассчитали остатки ε_i. После того они приступили к диагностике возможных недостатков модели, обнаружили гетероскедастичность и решили её побороть.

1. Самый младший научный сотрудник выдвинул предположение, что стандартное отклонение случайной составляющей может быть выражено так: $\sigma_{\varepsilon,i} = ax_i$, где a — неизвестный коэффициент. Каким образом нужно преобразовать исходное уравнение регрессии, чтобы избавиться от гетероскедастичности?
2. Профессор решил перепроверить результаты и оценил регрессию:
$$\hat{e}_i^2 = -0.3 + 0.08x_i - 0.01x_i^2, R^2 = 0.15.$$
Свидетельствует ли полученный профессором результат о наличии гетероскедастичности?