Wykłady Jacka Osiewalskiego

z Ekonometrii

zebrane ku pouczeniu i przestrodze

UWAGA!! (listopad 2003) to jest wersja nieautoryzowana, spisana przeze mnie dawno temu i od tego czasu nie przejrzana; ma status wersji roboczej, może zawierać błędy i nieścisłości zawinione wyłącznie przez przepisującego. Wykłady były przepisywane (i nieco komentowane) w ułomny sposób i taką wersję udostępniam na zasadzie „tak jak jest" na własną odpowiedzialność czytającego. Notatki obejmują kilka wykładów z semestru zimowego w nieco innym zakresie programowym niż w bieżącym roku - Błażej Mazur

CZĘŚĆ PIERWSZA: Modele Regresji

Wersja robocza 20 XI 2001

Spis treści:

UWAGA!! (listopad 2003) to jest wersja nieautoryzowana, spisana przeze mnie dawno temu i od tego czasu nie przejrzana; ma status wersji roboczej, może zawierać błędy i nieścisłości zawinione wyłącznie przez przepisującego. Wykłady były przepisywane (i nieco komentowane) w ułomny sposób i taką wersję udostępniam na zasadzie „tak jak jest" na własną odpowiedzialność czytającego. Notatki obejmują kilka wykładów z semestru zimowego w nieco innym zakresie programowym niż w bieżącym roku - Błażej Mazur

1.1 Wprowadzenie

Tu następują wstępne uprzejmości, próba zarysowania przedmiotu, odpowiedzi na pytanie, co to jest ekonometria, jak się ma do innych eko, statystyki matematycznej itd. idt
Co to jest teoria ekonometrii?

Modele statystyczne + metody wnioskowania
Co to jest model ekonometryczny:

Model statystyczny: (intuicja a nie formalny opis) {Kiedy Profesor mówi o intuicjach to nie znaczy, że to jest nieważne, tylko że zrobione tak jak trzeba, czyli opisem formalnym, byłoby za trudne. Opis formalny to trudne narzędzie, ale ekonomiczne w użyciu i niezastąpione BM}a więc:

Model statystyczny:

Układ założeń stochastycznych (probabilistycznych) opisujący hipotetyczny mechanizm generowania obserwacji (powstawania danych gospodarczych)

Na etapie modelowania obserwacje mogą być rozumiane jako „potencjalnie możliwe obserwacje" (ujęcie ex ante, bez znajomości konkretnych liczb reprezentujących zjawiska gospodarcze), modeluje się je za pomocą zmiennych losowych.

Wychodząc od modelu ekonomicznego uzupełnia się go o założenia stochastyczne.

1.2 Klasyczny Model Regresji Liniowej (KMRL)

Jest to model przeniesiony z doświadczalnictwa przyrodniczego, warunkowo tylko adekwatny w badaniach ekonomicznych.

Jest to model jednorównaniowy, układ wyjaśniający powstawanie obserwacji na jednej zmiennej endogenicznej; y_t to obserwacja o numerze t na wyróżnionej zmiennej endogenicznej y zwanej w kontekście regresji zmienną objaśnianą.

$$y_t \qquad (t = 1,2, \ldots ,T)$$

Zakłada się, że model wyjaśnia powstawanie T obserwacji o numerach od 1 do T.
Obserwacje te ustawione są w wektor kolumnowy: wektor „obserwacji na zmiennej objaśniającej igrek" **y**.

{Tu spróbuję trzymać się następującej konwencji: macierze i wektory będą oznaczane wytłuszczeniem (**bold**), wektory oznaczam małymi literami, (przyjmuję, że wektory są zawsze wektorami kolumnowymi), macierze wielkimi literami, nazwy zmiennej nie wytłuszczam; zmienne losowe nie będą w zapisie odróżniane od ich realizacji! BM}

$$\underset{(T\times 1)}{\mathbf{y}} = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_T \end{bmatrix}$$

Kolejne y_t to nie liczby z roczników czy księgowości czy skądkolwiek. Na tym etapie trzeba o nich myśleć zupełnie inaczej. To są możliwe obserwacje (ujęcie ex ante). Tworzony jest schemat, który mówi jak rzeczywiste obserwacje mogłyby się pojawić (patrz „model" wyżej) – obserwacja jest tu zmienną losową: reprezentuje to, co mogłoby się wydarzyć, a nie to, co zarejestrowano. Zarejestrowane liczby to realizacje zmiennej losowej, a na etapie modelowania „możliwa realizacja" zjawiska empirycznego, czyli wektor **y** to wektor losowy, który będzie dalej opisywany.

Zmienne zewnętrzne (w kontekście regresji „objaśniające" <explanatory variables> będą zwykle dobierane w oparciu o modele ekonomiczne. Zakłada się, że aby opisać powstawanie **y** trzeba i wystarczy wyróżnić k („ka małe") odrębnych zmiennych objaśniających x. Każdej obserwacji na zmiennej objaśnianej y odpowiada po jednej wartości każdej z k zmiennych objaśniających x_i ; (i = 1,...k).

Macierz **X** (macierz zmiennych objaśniających, macierz regresorów, macierz planu eksperymentu <design matrix>) tworzona jest w ten sposób, że kolejne kolumny odpowiadają różnym zmiennym objaśniającym a w ramach kolumny (w kolejnych wierszach) występują wartości danej zmiennej objaśniającej.

$$\underset{(T\times k)}{\mathbf{X}} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ x_{T1} & x_{T2} & \cdots & x_{Tk} \end{bmatrix}$$

T obserwacji na zmiennej objaśnianej tworzy wektor **y**, a w macierzy **X** jest po T wartości wszystkich k zmiennych objaśniających, każdej zmiennej odpowiada jedna kolumna macierzy **X**, a kolejnym jej wartościom (odpowiadającym kolejnym obserwacjom na y) posuwanie się do kolejnego wiersza. Przy małych iksach pierwsza liczba to numer obserwacji a druga to numer zmiennej. Dobrze jest się oswoić z tym zapisem, opłaca się wiedzieć bez wahania co jest „Te duże" a co „ka małe". Przy każdej macierzy warto odruchowo pomyśleć o wymiarze, co w wierszach, co w kolumnach i co to jest element a_{ij} .

Czy można powiedzieć, że macierz **X** grupuje obserwacje na zmiennych objaśniających? Tu są zastrzeżenia. Bo trzeba by powiedzieć, że są to obserwacje bez błędów – gdyby założyć, że są to realne obserwacje, obarczone np. błędem pomiaru, trzeba byłoby założyć ich losowość – będzie to przedmiotem jednego z kolejnych podrozdziałów. Na tym etapie bezpieczniej jest nazywać macierz **X** macierzą wartości zmiennych objaśniających, a nie obserwacji.

Ponadto praktycznie zawsze (ale niekoniecznie na kolokwium ☺ to dopisek AD 2003 kiedy przepisywacz niniejszego ze sprawdzanego kolokwiami stał się sprawdzającym kolokwia) w dalszej fazie, post-modelowej (czyli na ćwiczeniach;-)) w macierzy **X** pierwsza kolumna będzie zawierać same jedynki, co odpowiada uwzględnieniu wyrazu wolnego.

1.2.1 Założenia Klasycznego Modelu Regresji Liniowej

1° $\underset{(T\times 1)}{\mathbf{y}} = \underset{(T\times k)}{\mathbf{X}}\,\underset{(k\times 1)}{\boldsymbol{\beta}} + \underset{(T\times 1)}{\boldsymbol{\varepsilon}}$

2° $\mathbf{X}$ jest znaną macierzą nielosową

3° $rz(\mathbf{X}) = k$

4° $E(\boldsymbol{\varepsilon}) = \underset{(T\times 1)}{\mathbf{0}}$

5° $V(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I_T}$ $\qquad \sigma^2>0$

To jest zapis formalny założeń. A co one po kolei znaczą i jak się je czyta?

1° $\underset{(T\times 1)}{\mathbf{y}} = \underset{(T\times k)}{\mathbf{X}}\,\underset{(k\times 1)}{\boldsymbol{\beta}} + \underset{(T\times 1)}{\boldsymbol{\varepsilon}}$

Wektor $\boldsymbol{\beta}$ (beta) to wektor kolumnowy grupujący nieznane stałe liczbowe. Stałe te wyznaczają liniowy wpływ zmiennych objaśniających na zmienne objaśniane. Tych stałych jest k, tyle, ile zmiennych objaśniających. Wektor $\boldsymbol{\beta}$ ma postać:

$$\underset{(k\times 1)}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ . \\ \beta_k \end{bmatrix}$$

{przy tym czasem zdarza się β_0, ale z formalnego punktu widzenia jest to β_1, itd... chodzi o to, że współczynników beta jest k „ka małe”}

Wektor losowy $\boldsymbol{\varepsilon}$ (epsilon) to wektor składników losowych poszczególnych obserwacji:

$$\underset{(T\times 1)}{\boldsymbol{\varepsilon}} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_T \end{bmatrix}$$

Jak widać całkiem podobnie do wektora beta, ale epsilonów jest Te duże; ponadto wektor $\boldsymbol{\beta}$ jest nieznanym wektorem nielosowym, natomiast $\boldsymbol{\varepsilon}$ jest wektorem losowym - wielowymiarową zmienną losową. Jako taka ma bardziej złożone charakterystyki od jednowymiarowej zmiennej losowej. Wielowymiarowa zmienna losowa może (ale nie musi) mieć wektor wartości oczekiwanych i macierz kowariancji, powyżej zakłada się (4°,5°) że $\boldsymbol{\varepsilon}$ je ma, że jest zmienną losową mającą momenty pierwszego i drugiego rzędu.

Ponieważ wektor losowy **y** tworzy się jako suma wielkości nielosowej **Xβ** i wektora losowego **ε**, jego charakterystyki stochastyczne będą odpowiednią transformacją charakterystyk **ε**.

Zakłada się tu, że możliwa do zaobserwowania wartość **y** tworzy się jako suma tzw. składowej deterministycznej i składowej losowej. Część deterministyczna to **Xβ**. To jest macierzowy zapis jednej strony układu T równań liniowych. Każde równanie odpowiada obserwacji o innym numerze i zawiera te same parametry (β) i odpowiednie (o numerze t) wartości wszystkich zmiennych objaśniających. Czyli np. druga możliwa obserwacja na y składa się ze składnika deterministycznego który powstaje tak: β_1 razy x_{21} plus β_2 razy x_{22} plus ... plus β_k razy x_{2k} (koniec składnika deterministycznego) plus składowa stochastyczna czyli ε_2 koniec. Składowa deterministyczna obserwacji numer j powstaje zawsze tak: j-ty wiersz macierzy **X** razy wektor **β**. Potem następna obserwacja: kolejny (j+1) wiersz macierzy **X** razy ten sam wektor **β**. Parametry beta odzwierciedlają za liniowy wpływ zmiennych objaśniających na zmienną objaśnianą. Zależność liniowa ze stałymi parametrami jest prostą konstrukcją o ograniczonych możliwościach, jednak na jej bazie konstruować można bardziej wyrafinowane modele. Ostatecznie:

$$y_t = \beta_{1*}x_{t1} + \beta_{2*}x_{t2} \ \ ... + \beta_{k*}x_{tk} \ + \varepsilon_t$$

{teraz wyjątkowo napisałem * razy, ale potem już nie, bo jak mawiał prof. Andrzej Malawski: ... nie piszemy, ale pamiętajmy, że ta kropka, której tu nie ma, to zupełnie inna kropka niż ta, której tam nie piszemy...) BM}

Wektor **ε** to wektor składników losowych (poszczególnych obserwacji) reprezentujących łączny wpływ wszystkich czynników drugorzędnych, przypadkowych, nie uwzględnionych wśród zmiennych objaśniających. Dodanie do składowej deterministycznej wektora zakłóceń losowych **ε** ma modelować fakt, że zarejestrowane obserwacje mogą różnić się co do wartości od wielkości wynikających z teoretycznej konstrukcji modelu ekonomicznego. Wektor **ε** grupuje składniki losowe które są z definicji nieobserwowalne, postulujemy ich istnienie, by wyjaśnić wszelkie rozbieżności między teoretycznymi wartościami zmiennej objaśnianej a wartościami zaobserwowanymi.

W teorii jeśli wiemy ile jest istotnych zmiennych objaśniających (k) i jeśli wiemy, jaka jest postać zależności (liniowa), czyli jeśli 1° jest prawdziwe, to epsilon obejmuje tylko czynniki przypadkowe, drugorzędne, zakłócające. W praktyce jednak trzeba liczyć się z tym, że składnik losowy obejmuje też konsekwencje następujących błędów:

1. błędu specyfikacji (pominięcie istotnej zmiennej, włączenie nieistotnej itd.)
2. błędu aproksymacji (jeśli postać zależności jest inna czyli np. istotnie nieliniowa, i nie jest dobrze przybliżana postacią liniową).

W praktyce liczymy się z tym, że 1° nie jest idealnie spełnione (ale możemy zakładać, że dobrze dobraliśmy zmienne objaśniające i że prawdziwa postać zależności jest dobrze przybliżana przez zależność liniową...). Dokładniej własności **ε** zostaną opisane w punktach 4° - 5°.

