

Mikroekonometria

3

Mikołaj Czajkowski
Wiktor Budziński

Zadanie 1. Wykorzystując dane me.wine.dta przygotuj model regresji hedonicznej dla cen wina

1. Przeprowadź regresję objaśniającą ceny wina skupiając się na wpływie liczbie wyprodukowanych skrzynek wina oraz ocenie jego smaku
 1. Zacznij od prostego modelu regresji liniowej
 - ▶ Czy założenia KMRL są spełnione?
 2. Dodaj do modelu efekty nieliniowe
 3. Dodaj do modelu interakcje między wybranymi zmiennymi



Zadanie 2. Wykorzystując dane me.wine.dta przygotuj model regresji hedonicznej dla cen wina

Przeanalizuj problem heteroskedastyczności wykorzystując test Breusha-Pagana w różnych wersjach

- ▶ Wynik testu wskazujący na heteroskedastyczność może być także spowodowany niepoprawną formą funkcyjną modelu lub 'nienormalnymi' resztami



Testowanie heteroskedastyczności w STATA

- ▶ Pomimo heteroskedastyczności estymator MNK wciąż jest zgodny i nieobciążony. Przestaje być on jednak efektywny (wartości błędów standardowych mogą być mylące)
- ▶ Rozwiązanie – tzw. odporne macierze kowariancji:
 - ▶ Macierz White'a: $(\mathbf{X}'\mathbf{X})^{-1} \times \sum_{i=1}^n e_i^2 X_i X_i' \times (\mathbf{X}'\mathbf{X})^{-1}$
 - ▶ Bądź inne alternatywy dostępne w STATA

Zadanie 3. Wykorzystując dane `me.wine.dta` przygotuj model regresji hedonicznej dla cen wina

Dokonaj estymacji modelu regresji liniowej wykorzystując odporne macierze kowariancji

- ▶ Czy wnioski z tak uzyskanych wyników się zmieniają?



Uogólniona Metoda Najmniejszych Kwadratów

- ▶ W MNK estymatorem jest: $b_{MNK} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$
- ▶ W UMNK estymatorem jest: $b_{UMNK} = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1} \mathbf{X}'\Omega^{-1}\mathbf{y}$
 - ▶ Ω to znana macierz kowariancji błędów losowych
 - ▶ W przypadku występowania heteroskedastyczności jest to macierz diagonalna z elementami na przekątnej w postaci $\sigma^2(\mathbf{Z}_i)$
 - ▶ Jeśli $\sigma^2(\mathbf{Z}_i)$ jest znane to następujący model heteroskedastyczny:

$$y_i = \mathbf{X}_i\boldsymbol{\beta} + \sigma(\mathbf{Z}_i)\varepsilon_i$$

można przekształcić do:

$$y_i/\sigma(\mathbf{Z}_i) = (\mathbf{X}_i\boldsymbol{\beta})/\sigma(\mathbf{Z}_i) + \varepsilon_i$$

- ▶ Powyższy model już nie jest heteroskedastyczny, więc estymator MNK jest efektywny. Estymację takiego modelu nazywamy Ważoną Metodą Najmniejszych Kwadratów (WMNK).
 - ▶ Jest to szczególny przypadek UMNK

Ważona Metoda Najmniejszych Kwadratów

- ▶ Załóżmy, że wiemy, że wariancja składnika losowego jest proporcjonalna do średniej zmiennej objaśnianej. To jest, mamy model następującej postaci: $y_i = \mathbf{X}_i\boldsymbol{\beta} + \sigma(\mathbf{X}_i)\varepsilon_i$
 - ▶ Gdzie $\sigma^2(\mathbf{X}_i) = \sigma^2[\mathbf{X}_i\boldsymbol{\beta}]^2$
 - ▶ Model ten można estymować dwustopniowo

Zadanie 4. Wykorzystując dane me.wine.dta przygotuj model regresji hedonicznej dla cen wina

Dokonaj estymacji modelu ważonej regresji liniowej

- ▶ Czy wnioski z tak uzyskanych wyników się zmieniają?



Modelowanie heteroskedastyczności

- ▶ W praktyce dokładna postać wariancji jest zazwyczaj nieznaną
- ▶ Możemy chcieć estymować zależność między wariancją składnika losowego, a zmiennymi objaśniającymi (tak samo jak w przypadku zmiennej objaśnianej)
- ▶ Stata to (częściowo) umożliwia pozwalając na estymację zależności wariancji składnika losowego od zmiennych objaśniających zgodnie z następującą postacią funkcyjną: $\sigma^2(\mathbf{Z}_i) = \sigma^2 \exp(\mathbf{Z}_i \boldsymbol{\gamma})$
- ▶ Model estymowany przy pomocy Metody Największej Wiarygodności
 - ▶ Stata nie posiada bezpośredniej funkcji pozwalającą na taką estymację modeli liniowych
 - ▶ Można to obejść wykorzystując model ARCH
 - ▶ Model autoregresji z heteroskedastycznością warunkową (ang. Auto-Regressive Conditional Heteroskedasticity model)
 - ▶ W modelu ARCH zakłada się, że wariancja błędu losowego w danym okresie jest funkcją wartości błędów losowych w okresach poprzednich
 - ▶ Stosuje się go głównie do analizy szeregów czasowych, np. w analizie finansowej zmienności cen instrumentów finansowych.
 - ▶ Zaproponowany w 1982 roku przez Roberta Englea (Nobel 2003)

Zadanie 5. Wykorzystując dane `me.wine.dta` przygotuj model regresji hedonicznej dla cen wina

Dokonaj estymacji modelu ARCH, w którym wariancja zależy (log) liniowo od zmiennych objaśniających

- ▶ Czy w analizowanym zbiorze danych występuje heteroskedastyczność?



