

1. Гомоскедастичность

Для проверки гипотез мы предполагали условную гомоскедастичность ошибок:

$$E(\varepsilon_i^2|X) = \sigma^2$$

Что произойдет если эта предпосылка будет нарушена?

2. Разница между условной и безусловной гетероскедастичностью

тут вставка

3. Когда логично ожидать гетероскедастичность?

- * безусловной в случайной выборке не бывает

- * условная присутствует почти всегда

- * наличие «размера» объекта

4. В остальном всё ок

Все остальные предпосылки классической модели со стохастическими регрессорами для случайной выборки выполнены.

(тут пачка предпосылок)

5. Мы используем прежние формулы:

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

$$\widehat{Var}(\hat{\beta}) = \frac{RSS}{n-k}(X'X)^{-1}$$

$$\text{В частности, } \widehat{Var}(\hat{\beta}_j) = \frac{\hat{\sigma}^2}{RSS_j}$$

6. Свойства для конечных выборок:

Без требования нормальности ε

* Линейность по y

* Условная несмещенность, $E(\hat{\beta}|X) = \beta$

* (—) Оценки неэффективны

7. С требованием нормальности ε

$$* \left(- \right) \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} | X \sim t_{n-k}$$

$$* \left(- \right) \frac{RSS}{\sigma^2} | X \sim \chi_{n-k}^2$$

$$* \left(- \right) \frac{(RSS_R - RSS_{UR})/r}{RSS_{UR}/(n-k)} \sim F_{r, n-k}$$

8. Асимптотические свойства:

$$* \hat{\beta} \rightarrow \beta$$

$$* \frac{RSS}{n-k} \rightarrow \sigma^2$$

$$* \left(- \right) \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \rightarrow N(0, 1)$$

$$* \left(- \right) \frac{RSS_R - RSS_{UR}}{RSS_{UR}/(n-k)} \rightarrow \chi_r^2$$

9. Мораль:

- * Сами $\hat{\beta}$ можно интерпретировать и использовать
- * Стандартные ошибки $se(\hat{\beta}_j)$ несостоятельны
- * Не можем строить доверительные интервалы

10. Что делать

- * Исправить стандартные ошибки!
- * Вместо $\widehat{Var}(\hat{\beta}) = \frac{RSS}{n-k}(X'X)^{-1}$ использовать
- * $\widehat{Var}(\hat{\beta}) = (X'X)^{-1}X'\hat{\Omega}X(X'X)^{-1}$
- * Уайт, 1980, НС0:

$$\hat{\Omega} = \text{diag}(\hat{\varepsilon}_1^2, \dots, \hat{\varepsilon}_n^2)$$

- * Современный вариант, НС3:

$$\hat{\Omega} = \text{diag}\left(\frac{\hat{\varepsilon}_1^2}{(1-h_{11})^2}, \dots, \frac{\hat{\varepsilon}_n^2}{(1-h_{nn})^2}\right)$$

11. С практической точки зрения:

- * Новая формула, $se_{HC}(\hat{\beta}_j)$

- * R: $vcovHC(model)$

- * С ней жизнь прекрасна!

$$\frac{\hat{\beta}_j - \beta_j}{se_{HC}(\hat{\beta}_j)} \rightarrow N(0, 1)$$

12. Практичный подход!

- * Как только есть случайная выборка и объекты могут быть разного «размера», использовать $se_{HC}(\hat{\beta}_j)$ для проверки гипотез

13. Обнаружение гетероскедастичности

- * Графическое

14. Формальные тесты на гск

- * Тест Уайта

- * Асимптотический, не требуется нормальность остатков

15. Тест Уайта начало

- * Оценить основную регрессию, получить $\hat{\varepsilon}_i$

- * Оценить вспомогательную регрессию:

$$\hat{\varepsilon}_i^2 = \gamma_1 + \gamma_2 z_{i2} + \dots + \gamma_{i,k} z_{im} + u_i$$

z_{i2}, \dots, z_{im} — факторы, определяющие форму гетероскедастичности

Посчитать $LM = nR_{aux}^2$

16. Тест Уайта продолжение

При верной H_0 об условной гомоскедастичности

$$H_0: E(\varepsilon_i^2|X) = \sigma^2$$

$LM \sim \chi_{m-1}^2$, где m — число параметров во вспомогательной регрессии

По умолчанию во вспомогательной регрессии берут исходные регрессоры, их квадраты и попарные произведения

17. вставка тест уайта

По 200 киоскам мороженого и исследователь оценил зависимость спроса (q) от цены (p), разнообразия ассортимента (a) и удаленности от метро (d).

Какой регрессор скорее всего влияет на условную дисперсию ошибок?

Исследователь провел классический тест Уайта и получил $R_{aux}^2 = 0.2$.

Как выглядит вспомогательная регрессия?

Имеет ли место условная гетероскедастичность?

18. Тест Голдфельда-Квандта

- * Есть переменная, от которой может зависеть условная дисперсия ошибок
- * Требуется нормальность ошибок
- * Тест подходит для малых выборок

19. Процедура теста ГК

- * Сортируем наблюдения по предполагаемому убыванию условной дисперсии
- * Выкидываем часть наблюдений посередине (20%)
- * оцениваем исходную модель отдельно по первым и по последним наблюдениям
- * Считаем $F = \frac{RSS_1/(n_1-k)}{RSS_2/(n_2-k)}$

20. Тест ГК-2

* При верной H_0 об условной гомоскедастичности

$$H_0: E(\varepsilon_i^2|X) = \sigma^2$$

$$F \sim F_{n_1-k, n_2-k}$$

21. Вставка

По 200 киоскам мороженого и исследователь оценил зависимость спроса (q) от цены (p), разнообразия ассортимента (a) и удаленности от метро (d).

Чтобы проверить наличие гетероскедастичности исследователь оценил эту модель отдельно по 80 самым удаленным от метро киоскам, получил, $RSS_2 = 120$. По 80 самым близки к метро киоскам, получил, $RSS_1 = 210$.

22. Эффективность?

- * Да, надо смириться с тем, что оценки неэффективны
- * Мы довольны несмещенностью, состоятельностью и возможностью проверять гипотезы
- * Для получения эффективных оценок нужно точно понимать как устроена гетероскедастичность. Это большая редкость.

23. Вставка

Если бы мы знали как устроена гетероскедастичность...

24. Мораль

- * Мы рассмотрели ситуацию нарушения предпосылки условной гомоскедастичности
- * Почти всегда нарушена
- * Неприятность мелкая, мы переживем используя состоятельные стандартные ошибки