2° **X** jest znaną macierzą nielosową

Założenie drugie jest przejęte z doświadczalnictwa przyrodniczego. To odpowiada sytuacji gdy eksperymentator „zadaje" pewne wartości „na wejściu" i obserwuje inne wartości „na wyjściu" czyli kontroluje przebieg eksperymentu. Wielkości wejściowe są zadane, nielosowe. W zagadnieniach ekonomicznych założenie o prowadzeniu kontrolowanego eksperymentu jest co najmniej dyskusyjne i można je – czyli też założenie 2° - przyjmować wyłącznie jako przybliżenie.

3° $rz(\mathbf{X}) = k$

Zakładamy, że rząd macierzy **X** jest równy k. Jest to założenie o charakterze technicznym {będziemy macierz **X**'**X** odwracać wielokrotnie, a $(\mathbf{X}'\mathbf{X})^{-1}$ istnieje gdy rz(**X**) = k BM} oznaczające, że kolumny macierzy **X** są liniowo niezależne. To odzwierciedla intuicję o niedublowaniu informacji – do modelu wprowadza się liniowych kombinacji uwzględnionych już zmiennych (ale nieliniowe można – np. w translogarytmicznej funkcji produkcji). To ustala też minimalną liczbę obserwacji: z 3° wynika, że T ≥ k (z definicji rzędu macierzy).

Kolejne założenia (4° i 5°) dotyczą wektora epsilon:

4° $\quad E(\boldsymbol{\varepsilon}) = \underset{(T\times 1)}{\mathbf{0}}$

Czytamy: wartość oczekiwana wektora losowego **ε** jest wektorem zerowym. Zapis $\mathbf{0}_{(T\times 1)}$ oznacza wektor kolumnę o wymiarach T na 1 zawierający wyłącznie zera. Wartość oczekiwana wektora losowego **ε** to wektor zawierający odpowiednio wartości oczekiwane składników losowych poszczególnych obserwacji:

$$\underset{(T\times 1)}{E(\boldsymbol{\varepsilon})} \overset{df}{=} \begin{bmatrix} E(\varepsilon_1) \\ E(\varepsilon_2) \\ \cdot \\ \cdot \\ \cdot \\ E(\varepsilon_T) \end{bmatrix} \overset{4^\circ}{=} \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} = \underset{(T\times 1)}{\mathbf{0}}$$

{to, co jest nad znakiem równości czytamy „na mocy" i będzie nadużywane na następnym wykładzie; df (czasem trzy poziome kreski albo :=) oznacza „na mocy definicji" czyli „jest zdefiniowany jako" a czwórka „na mocy założenia numer 4" - to tak dla jasności.... BM}

Założenie to oznacza, że łączny wpływ czynników drugorzędnych, przypadkowych nie może mieć tendencji, składnika systematycznego. Oczekujemy, że średnio rzecz biorąc, łączny wpływ tych czynników będzie odchylał wartość y_t to w górę, to w dół, i że wartość oczekiwana tych odchyleń powinna być zerowa.

5° $\quad V(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_T \qquad$ i $\qquad \sigma^2 > 0$

Założenie 5° mówi o postaci macierzy wariancji-kowariancji wektora losowego **ε**. Macierz wariancji – kowariancji zwana dalej macierzą kowariancji to macierz zawierająca podstawowe charakterystyki rozproszenia wielowymiarowej zmiennej losowej. Jednowymiarowa zmienna losowa zwykle może być charakteryzowana miarą położenia (wartość oczekiwana) i rozproszenia (wariancja bądź odchylenie standardowe). [„Zwykle", gdyż wielkości te mogą nie istnieć w wypadku niektórych zmiennych losowych. Tu założenia 4° i 5° mówią, że **ε** charakteryzuje się rozkładem posiadającym momenty pierwszego i drugiego rzędu]. Wektor losowy **ε** jest wielowymiarową zmienną losową charakteryzowaną miarą położenia – omówioną w poprzednim punkcie wartością oczekiwaną będącą wektorem – oraz miarą rozproszenia – macierzą kowariancji, która określa wariancję poszczególnych składowych wektora losowego i zależności pomiędzy nimi – odpowiednie kowariancje. Ponieważ wariancję możemy uważać za kowariancję pomiędzy zmienną a nią samą, macierz w której zebrane są wszystkie kowariancje charakteryzujące wektor losowy **ε** nazywamy macierzą kowariancji i oznaczamy jako **V**(**ε**); macierz kowariancji musi być dodatnio określona (więc i symetryczna, skoro symetryczna to i kwadratowa oczywiście); σ^2 (sigma kwadrat) to skalar, nieznany większy od zera parametr struktury stochastycznej. Kwadrat jest tu tylko dlatego, że sama sigma to tradycyjnie nieznane odchylenie standardowe czyli pierwiastek z wariancji. $\mathbf{I}_T$ to macierz jednostkowa stopnia T, czyli

macierz kwadratowa, mająca jedynki na przekątnej głównej i zera poza nią, jest to element neutralny w mnożeniu macierzy.

Co oznacza zapis: $\mathbf{V}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_T$? Macierz opisana tym wzorem ma na przekątnej σ^2 i zera wszędzie indziej.

$$\underset{(T\times T)}{\mathbf{V}(\boldsymbol{\varepsilon})} \overset{df}{=} \begin{bmatrix} var(\varepsilon_1) & cov(\varepsilon_1,\varepsilon_2) & \dots & cov(\varepsilon_1,\varepsilon_T) \\ cov(\varepsilon_2,\varepsilon_1) & var(\varepsilon_2) & \dots & \dots \\ \dots & \dots & \dots & \dots \\ cov(\varepsilon_T,\varepsilon_1) & \dots & \dots & var(\varepsilon_T) \end{bmatrix} \overset{5^\circ}{=} \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & \dots \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix}$$

Założenie piąte można przeczytać następująco: macierz kowariancji $\boldsymbol{\varepsilon}$ jest znana z dokładnością do czynnika skalującego, czyli do nieznanego skalara przez który ją mnożymy. Taka postać założenia o macierzy kowariancji składnika losowego jeszcze się pojawi. Określamy tu strukturę macierzy kowariancji wektora losowego $\boldsymbol{\varepsilon}$ (epsilon). Określamy ją bardzo restrykcyjnie: kowariancje (i korelacje) między składnikami losowymi różnych obserwacji są znane i wynoszą 0, wariancje składników losowych poszczególnych obserwacji są nieznane, jednakowe i wynoszą σ^2. Założenie o stałości wariancji $\boldsymbol{\varepsilon}$ to założenie homoskedastyczności (w przeciwieństwie do heteroskedastyczności, która ma miejsce, gdy uchylimy homoskedastyczność). Oznacza ono, że czynniki przypadkowe, drugorzędne powodują podobne odchylenie (co do wielkości) dla każdej obserwacji. Założenie o zerowych kowariancjach między składnikami różnych obserwacji to założenie braku autokorelacji: składniki losowe różnych obserwacji są nieskorelowane. Współczynnik korelacji liniowej pomiędzy składnikami losowymi różnych obserwacji wyraża się następująco:

$$r(\varepsilon_t,\varepsilon_s) \overset{df}{=} \frac{cov(\varepsilon_t,\varepsilon_s)}{\sqrt{var(\varepsilon_t)*var(\varepsilon_s)}} \overset{5^\circ}{=} \frac{0}{\sigma^2} = 0 \text{ , dla } t \neq s$$

Współczynnik r określa kierunek i siłę przybliżonej zależności liniowej pomiędzy zmiennymi – tu jego wartość oznacza: nie ma przybliżonej zależności liniowej pomiędzy składnikami losowymi poszczególnych obserwacji. Gdyby to założenie uzupełnić o założenie o normalnym rozkładzie $\boldsymbol{\varepsilon}$, byłoby to równoznaczne z niezależnością stochastyczną składników losowych poszczególnych obserwacji. (niezależność jest równoznaczna z zerową korelacją w rozkładzie normalnym, w KMRL normalności $\boldsymbol{\varepsilon}$ nie zakładamy, jednak zerowe korelacje sugerują intuicje biegnące w podobnym kierunku) Niezależność stochastyczna jest formalnym ujęciem przekonania o niezależności w sensie potocznym: każda kolejna obserwacja jest niezależna od przeszłości. Tu zakładamy nieskorelowanie co może również odzwierciedlać potoczne pojęcie niezależności. Zauważmy, że założenie to („niezależność" zakłóceń losowych poszczególnych obserwacji) może łatwo nie być spełnione nawet w warunkach eksperymentu kontrolowanego, a co dopiero w ekonomii. Jest to bardzo mocne założenie, i tak jak pozostałe implikacje punktu 5 może podlegać weryfikacji i może zostać uchylone.

***

Ponadto można zauważyć, że KMRL nie ma struktury obserwacji tzn. można by przemieszać kolejność obserwacji (poprzestawiać wiersze macierzy $\mathbf{X}$ i tak samo wiersze wektora $\mathbf{y}$) i nic by się nie zmieniło. Czyli kiedy wprowadzamy do modelu szereg czasowy, to stosując KMRL tracimy całą strukturę czasową, skazujemy się na utratę części informacji, bo szereg czasowy jest z natury sztywno uporządkowany a używamy go w modelu gdzie porządek obserwacji nie ma znaczenia i można go co najwyżej wprowadzić sztucznie (zmienna czasowa).

Model KMRL jest mimo swych mocnych, nierealistycznych założeń modelem bardzo ważnym. Jego modyfikacje polegające na uchyleniu czy uogólnianiu założeń prowadziły do bardziej złożonych i bardziej realistycznych modeli. (aby je zrozumieć, dobrze jest wyjść od KMRL)

Omówiony układ założeń prowadzi do postawienia pytania o $\boldsymbol{\beta}$ (wektor nieznanych parametrów regresji). Jest on naturalnym przedmiotem zainteresowania, jego elementy określają wielkość i kierunek wpływu (liniowego, ale zakładamy, że tylko on zachodzi – poza tym są tylko „czynniki przypadkowe i drugorzędne") zmiennych objaśnianych na zmienną objaśniającą; wpływ ten chcemy oszacować, w związku z tym konieczne jest wnioskowanie o wektorze nieznanych stałych parametrów $\boldsymbol{\beta}$. Zagadnieniem jest więc estymacja (szacowanie) wektora współczynników regresji $\boldsymbol{\beta}$. Podstawowe twierdzenie w tym zakresie to twierdzenie Gaussa i Markowa o estymatorze MNK $\hat{\boldsymbol{\beta}}$:

1.2.2 Twierdzenie Gaussa i Markowa o estymatorze MNK

Twierdzenie Gaussa i Markowa o estymatorze MNK (metody najmniejszych kwadratów) :

W KMRL (czyli przy prawdziwości założeń 1° - 5°) :
Najlepszym estymatorem wektora współczynników regresji β w klasie estymatorów liniowych i nieobciążonych jest estymator MNK dany następującym wzorem: [wersja Profesora: dany *prostym* wzorem macierzowym:]

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

Macierz kowariancji estymatora MNK dana jest wzorem:

$$\mathbf{V}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

{' oznacza transpozycję macierzy}

Dowód tego twierdzenia podany będzie poniżej i poprzedzony zostanie komentarzem odnoszącym się do jego treści i szeregiem kroków przygotowawczych. W tytule i w treści twierdzenia GM występuje pojęcie estymatora. Co to jest estymator?

Estymator definiujemy następująco: jest to dowolna mierzalna funkcja wektora obserwacji której wartości służą jako oceny nieznanej wartości parametru:

$$\tilde{\boldsymbol{\beta}} = \mathbf{f(y)}$$

Funkcja f musi być funkcją mierzalną (to jest określone formalnie, ale trudne {podobno za trudne}) odwzorowującą przestrzeń T-wymiarową (liczba obserwacji) w k-wymiarową (liczba nieznanych stałych - wartości parametrów)

Ustalmy następujące oznaczenia: $\tilde{\boldsymbol{\beta}}$ to estymator w ogóle (jak powyżej) lub estymator – członek klasy o której akurat mowa; $\hat{\boldsymbol{\beta}}$ to zawsze i tylko estymator MNK dany wzorem jak w twierdzeniu GM; z kolei $\boldsymbol{\beta}$ samo to nieznany parametr który podlega szacowaniu, przy czym $\boldsymbol{\beta}, \tilde{\boldsymbol{\beta}}$ i $\hat{\boldsymbol{\beta}}$ to wszystko wektory (kolumnowe).

Twierdzenie GM mówi, że estymator MNK $\hat{\boldsymbol{\beta}}$ jest najlepszy wśród estymatorów liniowych i nieobciążonych. Co oznaczają te określenia?

Estymator liniowy to estymator który jest liniową funkcją obserwacji, w zapisie macierzowym:

$$\underset{(k\times 1)}{\mathbf{f}(\mathbf{y})} = \underset{(k\times T)}{\mathbf{G}}\ \underset{(T\times 1)}{\mathbf{y}} + \underset{(k\times 1)}{\mathbf{d}}$$

gdzie G i d to znane macierze (?)