Modelowanie heteroskedastyczności

- ▶ Po co w ogóle modelować wariancję, skoro oszacowania parametrów są zgodne i nieobciążone, a można zastosować odporne macierze kowariancji?
 - ▶ Odporne macierze kowariancji nie zmieniają oszacowań parametrów, a regresja z heteroskedastycznością tak
 - ▶ Teoretycznie, dla dużych prób, powinny być bardzo zbliżone, ale w praktyce różnice mogą być znaczne
 - ▶ Ma to znaczenie jeśli interesuje nas elastyczność (model na logarytmach)
 - ▶ Jeśli heteroskedastyczność występuje, to jej uwzględnienie ma znaczenie dla prognozowania
 - ▶ Np. rozważmy model w postaci: $\log(y_i) = \mathbf{X}_i\boldsymbol{\beta} + \sigma(\mathbf{X}_i)\varepsilon_i$
 - ▶ Jeśli chcielibyśmy prognozować średnie y_i przy pewnym poziomie zmiennych objaśnianych to musimy także uwzględnić ich wpływ na średnie wartości składnika losowego: $E(y_i | \mathbf{X}_i^*) = \exp(\mathbf{X}_i^*\boldsymbol{\beta} + 0,5\sigma^2(\mathbf{X}_i^*))$
 - Bo średnia rozkładu lognormalnego to $\exp(\mu + \sigma^2/2)$

Praca domowa ME.3

Wykorzystując dane `me.juice.dta` oraz model regresji liniowej sprawdź jaki jest wpływ ceny na sprzedaż danego soku

1. Jako zmienną objaśnianą wybierz liczbę sprzedanych sztuk soku wybranej firmy (zmiennie zaczynające się od `sal_`)
 2. Zbuduj model regresji liniowej w którym cena tego soku (zmiennie zaczynające się od `pri_`) oraz inne wybrane zmiennne tłumaczą sprzedaż
 1. Proszę nie uwzględniać w modelu zmiennych opisujących id sklepu
 3. Przeprowadź diagnostykę heteroskedastyczności, formy funkcyjnej modelu oraz normalności błędu losowego
 4. Sprawdź czy związek między liczbą sprzedanych sztuk soku wybranej firmy a ceną tego soku (zmienna zaczynająca się od `pri_`) jest liniowy
 1. Analogicznie jak na zajęciach, proszę spróbować zlogarytmować zmiennne (zarówno zmienną objaśnianą jak i objaśniającą), uwzględnić ich kwadraty, dodać jakieś interakcje, a następnie sprawdzić czy poprawia to formę funkcyjną lub czy dodatkowe przekształcenia wychodzą istotne.
 5. Przygotuj model (ARCH), w którym dodatkowo wariancja błędu losowego będzie objaśniana za pomocą wybranych zmiennych
- ▶ Do przygotowania w grupach trzyosobowych

Praca domowa

- ▶ Rozwiązania pracy domowej wyślij na wbudzinski@wne.uw.edu.pl
 - ▶ Termin – niedziela (przed kolejnymi zajęciami) do 23:59
 - ▶ W tytule emaila wpisz tylko
 - ▶ Praca domowa ME.<nr zajęć>
 - ▶ W treści emaila podaj wyłącznie osoby przygotowujące tę pracę, w następującym formacie:
 - <nr grupy1> <imię1> <nazwisko1> <nr indeksu1>
 - <nr grupy2> <imię2> <nazwisko2> <nr indeksu2>
 - <nr grupy3> <imię3> <nazwisko3> <nr indeksu3>
 - ...
 - ▶ Nr grupy odnosi się do list studentów podanych na stronie przedmiotu
 - ▶ Rozwiązania wszystkich zadań (jeśli jest więcej niż jedno) zamieść w jednym pliku .do załączonym do emaila
 - ▶ Plik do nazwij według schematu:
 - <nazwisko osoby przysyłającej>.do
 - ▶ Kod programu opatrz niezbędnymi komentarzami, w tym oznaczając kolejność wykonywanych zadań, **podając interpretację wyników estymacji i przeprowadzanych testów**, odpowiedzi na pytania (jeśli były zadane) i inne wyjaśnienia niezbędne do prześledzenia rozumowania. Na koniec przedstaw wnioski z przeprowadzonej analizy.
- ▶ Ścisłe zastosowanie się do powyższych reguł jest warunkiem przyznania pełnej liczby punktów z pracy domowej

Praca domowa

▶ Przykład maila:

- ▶ Od: Mikołaj Czajkowski
- ▶ Do: Wiktor Budziński (wbudzinski@wne.uw.edu.pl)
- ▶ Temat: Praca domowa ME.3
- ▶ Treść:
 - ▶ 107 Mikołaj Czajkowski 100001
 - ▶ 107 Kamil Czajkowski 100002
 - ▶ 108 Andrzej Czajkowski 100003
- ▶ Załącznik: Czajkowski.do
- ▶ Analizę przeprowadź wykorzystując losową podpróbę z zapewnionego zbioru danych (zachowaj 90% obserwacji)
 - ▶ Ustaw ziarno jako „10 + nr indeksu osoby z grupy”
 - ▶ Przykład obok dla indeksu 311563
 - ▶ W pracy domowej ME.3 chcemy zachować w próbie obserwacje dla wszystkich sklepów
 - ▶ Opcja *by()*

```
set seed 10311563
sample 90, by(store2)
```