Estymator nieobciążony to taki, którego wartość oczekiwana jest równa wartości nieznanego parametru:

$$E(\tilde{\boldsymbol{\beta}}) = E[f(\mathbf{y})] = \boldsymbol{\beta}$$

W ujęciu preeksperymentalnym, **y** to wektor możliwych obserwacji, na tym poziomie jest to wektor losowy : wektor możliwych wyników badania (patrz powyżej). Wobec tego $\tilde{\boldsymbol{\beta}}$ (oraz $\hat{\boldsymbol{\beta}}$) jako funkcja wektora losowego jest również wektorem losowym, co oznacza, że możemy opisywać $\tilde{\boldsymbol{\beta}}$ jako wielowymiarową zmienną losową. Możemy więc badać charakterystyki probabilistyczne estymatora $\tilde{\boldsymbol{\beta}}$, w tym jego wartość oczekiwaną. Używanie do szacowania parametru $\boldsymbol{\beta}$ estymatora którego wartość oczekiwana jest różna od $\boldsymbol{\beta}$ wydaje się intuicyjnie problematyczne, co prowadzi do badania nieobciążoności. Problem efektywności estymatora intuicyjnie odpowiada zagadnieniu: jak daleko przeciętnie jest $\tilde{\boldsymbol{\beta}}$ od $\boldsymbol{\beta}$?. (czyli nasz estymator od prawdziwej nieznanej wartości, czyli badamy np. „czy przeciętnie biorąc nasz estymator dobrze <trafia>"

Przykład: mamy dwa estymatory i przeprowadzamy dwa doświadczenia:
Pierwszy estymator dał w 1 doświadczeniu wartość: 0,9; w 2 doświadczeniu wartość: 1,1
Drugi estymator dał w 1 doświadczeniu wartość: 0,5; w 2 doświadczeniu wartość: 1,5
Oba mają tę samą wartość oczekiwaną (są nieobciążone czyli patrz definicja).
Wolimy pierwszy, bo ma mniejsze rozproszenie i stosując go zyskujemy lepsze (dokładniejsze) oszacowanie nieznanego parametru.

Estymator najlepszy w danej klasie to estymator efektywny, czyli o najmniejszym rozproszeniu (co zostanie dokładnie zdefiniowane poniżej).

1.2.2.1 Dowód Twierdzenia Gaussa i Markowa

Chcemy dowieść twierdzenia GM tj. pokazać, że w klasie estymatorów liniowych i nieobciążonych najlepszy (efektywny) jest estymator MNK $\hat{\boldsymbol{\beta}}$. Sam dowód nie jest trudny, tylko dużo jest kroków przygotowawczych:

Musimy skonstruować klasę, czyli znaleźć ogólną postać estymatora liniowego i nieobciążonego.

Musimy dowieść, że $\hat{\boldsymbol{\beta}}$ jest liniowy i nieobciążony – czyli że należy w ogóle do klasy w której ma być najlepszy.

Potem zapiszemy estymator $\hat{\boldsymbol{\beta}}$ jako członka tej klasy i wykorzystamy dobrze dobrany zapis do wykazania, że jest on najlepszy.

Zapis klasy estymatorów:

$\tilde{\mathbf{B}}$ - rozważana klasa estymatorów parametru wektorowego $\boldsymbol{\beta}$; $\tilde{\boldsymbol{\beta}} \in \tilde{\mathbf{B}}$ to element tej klasy.

Zapis efektywności w klasie:

Definicja: $\hat{\boldsymbol{\beta}}$ najlepszy w klasie $\tilde{\mathbf{B}}$ wtedy i tylko wtedy, gdy:

$$\underset{\tilde{\boldsymbol{\beta}}\in\tilde{\mathbf{B}}}{\forall}\ A = V(\tilde{\boldsymbol{\beta}}) - V(\hat{\boldsymbol{\beta}}) \text{ jest macierzą określoną nieujemnie}$$

Macierz **A** to różnica pomiędzy macierzami kowariancji odpowiednio $\tilde{\boldsymbol{\beta}}$ (każdego estymatora rozważanej klasy, w wypadku twierdzenia GM – klasy estymatorów liniowych i nieobciążonych) a $\hat{\boldsymbol{\beta}}$ (estymatora-kandydata do najlepszości, w naszym wypadku estymatora MNK)

Co to jest macierz określona nieujemnie? Jeśli macierz **A** jest symetryczna (więc i kwadratowa; dla takich macierzy definiuje się określoność) stopnia k, to:

$$\text{A jest określona nieujemnie} \Leftrightarrow \underset{\mathbf{z}\in\Re^k\setminus\{\mathbf{0}\}}{\forall} \mathbf{z'Az} \geq 0$$

W szczególności nieujemnie określona macierz **A** ma wszystkie przekątniowe elementy większe lub równe zero. Na przekątnej macierzy kowariancji są wariancje, czyli definicja efektywności mówi że:

$\mathbf{var}(\tilde{\beta}_i) \geq \mathbf{var}(\hat{\beta}_i)$ dla $i = 1, \ldots, k$ (β_i oznacza i-ty element wektora $\boldsymbol{\beta}$)

Oznacza to, że estymator jest najlepszy w danej klasie jeśli po odjęciu jego wariancji od wariancji każdego innego członka klasy zostaje zero lub więcej. Dla najlepszego estymatora $\hat{\boldsymbol{\beta}}$ wariancja każdego elementu wektora ($\mathbf{var}(\hat{\beta}_i)$) jest mniejsza (lub równa) od odpowiednich wariancji wszystkich innych estymatorów danej klasy ($\mathbf{var}(\tilde{\beta}_i)$). Istnienie najlepszego estymatora jest nietrywialne i w szerokich klasach może go nie być wcale.

Przystąpmy do realizacji następującego planu: dowiedziemy liniowości i nieobciążoności $\hat{\boldsymbol{\beta}}$, a następnie pokażemy, że $\hat{\boldsymbol{\beta}}$ jest najlepszy w klasie $\tilde{\mathbf{B}}$ estymatorów liniowych i nieobciążonych, czyli będziemy musieli dowieść, że macierz będąca różnicą macierzy kowariancji $\tilde{\boldsymbol{\beta}}$ i $\hat{\boldsymbol{\beta}}$ jest nieujemnie określona, wszystko to przy prawdziwości założeń KMRL (1° – 5°):

1° $\underset{(T\times 1)}{\mathbf{y}} = \underset{(T\times k)}{\mathbf{X}}\,\underset{(k\times 1)}{\boldsymbol{\beta}} + \underset{(T\times 1)}{\boldsymbol{\varepsilon}}$

2° **X** jest znaną macierzą nielosową

3° $rz(\mathbf{X}) = k$

4° $E(\boldsymbol{\varepsilon}) = \underset{(T\times 1)}{\mathbf{0}}$

5° $V(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I_T}$ $\qquad \sigma^2>0$

Chcemy dowieść że:

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

jest efektywny w klasie estymatorów liniowych i nieobciążonych, i ma to wynikać z 1°-5°

W tezie (czyli w powyższym wzorze) wykorzystano już założenie 3, bo ono gwarantuje że $(\mathbf{X'X})^{-1}$ istnieje i można tak zapisać. (3°)

Przypomnijmy definicję estymatora liniowego:

$$\underset{(k\times 1)}{\mathbf{f}(\mathbf{y})} = \underset{(k\times T)}{\mathbf{G}}\,\underset{(T\times 1)}{\mathbf{y}} + \underset{(k\times 1)}{\mathbf{d}};$$

Możemy zapisać, że:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'y} = \mathbf{Cy} \text{ ;gdzie :}$$

$\underset{(k\times T)}{\mathbf{C}} = \mathbf{(X'X)^{-1}X'}$ (korzystamy z założenia 2, że **X** jest znaną macierzą nielosową)($2°$)

Macierz **C** spełnia warunki macierzy **G**, a za wektor **d** przyjmujemy wektor zerowy, wobec czego $\hat{\boldsymbol{\beta}}$ spełnia warunki estymatora liniowego.

Zajmijmy się nieobciążonością: chcemy dowieść że $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ czyli że wartość oczekiwana estymatora MNK $\hat{\boldsymbol{\beta}}$ to wartość nieznanego parametru wektorowego $\boldsymbol{\beta}$.

$$E(\hat{\boldsymbol{\beta}}) \underset{}{=} E(\mathbf{Cy}) \overset{2°}{=} \mathbf{C^*}\, E(\mathbf{y}) \overset{1°}{=} \mathbf{C^*}\, E(\mathbf{X\beta+\varepsilon}) = \mathbf{C}[E(\mathbf{X\beta}) + E(\boldsymbol{\varepsilon})] \overset{4°}{=} \mathbf{C^*}\, E(\mathbf{X\beta}) = \mathbf{CX\beta} = \mathbf{I}_k\boldsymbol{\beta} = \boldsymbol{\beta}$$

Kolejne przejścia mają następujące uzasadnienie:

Pierwsza równość jest z definicji i z zapisu $\hat{\boldsymbol{\beta}}$ =**Cy**.

Druga równość to wyciągnięcie stałej przed wartość oczekiwaną :($2°$) **X** jest znany więc $\underset{(k\times T)}{\mathbf{C}} = \mathbf{(X'X)^{-1}X'}$ jest stałą i można ją wyprowadzić przed operator wartości oczekiwanej.

Trzecia równość jest z założenia $1°$ - podstawiamy za y.

Czwarta równość wynika z tego, że wartość oczekiwana sumy to suma wartości oczekiwanych. (suma musi być skończona, ale tu jest)

Piąta równość to z $4°$ E(**ε**)=**0**.

Szósta bo wartość oczekiwana stałej to stała, **X** jest znany, a **β** nieznany, ale też stały.

Siódma równość wynika z tego, że **CX** = I tożsamościowo (**X**'**X**)$^{-1}$(**X**,**X**) to macierz razy swoja odwrotność, macierz jednostkowa to element neutralny w mnożeniu macierzy; CBDO.

Uwaga: dowód nieobciążoności $\hat{\boldsymbol{\beta}}$ nie wykorzystuje założenia $5°$, estymator MNK jest nieobciążony niezależnie od postaci macierzy kowariancji składników losowych.

Przystąpmy do wyprowadzenia zapisu klasy estymatorów liniowych nieobciążonych:

Element tej klasy to $\tilde{\boldsymbol{\beta}} = \mathbf{Gy} + \mathbf{d}$, ale chcemy to zapisać jako $\hat{\boldsymbol{\beta}}$ plus pewna poprawka.
Bierzemy macierz F taką, że **F = G – C**
Czyli **G = F + C**

$$\tilde{\boldsymbol{\beta}} = \mathbf{Gy} + \mathbf{d} = \mathbf{(F + C)y} + \mathbf{d} = \hat{\boldsymbol{\beta}} + \mathbf{Fy} + \mathbf{d}$$

czyli zapisaliśmy estymator liniowy jako $\hat{\boldsymbol{\beta}} + \mathbf{Fy} + \mathbf{d}$.

Teraz szukamy warunków na **F** i **d** żeby był on nieobciążony:

$$E(\tilde{\boldsymbol{\beta}}) = E(\hat{\boldsymbol{\beta}}) + E(\mathbf{Fy}) + E(\mathbf{d}) = \boldsymbol{\beta} + E(\mathbf{Fy}) + E(\mathbf{d}) = \boldsymbol{\beta} + \mathbf{F}^*\, E(\mathbf{y}) + \mathbf{d}$$

$$= \boldsymbol{\beta} + \mathbf{F}E(\mathbf{X\beta} + \boldsymbol{\varepsilon}) + \mathbf{d} = \boldsymbol{\beta} + \mathbf{FX\beta} + \mathbf{d}$$

F i **d** są stałe, skorzystano z dowodu nieobciążoności $\hat{\boldsymbol{\beta}}$ w drugim kroku, reszta wnioskowania jak w poprzednio.

Wniosek:

Żeby $\tilde{\boldsymbol{\beta}}$ był nieobciążony **FXβ + d** musi być równe **0** (wymagamy, by $E(\tilde{\boldsymbol{\beta}}) = \boldsymbol{\beta}$), by $\tilde{\boldsymbol{\beta}}$ był nieobciążony niezależnie od konkretnej wartości **X**, trzeba by **FX = 0** i **d = 0**

Wobec tego klasę estymatorów liniowych i nieobciążonych zapiszemy:

$$\tilde{\mathrm{B}} = \left\{ \tilde{\beta} : \tilde{\beta} = \hat{\beta} + \mathrm{Fy} \wedge \mathrm{FX} = \underset{(k\times k)}{0} \right\}$$

Żeby dowieść nieobciążoności potrzebujemy macierzy $\mathbf{V}(\tilde{\boldsymbol{\beta}})$. Ogólnie macierz kowariancji wektorowej zmiennej losowej to wartość oczekiwana iloczynu wektora kolumny odchyleń (wartość zmiennej minus jej wartość oczekiwana) przez siebie po transpozycji:

$$\mathbf{V}(\tilde{\boldsymbol{\beta}}) = \mathbf{E}[(\tilde{\boldsymbol{\beta}} - \mathbf{E}(\tilde{\boldsymbol{\beta}}))(\tilde{\boldsymbol{\beta}} - \mathbf{E}(\tilde{\boldsymbol{\beta}}))']$$

czyli po podstawieniu $\mathbf{E}(\tilde{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ z definicji nieobciążoności i $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} + \mathbf{Fy}$ z konstrukcji klasy:

$$\mathbf{V}(\tilde{\boldsymbol{\beta}}) = \mathbf{E}[(\hat{\boldsymbol{\beta}} + \mathbf{Fy} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} + \mathbf{Fy} - \boldsymbol{\beta})']$$

a wektor tworzący tą macierz to:

$$\hat{\boldsymbol{\beta}} + \mathbf{Fy} - \boldsymbol{\beta} = \mathbf{Cy} + \mathbf{Fy} - \boldsymbol{\beta} = \mathbf{C}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) + \mathbf{F}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) - \boldsymbol{\beta} = \mathbf{CX}\boldsymbol{\beta} + \mathbf{C}\boldsymbol{\varepsilon} + \mathbf{FX}\boldsymbol{\beta} + \mathbf{F}\boldsymbol{\varepsilon} - \boldsymbol{\beta} = (\mathbf{C} + \mathbf{F})\boldsymbol{\varepsilon}$$

(bo $\mathbf{CX}=\mathbf{I}$ z czego już korzystaliśmy, a $\mathbf{FX}=\mathbf{0}$ z konstrukcji klasy)

zapiszmy:

$$\mathbf{V}(\tilde{\boldsymbol{\beta}}) = \mathrm{E}[((\mathbf{C} + \mathbf{F})\boldsymbol{\varepsilon})((\mathbf{C} + \mathbf{F})\boldsymbol{\varepsilon})'] = \mathrm{E}[(\mathbf{C} + \mathbf{F})\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'(\mathbf{C} + \mathbf{F})'] = (\mathbf{C} + \mathbf{F})\mathbf{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')(\mathbf{C} + \mathbf{F})' =$$
$$= (\mathbf{C} + \mathbf{F})\mathbf{V}(\boldsymbol{\varepsilon})(\mathbf{C}' + \mathbf{F}') = (\mathbf{C} + \mathbf{F})(\sigma^2\mathbf{I}_T)(\mathbf{C}' + \mathbf{F}') = \sigma^2(\mathbf{C} + \mathbf{F})(\mathbf{C}' + \mathbf{F}') = \sigma^2(\mathbf{C}'\mathbf{C} + \mathbf{CF}' + \mathbf{FC}' + \mathbf{FF}')$$

Tu drugi krok wynika z własności transpozycji iloczynu macierzy (**AB**)'=**B**'**A**', następnie wyciąga się stałe poza wartość oczekiwaną a $\mathbf{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')$to jest definicja $\mathbf{V}(\boldsymbol{\varepsilon})$(bo V(**ε**)=E[(**ε**-E(**ε**))(**ε**-E(**ε**))'] a E(**ε**)=**0** na mocy 4°), która na mocy założenia 5° ma postać $\mathrm{V}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}_T$.

Wartości poszczególnych składników sumy w ostatnim nawiasie to:
CC'=(**X**'**X**)$^{-1}$**X**'**X**(**X**'**X**)$^{-1}$=(**X**'**X**)$^{-1}$ {z własności transpozycji iloczynu i ponieważ macierz symetryczna jest równa swej transpozycji}
CF'=(**X**'**X**)$^{-1}$**X**'**F**'=(**X**'**X**)$^{-1}$(**FX**)'=(**X**'**X**)$^{-1}$***0=0** (j.w. i z własności klasy: **FX**=0)
FC'=(**CF**')'=**0**

Ostatecznie:

$$\mathbf{V}(\tilde{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} + \sigma^2\mathbf{FF}'$$

Po podstawieniu macierz $\mathbf{A}=\sigma^2\mathbf{FF}'$,czyli jest nieujemnie określona na mocy konstrukcji.

Dla $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}$ macierz **F** jest macierzą zawierającą wyłącznie zera (patrz konstrukcja macierzy **F**).

(na wykładzie 2001/2002 czyli obecnym była podobno inna końcówka dowodu, co warto by uzupełnić.)

Z powyższego wzoru widać, że macierz kowariancji dowolnego estymatora liniowego i nieobciążonego to macierz kowariancji $\hat{\boldsymbol{\beta}}$ plus pewna macierz mająca elementy większe lub równe zeru na przekątnych, czyli żaden estymator liniowy i nieobciążony nie może mieć mniejszej wariancji od $\hat{\boldsymbol{\beta}}$ czyli QED.

1.2.3 Twierdzenie o wariancji resztowej w KMRL

W KMRL z T>k (T=k odrzucamy) nieobciążonym estymatorem parametru σ^2 jest tzw. wariancja resztowa dana wzorem:

$s^2 = \frac{1}{T-k}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})$, i $E(s^2) = \sigma^2$

Estymatorem (nieobciążonym) macierzy kowariancji estymatora MNK $\mathbf{V}(\hat{\boldsymbol{\beta}})$ jest:

$$\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}) = s^2(\mathbf{X}'\mathbf{X})^{-1}, \qquad E\left(\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}})\right) = \mathbf{V}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

{tu drobna uwaga: rozróżniajmy dokładnie V z daszkiem od V bez daszka, jeśli ktoś dokładnie wie, to niech opuści, ale jak ktoś nie wie, to niech przeczyta, bo inaczej dalej będzie miał mętlik że szkoda gadać. Otóż estymator MNK $\hat{\boldsymbol{\beta}}$ traktujemy tu jako wektor losowy zgodnie z uzasadnieniem danym wyżej, wobec czego ma on macierz kowariancji $\underset{(k\times k)}{\mathbf{V}}(\hat{\boldsymbol{\beta}})$. To jest jego charakterystyka, która istnieje i którą chcemy szacować. Dana jest wzorem jak w twierdzeniu GM i nie znamy jej, bo nie znamy σ^2, kiedy jednak za σ^2 podstawimy oszacowanie s^2, według powyższego twierdzenia, dostaniemy (estymator/oszacowanie $\underset{(k\times k)}{V}(\hat{\beta})$) czyli ($\underset{(k\times k)}{\hat{\mathbf{V}}}(\hat{\boldsymbol{\beta}})$); to ostatnie V to estymator macierzy kowariancji (lub jego realizacja, czyli oszacowanie tej macierzy) a to poprzednie V to macierz kowariancji estymatora MNK $\hat{\boldsymbol{\beta}}$. Przy czym oczywiście jest to zupełnie co innego niż $\underset{(T\times T)}{\mathbf{V}}(\boldsymbol{\varepsilon})$. }

Daszków ciąg dalszy:

$\underset{(T\times 1)}{\hat{\mathbf{y}}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ to wektor „wartości teoretycznych” zmiennej y przewidywanych przez oszacowany model;

$\underset{(T\times 1)}{\hat{\boldsymbol{\varepsilon}}} = \mathbf{y} - \hat{\mathbf{y}}$ to wektor „reszt MNK”.

Jeśli podstawimy dwa poprzednie wzory do wzoru na wariancję resztową otrzymamy:

$$s^2 = \frac{1}{T-k}\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} = \frac{\sum_{t=1}^{T}\hat{\varepsilon}_t^2}{T-k}$$

a to jest wariancja próbkowa reszt MNK.

Wzór obliczeniowy na s^2 jest natomiast następujący:

$$s^2 = \frac{1}{T-k}(\mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\hat{\boldsymbol{\beta}})$$

Przy zastosowaniu zbyt poważnych zaokrągleń można otrzymać tu wartość ujemną, co jest oczywistym komunikatem błędu.

1.3 Klasyczny Model Normalnej Regresji Liniowej (KMNRL)

Model KMNRL to modyfikacja KMRL: przyjmujemy dodatkowe założenie dotyczące wektora $\boldsymbol{\varepsilon}$ nieobserwowalnych składników losowych. Przejście z KMRL do KMNRL to wzmocnienie założeń. A przy okazji: co to są mocne a co to słabe założenia? I czy to „dobrze", że założenie jest mocne/słabe? Otóż mocne założenie to takie które nakłada dodatkowe wymagania, warunki (mówiąc potocznie) na modelowany obiekt. Dzięki nałożeniu dodatkowych warunków, czyli wzmocnieniu założeń uzyskujemy też wzmocnienie tezy – możemy dowieść „więcej". Czyli mocne założenia są lepsze bo w mocnym układzie założeń możemy więcej zdziałać (mając do dyspozycji mocniejszą tezę). Ale stosując model przyjmujemy jego założenia bez możliwości sprawdzenia (w ramach tego modelu). Zakładamy, że nasz badany obiekt posiada wymienione właściwości. Tutaj założenia mocne są restrykcyjne, bo przecież dużo wymagają, a założenia słabe są mało wymagające, ogólne. Z przyjęcia założeń powinniśmy się wytłumaczyć, założenia słabe łatwiej przyjąć – bo niewiele wymagają, założenia mocne są kontrowersyjne. Czyli założenia słabe są lepsze, bo są bardziej ogólne. Mamy tu przypadek „trade – off": musimy przyjąć założenia na tyle mocne, by dało się coś osiągnąć, i na tyle słabe, by dało się rozsądnie przypuszczać (lub formalnie testować w szerszym modelu), że rzeczywiście obejmują badany obiekt.

Będziemy się więc zajmować wektorem $\boldsymbol{\varepsilon}$ nieobserwowalnych składników losowych : „reprezentującym łączny wpływ wszystkich czynników drugorzędnych, przypadkowych, nie uwzględnionych...", czyli ε_t, składnik losowy obserwacji o numerze t jest skończoną sumą m indywidualnych zakłóceń losowych:

$$\varepsilon_t = a_{t1} + a_{t2} + a_{t3} + ... + a_{tm}$$

Jeśli założymy, że poszczególne losowe zakłócenia a_{tj} mają identyczne niezależne rozkłady o zerowej wartości oczekiwanej i wariancji σ^2_a, co zapisujemy:

$$a_{tj} \sim {}_{ii}D(0, \sigma^2_a)$$

to na podstawie CTG (Centranych Twierdzeń Granicznych) możemy dowieść, że ich suma w przybliżeniu ma rozkład normalny o zerowej wartości oczekiwanej i wariancji równej $m\sigma^2_a$:

$$\varepsilon_t = \sum_{j=1}^{m} a_{tj} \dot{\sim} N(0, m\sigma^2_a)$$

czyli mamy podstawy przyjąć, że wektor $\boldsymbol{\varepsilon}$ ma wielowymiarowy rozkład normalny.

Tylda ~ oznacza: „ma rozkład", tylda z kropką: „w przybliżeniu ma rozkład", iiD to identically independently distributed czyli „mają identyczne niezależne rozkłady"; iiN(,) czytamy: mają (identyczne) niezależne rozkłady normalne o wartości oczekiwanej i wariancji – podane w nawiasie, czyli mamy pełną charakterystykę normalnej zmiennej losowej, bo znamy wartości oczekiwane (pierwsze w nawiasie), wariancje (drugie w nawiasie) i dla rozkładu normalnego kowariancje (zerowe, bo independently – niezależnie).

σ^2_a ma subskrypt „a" żeby się nie myliło z wariancją $\boldsymbol{\varepsilon}$ równą σ^2, a wariancja $\boldsymbol{\varepsilon}$ w powyższym wzorze równa jest $m\sigma^2_a$, bo zmienne są niezależne i nie ma co odejmować (gdyby były zależne to odejmuje się odpowiednie iloczyny kowariancji, które tu się zerują).}

Pokazaliśmy powyżej, że z rozumienia $\boldsymbol{\varepsilon}$ jako „sumy indywidualnych zakłóceń losowych", przy przyjęciu jednakowego (nieznanego) rozkładu, jednakowej wariancji i zerowej wartości oczekiwanej składników tej sumy – czyli owych „indywidualnych zakłóceń" wynika, że $\boldsymbol{\varepsilon}$ ma w przybliżeniu rozkład normalny. Wobec tego obok założeń 1° - 5° przyjmiemy dodatkowe założenie 6° mówiące, że $\boldsymbol{\varepsilon}$ ma wielowymiarowy rozkład normalny. Uzyskamy w ten sposób model KMNRL, o czym dalej.

Trochę komentarza do założenia 6° i jego uzasadnienia.

1. Przedstawione uzasadnienie w postaci implikacji mogłoby być rozszerzone. Intuicyjnie mówiąc są warianty CTG wspierające tezę o normalności powyższej sumy także w przypadku gdy jej składniki nie są niezależne i mają różną wariancję. Nasze „indywidualne zakłócenia" mogą więc być zależne, mogą też mieć różną wariancję (byle żadna z cząstkowych wariancji nie dominowała nad wariancją sumy), i nawet wtedy wspierać założenie o wielowymiarowym rozkładzie normalnym dla $\boldsymbol{\varepsilon}$. To argumenty mogące skłaniać do przyjęcia założenia 6°.

2. Zauważmy jednak, że w podanym uzasadnieniu występuje nie rzucające się w oczy założenie, że rozkłady owych indywidualnych zakłóceń mają wariancję i wartość oczekiwaną. Co z tego wynika? Otóż są rozkłady które nie mają tych momentów (wartość oczekiwana i wariancja to momenty). Ulubionym kontrprzykładem Profesora jest rozkład Cauchy'ego. (zmienna mająca rozkład t-Studenta o 1 stopniu swobody ma rozkład Cauchy'ego, w rozkładzie t-Studenta o n stopniach swobody istnieją momenty rzędu niższego niż n.). Jak technicznie wygląda „niemanie" momentów? Taki moment to odpowiednia całka, która może nie istnieć, jeśli rozkład ma tzw. grube ogony, czyli wolno maleje do zera w +- ∞. Grube ogony (heavy tails) to cecha rozkładu mówiąca o prawdopodobieństwie wystąpienia realizacji daleko odbiegającej od tendencji centralnej. Czyli jeśli obserwacje „nietypowe" zdarzają się często, to przypuszczamy, że rozkład charakteryzujący losowe zakłócenia ma grube ogony jak np. rozkład Cauchy'ego, czyli nie spełnia naszych założeń, i nie ma podstaw aby przyjąć, że suma takich zakłóceń miała rozkład normalny. Rozkład Cauchy'ego jest α-stabilny, czyli suma zmiennych o tym rozkładzie jest również zmienną Cauchy'ego, a nie normalną. W takich okolicznościach założenie 6° okazuje się założeniem „zbyt mocnym", nie dającym się utrzymać, musimy je odrzucić i w związku z tym tracimy wszelkie korzyści wynikające z jego przyjęcia (korzyści zostaną opisane poniżej). Jest to przypadek np. danych finansowych wysokiej częstotliwości, gdzie „obserwacje nietypowe są typowe", nie są spełnione założenia CTG wobec czego stosowanie technik KMNRL nie ma uzasadnienia i musimy odwołać się do sposobów wnioskowania zakładających możliwość występowania zakłóceń „gruboogoniastych". Po tym przydługim nieco wstępie przejdźmy wreszcie do modelu KMNRL.

1.3.1 Założenia Klasycznego Modelu Normalnej Regresji Liniowej

1°	$\underset{(T\times 1)}{\mathbf{y}} = \underset{(T\times k)}{\mathbf{X}}\,\underset{(k\times 1)}{\boldsymbol{\beta}} + \underset{(T\times 1)}{\boldsymbol{\varepsilon}}$
2°	$\mathbf{X}$ jest znaną macierzą nielosową
3°	$rz(\mathbf{X}) = k$
4°	$E(\boldsymbol{\varepsilon}) = \underset{(T\times 1)}{\mathbf{0}}$
5°	$V(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I_T}$ $\qquad \sigma^2>0$
6°	$\boldsymbol{\varepsilon}$ ma T-wymiarowy rozkład normalny

założenia 4°-6° zgodnie z podaną wcześniej i nieco rozszerzoną notacją możemy zapisać następująco:

$$\boldsymbol{\varepsilon} \sim N^T(\mathbf{0}, \sigma^2 \mathbf{I})$$

co czytamy: epsilon ma T-wymiarowy łączny rozkład normalny o wartości oczekiwanej będącej wektorem zerowym i macierzy kowariancji danej wzorem $\sigma^2 \mathbf{I}$. W stosunku do poprzednio wprowadzonego zapisu to jest uogólnienie na przypadek wielowymiarowy, czyli w nawiasie podajemy wektor wartości oczekiwanych (oznaczany $\boldsymbol{\mu}$) i macierz kowariancji, najczęściej oznaczaną jako $\boldsymbol{\Sigma}$. Ogólnie:

$$\mathbf{z} \sim N^T(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

A co oznacza: „ma łączny wielowymiarowy rozkład normalny”?

{dygresja: w ekonometrii nie ma słów „niewinnych”. Najczęściej jest tak, że każde słowo na swoim miejscu coś znaczy, i zastąpienie go podobnym lub pominięcie odwraca cały sens. Powoduje to przykre niespodzianki na egzaminie przy punktowaniu interpretacji. Jeśli coś jest jakoś napisane, to znaczy, że ma być dokładnie tak, i warto się zastanowić dlaczego. Albo, że Profesor się pomylił {tu jeszcze ja się mogę pomylić BM}, ale to nie zmienia reguły, bo Profesor zdecydowaną większość czasu mówi do rzeczy i tylko czasem się myli.}

{dygresja 2. Pytanie „co oznacza?” czytaj zazwyczaj: „Jak jest zdefiniowany?”}

Definicja:

Wielowymiarowa zmienna losowa $\mathbf{z}$ (będąca wektorem losowym) ma T-wymiarowy rozkład normalny wtedy i tylko wtedy, gdy wszystkie niezerowe kombinacje liniowe jej elementów mają jednowymiarowy rozkład normalny.

$\underset{(T\times 1)}{\mathbf{z}}$ ma wielowymiarowy rozkład normalny ⇔

$\underset{\mathbf{c}\in\Re^T\setminus\{\mathbf{0}\}}{\forall} \mathbf{c}'\mathbf{z} = \sum_{t=1}^{T} c_t z_t$ ma 1-wymiarowy rozkład normalny.

W szczególności z tej definicji wynika, że wszystkie składowe mają rozkład normalny, jednak nie odwrotnie, tzn. z normalności wszystkich składowych nie wynika wielowymiarowa normalność.

Funkcja gęstości prawdopodobieństwa (p.d.f. : probability density function – nie mylić z *.pdf) T-wymiarowego nieosobliwego rozkładu normalnego z parametrami $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$:

$$p(\mathbf{z};\boldsymbol{\mu},\boldsymbol{\Sigma}) = f_N^T(\mathbf{z} \mid \boldsymbol{\mu},\boldsymbol{\Sigma}) = (2\pi)^{(-T/2)}(\det \boldsymbol{\Sigma})^{-1/2}\exp\left[-1/2(\mathbf{z}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu})\right]$$

gdzie (bezpośrednia interpretacja probabilistyczna):

$$V(\mathbf{z}) = \boldsymbol{\Sigma} \text{ i } E(\mathbf{z}) = \boldsymbol{\mu}$$

$\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ to odpowiednio (T×1) wektorowa wartość oczekiwana i (T×T) macierz kowariancji (symetryczna i dodatnio określona, dająca się odwracać, dzielić, pierwiastkować na przekątnej)

Co oznacza rozkład „nieosobliwy”? (ograniczamy się do rozkładów mających momenty 2-go rzędu). Rozkład normalny osobliwy ma osobliwą macierz $\boldsymbol{\Sigma}$, czyli nie da się zapisać powyższej funkcji gęstości, bo nie da się odwrócić macierzy $\boldsymbol{\Sigma}$. W rozkładzie nieosobliwym da się odwrócić tą macierz, czyli można zapisać funkcję gęstości {*trywializując*, jak mawia Profesor}.

W rozkładzie nieosobliwym każda badana zmienna wnosi swoistą zmienność do układu, w rozkładzie osobliwym pewne zmienne są funkcyjnie zależne od innych, nie wnoszą zmienności. Z rozkładami osobliwymi mamy do czynienia przy modelowaniu właśnie związków funkcyjnych między zmiennymi tworzącymi wektor.

{Wzór w powyższej ramce (funkcja gęstości wielowymiarowego rozkładu normalnego) jest jednym z najważniejszych wzorów tego kursu, trzeba go znać na pamięć bez cienia błędu. Co oznacza zapis $p(\mathbf{z} ; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = ...$? a konkretnie co w tym nawiasie robi średnik? Ogólnie w nawiasie są argumenty, ale po średniku są parametry które chcemy traktować jako ustalone – (wektor wartości oczekiwanej i macierz kowariancji) a przed średnikiem jest wektor wartości zmiennej losowej $\mathbf{z}$, który jest podstawowym argumentem funkcji gęstości. Zapisujemy w ten sposób, bo interesuje nas wartość funkcji p dla dowolnego wektora $\mathbf{z}$ przy ustalonych wartościach $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, co jeszcze bardziej widać w zapisie z f, tam pionowa kreska oznacza: warunkowo względem ustalonej wartości tego, co po kresce. Przy tym zapisie jest też ewidentne, że interesuje nas wielowymiarowa zmienna losowa z dowolną (niekoniecznie diagonalną) macierzą kowariancji, czyli mamy wektor zmiennych które mogą być zależne. Gdyby macierz kowariancji była diagonalna to mielibyśmy iloczyn T jednowymiarowych funkcji gęstości, a gdyby jeszcze wszystkie jej przekątniowe elementy były równe np. σ^2, to byłaby T-ta potęga jednowymiarowej funkcji. Ponieważ na statystyce matematycznej był chyba tylko przypadek jednowymiarowy a tu jest cały wzór macierzowy to warto do niego podstawić różne (np. powyższe) postacie macierzy $\boldsymbol{\Sigma}$ i spróbować dojść do odpowiedniej modyfikacji przypadku jednowymiarowego. }

Teraz wreszcie te obiecane a wielkie korzyści przyjęcia 6-go założenia. Są one dwojakiego rodzaju: estymator MNK nabiera dodatkowych (w porównaniu z KMRL) własności teoretycznych, oraz dostępne się stają dodatkowe możliwości wnioskowania statystycznego: estymacja przedziałowa i testowanie hipotez.

1.3.2 Własności estymatora MNK w KMNRL

W KMNRL (przy założeniach 1°-6°) estymator MNK $\hat{\boldsymbol{\beta}}$ jest najlepszy w klasie wszystkich estymatorów nieobciążonych.

Przy czym jest to teoretyczna własność stosowanej procedury (estymatora MNK) która nie ma zazwyczaj zbyt wielkiego wpływu na nasz sposób postępowania. {Miałoby ono być może techniczne znaczenie, gdybyśmy mieli „na widoku” jakiś inny estymator.}

Kosztem jest tu przyjęcie mocnego założenia o normalności $\boldsymbol{\varepsilon}$, a rezultat powyższy jest warunkowany prawdziwością tego założenia.

Z praktycznego punktu widzenia dużo ważniejsza jest kolejna własność:

W KMNRL zachodzą twierdzenia o rozkładach związanych z normalnym, które pozwalają nam na budowanie przedziałów ufności i testowanie hipotez o współczynnikach regresji, czyli mamy do dyspozycji nowe możliwości wnioskowania statystycznego.

1.3.3 Estymacja przedziałowa i testowanie hipotez

1.3.3.1 Wnioskowanie o pojedynczym parametrze regresji.

Wnioski z twierdzeń o rozkładach związanych z normalnym:

a) $$\frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{\sigma^2}=\frac{(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})}{\sigma^2}=\frac{(T-k)*s^2}{\sigma^2}\sim\chi^2(T-k)$$

b) $$\frac{\hat{\beta}_i-\beta_i}{D(\hat{\beta}_i)}\sim St(T-k)$$

(oczywiście χ^2 oznacza rozkład chi-kwadrat a St rozkład t-Studenta, w nawiasach stopnie swobody)

W powyższych punktach mamy podane rozkłady zmiennych losowych – tzw. statystyk (statistic nie mylić ze statistics która jest nauką) które są znanymi funkcjami z jednym nieznanym parametrem. Możemy je tak przekształcić, żeby móc testować hipotezy i budować przedziały ufności dotyczące owego nieznanego parametru. {jak na statystyce matematycznej}.

Korzystając ze statystyki a) moglibyśmy budować przedziały ufności i testować hipotezy dotyczące nieznanego parametru struktury stochastycznej σ^2. Moglibyśmy, ale nie będziemy (chyba, że Profesor zmienił zdanie).

Wykorzystamy jednak statystykę daną w punkcie b). W tym celu przyjrzyjmy się jej bliżej. W liczniku jest $\hat{\beta}_i-\beta_i$, czyli „prawdziwy błąd estymacji" (różnica między oszacowaniem MNK a prawdziwą nieznaną wartością parametru). W mianowniku jest $D(\hat{\beta}_i)$, czyli błąd średni szacunku MNK, to jest *ocena* błędu estymacji (czyli tego, co w liczniku, a czego oczywiście nie znamy). A skąd wziąć błąd średni szacunku? Z oszacowanej macierzy kowariancji estymatora MNK. Pierwiastki kwadratowe elementów przekątniowych tej macierzy to błędy średnie szacunku odpowiednich parametrów

Pamiętamy: $\mathbf{V}(\hat{\boldsymbol{\beta}})=\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ natomiast $\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}})=s^2(\mathbf{X}'\mathbf{X})^{-1}$ i

$$\underset{(T\times T)}{\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}})}\overset{df}{=}\begin{bmatrix}\hat{var}(\hat{\beta}_1) & \hat{cov}(\hat{\beta}_1,\hat{\beta}_2) & \dots & \hat{cov}(\beta_1,\beta_k)\\ \hat{cov}(\hat{\beta}_2,\hat{\beta}_1) & \hat{var}(\hat{\beta}_2) & \dots & \dots\\ \dots & \dots & \dots & \dots\\ \hat{cov}(\hat{\beta}_k,\hat{\beta}_1) & \dots & \dots & \hat{var}(\hat{\beta}_k)\end{bmatrix}$$

następnie $D(\hat{\beta}_i)=\sqrt{\hat{var}(\hat{\beta}_i)}$.

Błąd średni szacunku i-tego parametru regresji $D(\hat{\beta}_i)$ mówi nam, o ile średnio mylimy się zastępując nieznaną wartość parametru jego oszacowaniem MNK.

1.3.3.1.1 Przedziały ufności dla pojedynczego parametru regresji

Przedziały ufności są narzędziami wnioskowania statystycznego, estymacja przedziałowa daje bogatszą informację dotyczącą szacowanego parametru niż estymacja punktowa. Szczególną wagę mają tu jednak zagadnienia związane z prawidłową interpretacją. Rozumienie ich jest warunkiem odniesienia korzyści z dodatkowych możliwości wnioskowania właściwych KMNRL.

W wypadku przedziału ufności zasadniczą wagę ma rozróżnienie pomiędzy zmienną losową a jej pojedynczą realizacją. Wymienione powyżej założenia i statystyka przedstawiona w punkcie b) pozwalają na wyprowadzenie własności teoretycznych przedziału ufności rozumianego jako przedział losowy (czyli przedział o końcach losowych). Jednak w badaniach czy w zadaniu będziemy mieli do czynienia z *pojedynczą realizacją* przedziału losowego i to ona podlegać będzie interpretacji. O rozróżnieniu przedziału ufności od realizacji przedziału ufności należy bezwzględnie pamiętać.

1. Przedział ufności jest najkrótszym przedziałem o końcach losowych który z określonym prawdopodobieństwem zawiera nieznaną wartość parametru. Prawdopodobieństwo to określamy z góry i nazywamy poziomem ufności. Poziom ufności oznacza się $(1 - \alpha)$, czyli 95%-wemu przedziałowi ufności odpowiada $\alpha = 5\% = 0{,}05$. Przedział ufności jest przedziałem o końcach losowych dlatego, że końce przedziału są znanymi funkcjami wektora losowego obserwacji **y**, a wektor obserwacji **y** jest zdefiniowany jako wektor losowy (patrz początek). Przedział ufności jest skonstruowany tak, żeby był najkrótszym (o tym dalej) przedziałem zawierającym nieznaną wartość parametru z określonym prawdopodobieństwem. Oznacza to (na podstawie częstościowej interpretacji prawdopodobieństwa na gruncie której się znajdujemy), że gdybyśmy obserwowali realizacje wektora **y** odpowiednio wiele razy i odpowiednio wiele razy wyliczali realizacje przedziału ufności, to przeciętnie w $(1 - \alpha)*100$ na 100 przypadków nasza realizacja przedziału ufności zawierałaby nieznany parametr. Na gruncie ekonomii mamy jednak do czynienia z danymi historycznymi: pojedynczą realizacją danych. Na tej podstawie obliczamy:

2. pojedynczą realizację przedziału ufności. *Pojedyncza realizacja* przedziału ufności to przedział którego końce są znanymi liczbami, są znanymi funkcjami *realizacji* wektora obserwacji, czyli ustalonych liczb. Ten przedział (już nie „o końcach losowych") albo zawiera nieznaną wartość szacowanego parametru, albo nie. Na gruncie interpretacji częstościowej jest błędem stwierdzenie, że przedział o końcach np. (-1,1) zawiera nieznany parametr z prawdopodobieństwem 95%. Prawdopodobieństwo odnosi się do przedziału ufności, jego *realizacja* (a przedział (-1,1) to może być tylko realizacja) nieznany parametr zawiera albo nie! Prawidłowa interpretacja wygląda następująco: przedział o końcach (-1,1) jest realizacją 95% przedziału ufności dla parametru β. Można też powiedzieć, że przedział (-1,1) zawiera nieznaną wartość parametru z *ufnością* 95%, ponieważ nie popadamy tu w kolizję z interpretacją prawdopodobieństwa, które jest ściśle zdefiniowane. (Powyżej –1 i 1 zastępowały wyniki rzeczywistych obliczeń.) Przystąpimy teraz do przedstawienia sposobu konstruowania przedziału ufności dla pojedynczego parametru regresji.

Konstrukcję przedziału ufności można przedstawić w 3 etapach:

1. Ustalamy poziom ufności, czyli ustalamy wartość parametru α ($\alpha \in (0,1)$, w praktyce α bliska 0).
2. Znajdujemy wartość t_α taką, że:

$$P\left\{\left|St(T-k)\right| > t_\alpha\right\} = \alpha$$

(rozkład t-Studenta jest rozkładem symetrycznym co pozwala na proste skonstruowanie przedziału ufności, który jest jednocześnie najkrótszy, wartość t_α znajdujemy w tablicach lub w odpowiednim programie komputerowym)

3. Używamy twierdzenia o rozkładach związanych z normalnym, wykorzystujemy punkt b).

$$P\left\{\left|\frac{\hat{\beta}_i - \beta_i}{D(\hat{\beta}_i)}\right| > t_\alpha\right\} = \alpha \Leftrightarrow P\left\{\left|\frac{\hat{\beta}_i - \beta_i}{D(\hat{\beta}_i)}\right| \le t_\alpha\right\} = 1-\alpha \Leftrightarrow P\left\{\left|\frac{\beta_i - \hat{\beta}_i}{D(\hat{\beta}_i)}\right| \le t_\alpha\right\} = 1-\alpha \Leftrightarrow$$

$$\Leftrightarrow P\left\{-t_\alpha \le \frac{\beta_i - \hat{\beta}_i}{D(\hat{\beta}_i)} \le t_\alpha\right\} = 1-\alpha \Leftrightarrow P\left\{\hat{\beta}_i - t_\alpha D(\hat{\beta}_i) \le \beta_i \le \hat{\beta}_i + t_\alpha D(\hat{\beta}_i)\right\} = 1-\alpha$$

przy powyższych przekształceniach wykorzystujemy fakt, iż $D(\hat{\beta}_i)$ jest większy od zera (jest zdefiniowany jako pierwiastek kwadratowy, elementy przekątniowe macierzy dodatnio określonej, a macierz kowariancji jest dodatnio określona, są większe od zera. Czasem jednak *oszacowanie* macierzy kowariancji może nie spełniać tego warunku, wtedy jest problem, bo nie da się policzyć błędów średnich szacunku.)

Zauważmy, że ostatni człon mówi nam, że nieznana wartość parametru β_i jest większa od $\hat{\beta}_i - t_\alpha D(\hat{\beta}_i)$ i mniejsza od $\hat{\beta}_i + t_\alpha D(\hat{\beta}_i)$ z prawdopodobieństwem równym (1 - α), co tworzy przedział ufności. Przypominamy, że zarówno $\hat{\beta}_i$ jak i $D(\hat{\beta}_i)$ są funkcjami wektora losowego **y**, a więc zmiennymi losowymi.

Ostatecznie: przedział o końcach losowych dany wzorem: $\left(\hat{\beta}_i - t_\alpha D(\hat{\beta}_i), \hat{\beta}_i + t_\alpha D(\hat{\beta}_i)\right)$ jest najkrótszym przedziałem który z prawdopodobieństwem (1 - α) pokrywa nieznaną wartość parametru β_i.

1.3.3.1.2 Testowanie hipotez o pojedynczym parametrze regresji

Weryfikacja hipotez o pojedynczym parametrze regresji jest jednym z podstawowych narzędzi wnioskowania statystycznego, ze względu na prostotę jest rutynowo stosowana np. występuje automatycznie w pakietach analizy statystycznej (w nieco zmienionej postaci – jako tzw. p-value). Jednak prawidłowa interpretacja wyników testowania nie jest trywialna i należy ją przeprowadzać zgodnie z wymogami teorii. Testowanie hipotez statystycznych jest narzędziem mającym pomóc w podejmowaniu decyzji, jednak nie oznacza to, że wynik testu statystycznego zawsze wskazuje konkretną decyzję.

Podstawowym układem w jakim przeprowadza się testowanie jest układ pary hipotez: hipotezy zerowej H_0 i hipotezy alternatywnej H_1, przy założonym z góry poziomie istotności α. Hipotezy mogą być proste tj. określające liczbową wartość parametru, lub złożone tj. określające zbiór wartości parametru. Ogólnie prostą hipotezę zerową i możliwe hipotezy alternatywne można zapisać:

$\mathbf{H_0}$: $\beta_i = \beta_i^*$

Wobec:

$\mathbf{H}_1$: $\beta_i \neq \beta_i^*$ albo $\mathbf{H}_1$: $\beta_i < \beta_i^*$ albo $\mathbf{H}_1$: $\beta_i > \beta_i^*$

Gdzie β_i^* oznacza konkretną, testowaną wartość parametru.

Poziom istotności α musi być przez badacza przyjęty z góry i oznacza wielkość prawdopodobieństwa popełnienia błędu I-go rodzaju tj. odrzucenia hipotezy prawdziwej (tu: przyjęcia H_1 w sytuacji prawdziwości H_0) Poziom ten podlega bezpośredniej kontroli badacza.

Hipoteza zerowa jest hipotezą podlegającą testowaniu, w razie odrzucenia hipotezy zerowej przyjmujemy odpowiednio dobraną hipotezę alternatywną. Typowy wynik testowania hipotez ma postać: „na poziomie istotności α=... nie ma podstaw do odrzucenia hipotezy zerowej H_0 mówiącej, że...” lub: „na poziomie istotności α=... odrzucamy hipotezę zerową mówiącą (...) wobec czego przyjmujemy hipotezę alternatywną H_1, mówiącą, że (...)”. Oczywiście wynik testowania zależy od przyjętego poziomu istotności, i jest warunkowy względem niego. Wybór poziomu istotności należy do badacza. Należy tu zaznaczyć rzecz zasadniczą: wynik testowania mówiący, że nie ma podstaw do odrzucenia hipotezy zerowej nie oznacza, że powinniśmy tę hipotezę zaakceptować. Mówi tylko tyle: nie potrafimy jej odrzucić. Zilustrujmy to przykładem: zwykle interesuje nas szczególnie istotność parametrów regresji, co odpowiada pytaniu: czy można przyjąć, że dany parametr ma wartość 0? (w powyższym zapisie odpowiada to β_i^*= 0). W kontekście regresji zerowa wartość parametru oznacza, że odpowiadająca mu zmienna objaśniająca nie ma żadnego liniowego wpływu na zmienną objaśnianą, czyli powinna być usunięta z modelu. Problem decyzyjny jest następujący: czy usunąć i-tą zmienną i reestymować model, czy pozostawić ją w modelu. W tej sytuacji

przeprowadzamy testowanie w następującym układzie hipotez: H_0: $\beta_i = 0$; H_1: $\beta_i \neq 0$ na poziomie istotności wynoszącym np. $\alpha = 0{,}05$. Przypuśćmy, że wynik testu mówi: na poziomie istotności $\alpha = 0{,}05$ odrzucamy hipotezę H_0: $\beta_i = 0$ wobec czego przyjmujemy hipotezę alternatywną: H_1: $\beta_i \neq 0$. Oznacza to, że na przyjętym poziomie istotności parametr β_i jest statystycznie istotny, czyli nie powinniśmy usuwać odpowiadającej mu zmiennej z modelu. Możemy (jeśli przyjęliśmy wynik testu jako kryterium) podjąć decyzję: i-ta zmienna objaśniająca pozostaje w modelu. Jednak jeśli napotkamy wynik testu mówiący: nie ma podstaw do odrzucenia hipotezy H_0: $\beta_i = 0$, decyzja jest trudniejsza. Może to bowiem odzwierciedlać sytuację kiedy nie będziemy w stanie odrzucić podobnej hipotezy dla dużego zakresu wartości różnych od zera. Np. kolejny test może dać wynik: na zadanym poziomie istotności nie ma podstaw do odrzucenia H_0: $\beta_i = 10$. W tym wypadku test nie daje nam odpowiedzi, co oznacza że musimy kierować się innym kryterium. Jest to sytuacja powszechnie spotykana, badacz musi w niej podjąć decyzję według innego kryterium, bowiem wniosek: „nie ma podstaw do odrzucenia hipotezy H_0: $\beta_i = 0$ więc przyjmujemy $\beta_i = 0$ i usuwamy zmienną" nie jest logiczną implikacją samego rezultatu testu, lecz wynikiem subiektywnych kryteriów i preferencji badacza, i powinien jako taki być rozumiany. W szczególności inny badacz na podstawie identycznego wyniku testu może podjąć decyzję: „pozostawiam i-tą zmienną w modelu".

Wynik takiego testu zależy oczywiście od przyjętego poziomu istotności. W pakietach statystycznych automatycznie przeprowadzane jest testowanie istotności poszczególnych parametrów regresji. Jego wynik jest jednak podany w postaci tzw. p-value, to jest najniższego poziomu istotności dla którego można odrzucić hipotezę H_0: $\beta_i = 0$. Wartość p-value wynosząca 0,001 oznacza, że przyjmując że wartość parametru jest różna od zera (czyli przyjmując statystyczną istotność parametru) pomylimy się (czyli prawdziwa wartość parametru wyniesie 0) „1 raz na 1000" (interpretacja jest w cudzysłowie, ponieważ jest to problem analogiczny do interpretacji konkretnych realizacji przedziałów ufności, czyli interpretacji częstościowej pojedynczej realizacji; można by powiedzieć, że gdybyśmy obserwowali „bardzo wiele" realizacji danych i przeprowadzali test, to pomylilibyśmy się przeciętnie 1 raz na 1000). Oczywiście znajomość p-value pozwala nam wyznaczyć wyniki testów dla wszystkich innych poziomów istotności. Dla p-value wynoszącego 0,001 jesteśmy wstanie odrzucić H_0: $\beta_i = 0$ na poziomie istotności $\alpha = 0{,}05$ (i wszystkich większych od 0,001) natomiast „nie ma podstaw do odrzucenia H_0: $\beta_i = 0$" na poziomie istotności $\alpha = 0{,}0005$ (i wszystkich mniejszych od 0,001). Rozumiejąc p-value jako prawdopodobieństwo uzyskania danej oceny parametru w sytuacji, gdy jego prawdziwa wartość wynosi 0, niskie wartości świadczą o statystycznej istotności parametru.

Należy tu zaznaczyć jeszcze jedną istotną kwestię. Wyżej opisana procedura odnosi się do pojedynczego testu istotności pojedynczego współczynnika regresji. Nie można analogicznie interpretować wyników serii takich testów przeprowadzonych na kolejnych współczynnikach. W szczególności oznacza to, że oddzielnie testowana statystyczna istotność wszystkich pojedynczych parametrów regresji na poziomie istotności α **nie** implikuje łącznej istotności tych współczynników na poziomie α. Wnioskowanie o więcej niż jednym parametrze regresji będzie omówione poniżej. Tu zaznaczamy że należy wyraźnie rozróżnić rezultat pojedynczego testu od rezultatu sekwencji testów, i w tym ostatnim przypadku stosować należy odpowiednie odmienne procedury.

Testowanie hipotez przeprowadzamy w oparciu o statystyki testowe. Związki testowania hipotez i budowy przedziałów ufności wynikają ze stosowania tych samych statystyk i są następujące: jeśli $(1 - \alpha)*100$ procentowy przedział ufności zawiera wartość β_i^*, to nie jesteśmy w stanie na poziomie istotności α odrzucić hipotezy zerowej: H_0: $\beta_i = \beta_i^*$. Możemy jednak odrzucić tę hipotezę na tym samym poziomie istotności jeśli przyjmiemy wartość β_i^* spoza odpowiedniego przedziału ufności. Widać stąd, że jeśli przedział istotności obejmuje 0, nie jesteśmy w stanie odrzucić hipotezy o statystycznej nieistotności parametru. Nie jest to równoznaczne z przyjęciem wartości $\beta_i = 0$.

Konstrukcja testu statystycznego.

Do skonstruowania testu konieczna jest statystyka testowa, czyli losowa funkcja o jednym nieznanym parametrze (podlegającym wnioskowaniu) i znanym rozkładzie przy prawdziwości hipotezy zerowej H_0. (jeśli rozkład statystyki testowej przy prawdziwości H_0 nie jest znany, nie możemy przeprowadzić testu, jest to np. przypadek testowania standardowym testem tzw. unit root czyli wielkości pewnego parametru = 1; stosowanie standardowego testu jest niemożliwe, ponieważ prawdziwości hipotezy zerowej odpowiadają drastycznie inne własności rozkładu statystyki testowej). Dla konstrukcji omawianego testu skorzystamy ze statystyki danej w punkcie b) twierdzenia o rozkładach związanych z normalnym. Jeśli testujemy parę hipotez:

$\mathbf{H_0}$: $\beta_i = \beta_i^*$

Wobec:

$\mathbf{H}_1$: $\beta_i \neq \beta_i^*$

1. Ustalamy poziom istotności testu α.
2. Znajdujemy wartość t_α taką, że:
$$P\left\{\left|St(T-k)\right| > t_\alpha\right\} = \alpha$$
3. Używamy twierdzenia o rozkładach związanych z normalnym, wykorzystujemy punkt b), wyliczamy wartość statystyki:

$$\left|\frac{\hat{\beta}_i - \beta_i^*}{D(\hat{\beta}_i)}\right|$$

{interesuje nas wartość statystyki testowej {i jej rozkład} przy prawdziwości hipotezy zerowej (podstawiamy $\beta_i = \beta_i^*$)},

jeśli wartość statystyki jest:

a) większa od t_α - odrzucamy hipotezę zerową i przyjmujemy hipotezę alternatywną.
b) mniejsza lub równa t_α - nie ma podstaw do odrzucenia hipotezy zerowej.

Testowanie prostych hipotez o pojedynczych współczynnikach regresji daje ograniczone możliwości wnioskowania, ograniczenia zarysowane są powyżej. Szersze możliwości wnioskowania daje nam łączne testowanie hipotez o parametrach regresji, omówione poniżej.

1.3.3.2 Łączne testowanie liniowych równościowych restrykcji na parametry regresji

Poniżej przedstawiony zostanie dość ogólny i bardzo użyteczny układ założeń umożliwiających testowanie liniowych równościowych warunków pobocznych dotyczących wielu współczynników regresji łącznie. Pozwoli to kontrolować łączny poziom istotności i wykonać odpowiedni test (przeprowadzenie serii testów o pojedynczych współczynnikach regresji miałoby trudny do kontrolowania łączny poziom istotności). Przedstawione zostaną też dwa ważne szczególne przypadki powyższego schematu oraz estymator MNK uwzględniający restrykcje. Szczególne przypadki będą odnosiły się do badania łącznej istotności wszystkich współczynników regresji poza wyrazem wolnym (podstawowy test w kontekście regresji) oraz do problemu istotności bloku współczynników regresji, co odpowiada testowaniu redukcji modelu. Postać ogólna daje szersze możliwości wnioskowania; w szczególności pozwala na wnioskowanie o liniowych kombinacjach współczynników regresji. Często właśnie liniowe kombinacje współczynników mają interpretację ekonomiczną i w związku z tym wymagają testowania. Może to wynikać z faktu, że ekonomiczna charakterystyka jest dana jako liniowa kombinacja parametrów, przy czym chcemy wnioskować o jej wielkości. Ponadto w sytuacji gdy z teorii ekonomii wynikają ograniczenia dotyczące postaci szacowanej

funkcji, to jeśli treść tych ograniczeń można sprowadzić do narzucenia równościowych warunków pobocznych (restrykcji) na liniowe kombinacje parametrów, opisany poniżej test pozwoli sprawdzić, czy oszacowany model spełnia wymagania stawiane przez teorię ekonomii. Do narzucenia a priori na model takich restrykcji możemy wykorzystać estymator MNK z restrykcjami opisany poniżej – pozwoli on na oszacowanie wartości parametrów przy założeniu prawdziwości restrykcji. Należy zaznaczyć, że schemat poniższy dotyczy wyłącznie restrykcji równościowych; nakładanie restrykcji nierównościowych i estymacja parametrów w ich obecności należą do bardziej zaawansowanych zagadnień.

Ogólny schemat łącznego testowania p liniowych restrykcji na parametry regresji.

Założenia:

- testowaniu podlega p liniowych równościowych restrykcji.
- liczba parametrów regresji wynosi k i $k \geq p$
- restrykcje przedstawiamy w postaci macierzowego zapisu układu równań liniowych:

$$\underset{(p\times k)}{\mathbf{S}}\ \underset{(k\times 1)}{\boldsymbol{\beta}} = \underset{(p\times 1)}{\mathbf{c}} \quad ; \mathrm{rz}(\mathbf{S}) = p$$

gdzie **S** to znana macierz rzędu p, a **c** to znany wektor.

- hipotezy testowe są postaci:

H_0: $\mathbf{S}\boldsymbol{\beta} = \mathbf{c}$ wobec
H_1: $\mathbf{S}\boldsymbol{\beta} \neq \mathbf{c}$

Co logicznie odpowiada układowi: H_0 : wszystkie restrykcje są prawdziwe - wobec H_1 : przynajmniej jedna restrykcja nie jest prawdziwa.

Powyższy układ równań specyfikuje wartości p liniowych kombinacji współczynników regresji. Wymaganie dotyczące rzędu macierzy **S** wymaga, aby były to odmienne (nie dublujące informacji) restrykcje.

Statystyka testowa (tzw. F empiryczne) w tym wypadku ma postać:

$$F_{emp} = \frac{(\mathbf{S}\hat{\boldsymbol{\beta}} - \mathbf{c})'[\mathbf{S}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{S}']^{-1}(\mathbf{S}\hat{\boldsymbol{\beta}} - \mathbf{c})/p}{s^2}$$

Przy prawdziwości hipotezy zerowej rozkład tej statystyki jest następujący:

$$H_0 \Rightarrow F_{emp} \sim F(p, T-k)$$

F oznacza rozkład Fischera – Snedecora o liczbie stopni swobody jak w nawiasie.

W liczniku wzoru na F_{emp} jest dodatnio określona macierz (w nawiasach kwadratowych) obłożona wektorem i jego transpozycją. Taka konstrukcja to forma kwadratowa i jej wartość jest skalarem. Wartość jaką przyjmuje forma kwadratowa dla dowolnego niezerowego wektora zależy od określoności macierzy. Tu macierz jest dodatnio (nieujemnie?) określona z konstrukcji; wobec tego cała forma kwadratowa będzie przyjmować wartości dodatnie (jak i cała statystyka, bo p i s^2 są również większe od 0). Wartości formy kwadratowej będą tym mniejsze, im mniejsze będą składowe wektora $(\mathbf{S}\hat{\boldsymbol{\beta}} - \mathbf{c})$. Ponieważ prawdziwość restrykcji odpowiada zerowej wartości wektora, małe wartości F empirycznego świadczyć będą za hipotezą zerową. Szukamy więc wartości krytycznej F_α powyżej której odrzucimy hipotezę zerową. Test F jest więc testem asymetrycznym. Należy również zaznaczyć, że jest to tzw. test dokładny, tzn. nie odwołujemy się tu do asymptotycznego rozkładu statystyki testowej. Testy asymptotyczne podają rozkład statystyki testowej przy liczbie obserwacji zmierzającej do nieskończoności i w ich praktycznym stosowaniu zawsze pojawia się

problem: na ile nieznany rozkład małopróbkowy (dotyczący realnej sytuacji posiadania skończonej i często niewielkiej liczby obserwacji) można przybliżyć znanym rozkładem asymptotycznym, inaczej: „jaki popełnimy błąd przyjmując rozkład asymptotyczny jako rozkład dokładny?". W celu rozwiązania tego problemu stosuje się rozmaite poprawki, wykonuje się symulacje itd. Stosowanie testu dokładnego pozwala uniknąć problemów związanych z możliwą rozbieżnością między rozkładem asymptotycznym a dokładnym.

Procedura testowa jest tu następująca:

1. Ustalamy poziom istotności testu α oraz postać restrykcji (wartości p, **S**, **c**).
2. Znajdujemy wartość F_α :
$$F_\alpha : P\{F_{(p,T-k)} > F_\alpha\} = \alpha$$
3. Wyliczamy wartość statystyki F_{emp},

jeśli wartość statystyki jest:

a) większa od F_α - odrzucamy hipotezę zerową i przyjmujemy hipotezę alternatywną (przynajmniej jedna restrykcja nie jest prawdziwa).
b) mniejsza lub równa F_α - nie ma podstaw do odrzucenia hipotezy zerowej.

Możemy zastosować test F do przetestowania hipotezy o pojedynczym parametrze regresji ($\beta_i = \beta_i^*$ wobec $\beta_i \neq \beta_i^*$) Mamy wówczas: p=1, **c** = c = [β_i^*] a macierz **S** staje się wektorem 1 na k składającym się z zer i z jedynką na i-tej pozycji. Wówczas (co warto przeliczyć) z wzoru na F empiryczne otrzymujemy kwadrat statystyki testowej z punktu dotyczącego testowania hipotez o pojedynczym współczynniku regresji. (forma kwadratowa w [] „wybiera" z macierzy $(\mathbf{X'X})^{-1}$ i-ty element i bierze jego odwrotność - ten element wraz z s^2 z mianownika daje błąd średni szacunku do kwadratu w mianowniku; w liczniku $(\mathbf{S\hat{\beta}} - \mathbf{c})$ daje $\hat{\beta}_i - \beta_i^*$ co jako skalar nie jest zmieniane przez transpozycję i podniesione do kwadratu). Natomiast rozkład F(1,T-k) jest rozkładem kwadratu zmiennej St(T-k), test t jest testem dwustronnym i symetrycznym wobec czego podniesienie statystyki testowej do kwadratu nie powoduje utraty informacji; widać wobec tego że test F zastosowany dla pojedynczego współczynnika sprowadza się do testu t-Studenta; obie te procedury są równoważne.

Inny ważny przypadek szczególny testu F odpowiada problemowi redukcji modelu. Załóżmy, że w modelu regresji zastanawiamy się nad istotnością całej grupy zmiennych objaśniających. Gdyby odpowiadające im współczynniki regresji były równocześnie zerami, model redukowałby się do prostszej postaci. Nie można testować osobno parametrów regresji odpowiadających grupie zmiennych, ponieważ interesuje nas test łączny. Mamy więc k wszystkich zmiennych objaśniających (i parametrów regresji) próbujemy testować łączną istotność p z nich (H_0: p odpowiednich współczynników równocześnie równych 0). Przyjmijmy następujące oznaczenia: podzielimy macierz **X** i wektor $\boldsymbol{\beta}$ odpowiednio na części odpowiadające obu grupom zmiennych (w zapisie blokowym macierzy):

$$\underset{(T\times k)}{\mathbf{X}} = [\mathbf{X_1}\ \mathbf{X_2}]; \quad \underset{(k\times 1)}{\boldsymbol{\beta}} = \begin{bmatrix} \boldsymbol{\beta}^{(1)} \\ \boldsymbol{\beta}^{(2)} \end{bmatrix} \quad ; \quad \underset{(T\times 1)}{\mathbf{y}} = \underset{(T\times(k-p))}{\mathbf{X_1}} * \underset{((k-p)\times 1)}{\boldsymbol{\beta}^{(1)}} + \underset{(T\times p)}{\mathbf{X_2}} * \underset{(p\times 1)}{\boldsymbol{\beta}^{(2)}} + \underset{(T\times 1)}{\boldsymbol{\varepsilon}} \ ;$$

bloki z indeksem „2" odpowiadają zmiennym, których istotność ma podlegać testowaniu, bloki z indeksem (1) opisują model zredukowany, ich istotność nie ma podlegać testowaniu. Postać macierzy **S** i wektora **c** odpowiadająca temu zagadnieniu jest następująca:

$$\mathbf{S} = [\underset{((k-p)\times p)}{\mathbf{0}} \quad \underset{(p\times p)}{\mathbf{I}}]; \quad \underset{(p\times 1)}{\mathbf{c}} = \underset{(p\times 1)}{\mathbf{0}}$$

gdzie **I** oznacza macierz jednostkową. Wobec tego układ restrykcji $\mathbf{S\beta} = \mathbf{c}$ oznacza $\boldsymbol{\beta}^{(2)} = \mathbf{0}$. Hipoteza zerowa mówi, że nie ma podstaw do odrzucenia stwierdzenia, że wszystkie p współczynników regresji odpowiadających zmiennym zebranym w macierzy $\mathbf{X}_2$ jest równa zero. Przyjęcie tej hipotezy jest równoznaczne z redukcją całego modelu do postaci $\mathbf{y} = \mathbf{X}_1 \boldsymbol{\beta}^{(1)} + \boldsymbol{\varepsilon}$. Hipoteza alternatywna mówi: przynajmniej jeden z badanych współczynników jest różny od zera, co oznacza, że przynajmniej jedna zmienna z macierzy $\mathbf{X}_2$ ma istotny liniowy wpływ na zmienną objaśniającą.

W tej sytuacji F_{emp} można wyliczyć z uproszczonego wzoru:

$$\mathbf{F_{emp}} = \frac{(SSE_0 - SSE_1)/p}{SSE_1/(T-k)}$$

gdzie SSE_i to suma kwadratów reszt w modelu odpowiadającym i-tej hipotezie; SSE_0 odpowiada modelowi zredukowanemu (policzenie SSE_0 wymaga oddzielnej estymacji modelu $\mathbf{y} = \mathbf{X}_1 \boldsymbol{\beta}^{(1)} + \boldsymbol{\varepsilon}$), natomiast SSE_1 odpowiada modelowi pełnemu $\mathbf{y} = \mathbf{X\beta} + \boldsymbol{\varepsilon}$, wobec czego $SSE_0 \geq SSE_1$ (przy czym równość jest możliwa z prawdopodobieństwem 1, co wynika z własności MNK; intuicyjnie model z większą ilością zmiennych objaśniających (nawet nieistotnych) nie może mieć gorszej (większej) sumy kwadratów reszt od modelu w którym tych zmiennych jest mniej). Należy pamiętać, że do skorzystania z tego wzoru konieczna jest dwukrotna estymacja – raz modelu pełnego, i raz modelu zredukowanego. Widać też, że w mianowniku jest po prostu s^2 dla pełnego modelu.

Szczególnym przypadkiem powyższej procedury jest łączne testowanie istotności wszystkich parametrów regresji z wyjątkiem wyrazu wolnego. Odpowiada to macierzy $\mathbf{X}_1$ będącej wektorem jedynek, macierzy $\mathbf{X}_2$ grupującej wszystkie inne zmienne objaśniające. Taki model zredukowany odpowiada modelowi gdzie $y_t = \beta_1 + \varepsilon_t$, i β_1 równa średniej arytmetycznej z elementów **y**. Gdy wynik testu sugeruje redukcję modelu, oznacza to, że ani jedna z wybranych zmiennych objaśniających nie ma statystycznie istotnego liniowego wpływu na zmienną objaśnianą. Stawia to sensowność modelu pod znakiem zapytania. Jeśli natomiast odrzucimy hipotezę zerową, oznacza to, że przynajmniej jedna zmienna objaśniająca ma istotny liniowy wpływ na zmienną objaśnianą. (i model nie musi być bezsensowny). W wypadku testowania łącznej istotności wszystkich parametrów regresji z wyjątkiem wyrazu wolnego można skorzystać z następującego wzoru na F_{emp}:

$$F_{emp} = \frac{R^2/(k-1)}{(1-R^2)/(T-k)}$$

1.3.4 Estymator MNK z narzuconymi liniowymi restrykcjami równościowymi

Estymator MNK z narzuconymi liniowymi równościowymi warunkami pobocznymi $\hat{\boldsymbol{\beta}}_r$ oraz jego macierz kowariancji mają postać:

$$\hat{\boldsymbol{\beta}}_r = \hat{\boldsymbol{\beta}} + (\mathbf{X'X})^{-1}\mathbf{S'}[\mathbf{S}(\mathbf{X'X})^{-1}\mathbf{S'}]^{-1}(\mathbf{S}\hat{\boldsymbol{\beta}} - \mathbf{c})$$
$$\mathbf{V}(\hat{\boldsymbol{\beta}}_r) = \sigma^2(\mathbf{X'X})^{-1} - \sigma^2(\mathbf{X'X})^{-1}\mathbf{S'}[\mathbf{S}(\mathbf{X'X})^{-1}\mathbf{S'}]^{-1}\mathbf{S}(\mathbf{X'X})^{-1}$$

Estymator ten jest nieobciążony jeśli restrykcje są prawdziwe. Przy prawdziwości restrykcji i założeniach KMRL jest on najlepszy w klasie estymatorów liniowych nieobciążonych spełniających restrykcje. Przy dodaniu założenia 6° staje się estymatorem najlepszym w klasie wszystkich estymatorów nieobciążonych spełniających restrykcje. Błędy średnie szacunku tego estymatora będą mniejsze, niż błędy średnie szacunku estymatora MNK, ale mają one interpretację tylko wtedy, kiedy estymator jest nieobciążony,

czyli kiedy restrykcje są prawdziwe. Aby stwierdzić, że $\hat{\boldsymbol{\beta}}_r$ istotnie spełnia restrykcje, warto podstawić $\mathbf{S}\hat{\boldsymbol{\beta}}_r$ i wyliczyć rezultat.

1.4 Model Normalnej Regresji Nieliniowej (MNRN)

1.4.1 Założenia Modelu Normalnej Regresji Nieliniowej

W modelu normalnej regresji nieliniowej (MNRN) wyróżniamy wektor obserwacji na zmiennej objaśnianej **y**, jednak bardziej ogólnie niż w KMNRL zakładamy:

1° $y_t = f(\mathbf{x}_t;\boldsymbol{\beta}) + \varepsilon_t$ gdzie $t = (1,2,....,T)$ oraz:

$\mathbf{x}_t' = \begin{pmatrix} x_{t1} & x_{t2} & ... & x_{tp} \end{pmatrix}$ gdzie $\mathbf{x}_t'$ jest (1×p) wektorem wartości p zmiennych objaśniających odpowiadających obserwacji o numerze t; natomiast:

$\underset{(k\times 1)}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_1 \\ \cdots \\ \beta_k \end{bmatrix}$ jest (k×1) wymiarowym wektorem nieznanych parametrów,

i $\boldsymbol{\beta} \in \mathbf{B} \subset \Re^k$, gdzie **B** jest zbiorem otwartym;

$f(\mathbf{x}_t;\boldsymbol{\beta})$ jest znaną funkcją klasy C^2 względem $\boldsymbol{\beta}$.

2° $\mathbf{x}_t$ są wektorami nielosowymi lub wektorami losowymi niezależnymi stochastycznie od składników losowych:

$\mathbf{x}_t \perp\!\!\!\perp \varepsilon_s \quad t, s = 1, 2, ... , T$

3° macierz $\mathbf{A}(\boldsymbol{\beta})$ pochodnych cząstkowych funkcji f względem parametrów β dana wzorem:

$$\underset{(T\times k)}{\mathbf{A}(\boldsymbol{\beta})} = \begin{bmatrix} \frac{\partial f(\mathbf{x}_1;\boldsymbol{\beta})}{\partial \beta_1} & \frac{\partial f(\mathbf{x}_1;\boldsymbol{\beta})}{\partial \beta_2} & \cdots & \frac{\partial f(\mathbf{x}_1;\boldsymbol{\beta})}{\partial \beta_k} \\ \frac{\partial f(\mathbf{x}_2;\boldsymbol{\beta})}{\partial \beta_1} & \frac{\partial f(\mathbf{x}_2;\boldsymbol{\beta})}{\partial \beta_2} & \cdots & \frac{\partial f(\mathbf{x}_2;\boldsymbol{\beta})}{\partial \beta_k} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial f(\mathbf{x}_T;\boldsymbol{\beta})}{\partial \beta_1} & \frac{\partial f(\mathbf{x}_T;\boldsymbol{\beta})}{\partial \beta_2} & \cdots & \frac{\partial f(\mathbf{x}_T;\boldsymbol{\beta})}{\partial \beta_k} \end{bmatrix}$$

jest macierzą rzędu k dla prawie wszystkich możliwych $\boldsymbol{\beta}$.

Pozostałe założenia mają taką samą postać jak w KMNRL:

4° $\underset{(T\times 1)}{E(\boldsymbol{\varepsilon})} = \mathbf{0}$

5° $V(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I_T}$ $\sigma^2>0$

6° $\boldsymbol{\varepsilon}$ ma T-wymiarowy rozkład normalny

Generalizacja KMNRL prowadzi do MNRN; relacja tych dwóch modeli dobrze obrazuje przypadek bardzo często w ekonometrii spotykany: uogólnienie modelu A prowadzi do modelu B o słabszych założeniach; A ma wszystkie własności B, bo jest jego szczególnym przypadkiem, może posiadać też dodatkowe własności wynikające z mocniejszych założeń. Uogólnienie modelu pozwala nam rozszerzyć jego stosowalność: możemy zastosować B tam, gdzie zastosowanie A nie byłoby możliwe. Ceną za to jest zwykle osłabienie mocy konkluzji modelu B. Z modelu B możemy też wyprowadzić przypadki szczególne inne niż A, o założeniach i tezie mocniejszej od B. Relacje te mogą być dość złożone, jednak dokładnie ich zrozumienie pozwala na właściwe i świadome dobranie założeń oraz zrozumienie ich konsekwencji. Jest to zagadnienie mające kluczowy wpływ na sensowność uzyskiwanych wyników. W celu zilustrowania tego zagadnienia najpierw omówmy dokładniej założenia MNRN.

Założenia 1°-3° odróżniające MNRN od KMNRL opisują generalizację deterministycznej części modelu; właściwości części stochastycznej (4°-6°) pozostają niezmienione.

Ad 1°. Model MNRN objaśnia powstawanie T obserwacji na zmiennej y, jednak dopuszcza, by liczba zmiennych objaśniających (p) była inna, niż liczba nieznanych parametrów modelu (k) (wynika to z potrzeb nieliniowej postaci f). Nieznane parametry zgrupowane są w wektor $\boldsymbol{\beta}$ należący do zbioru **B** będącego otwartym podzbiorem k-wymiarowej przestrzeni rzeczywistej. Zbiór otwarty to taki, w którym dla każdego punktu $\boldsymbol{\beta}\in \mathbf{B}$ istnieje otoczenie należące w całości do zbioru **B**, przykłady zbiorów otwartych to: koło minus okrąg (koło bez brzegu), czy kula minus sfera, (patrz skrypt prof. Malawskiego ☺). Jest to założenie mające wykluczyć tzw. rozwiązania brzegowe tj sytuacje, gdy punkt $\boldsymbol{\beta}$ spełniający kryteria estymacji (np. minimalizujący sumę kwadratów reszt) leży na brzegu zbioru **B**, co owocuje trudnościami teoretycznymi głębszej natury. { w praktyce nie wiedzieć czemu przyjmuje się $\beta_i \geq 0$ niby z założeń ekonomicznych (zamiast $\beta_i > 0$, co spełnia wymóg otwartości), wychodzi $\beta_1 = 0$ i niby jest OK., ale tak uzyskane oszacowanie nie spełnia prawidłowo postawionych wymogów modelu. Żeby to rozstrzygnąć, trzeba by wiedzieć dokładnie, jakie są własności rozwiązań brzegowych i co z tego. Ja nie mam pojęcia. BM}. Funkcja f ma być klasy C^2 względem β co oznacza, że drugie pochodne funkcji f względem każdego elementu $\boldsymbol{\beta}$ muszą istnieć i być ciągłe w każdym punkcie. Zapis założenia 1° można przedstawić następująco:

$$\underset{(T\times 1)}{\mathbf{y}} = \underset{(T\times 1)}{\mathbf{g}}\left(\underset{(T\times p)}{\mathbf{X}}; \underset{(k\times 1)}{\boldsymbol{\beta}}\right) + \underset{(T\times 1)}{\boldsymbol{\varepsilon}};$$

$$\text{gdzie } \underset{(T\times 1)}{\mathbf{g}}(\mathbf{X};\boldsymbol{\beta}) = \begin{bmatrix} f(\mathbf{x}_1;\boldsymbol{\beta}) \\ f(\mathbf{x}_2;\boldsymbol{\beta}) \\ \cdots \\ f(\mathbf{x}_T;\boldsymbol{\beta}) \end{bmatrix} \text{ oraz } \underset{(T\times p)}{\mathbf{X}} = \begin{bmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \cdots \\ \mathbf{x}_T' \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ x_{T1} & x_{T2} & \cdots & x_{Tp} \end{bmatrix}$$

Przejście od liniowej do nieliniowej postaci zależności ogromnie rozszerza możliwości zastosowania omawianego modelu. Modele dające się sprowadzić do postaci liniowej lub dobrze aproksymować postacią liniową łatwiej jest szacować w układzie założeń KMRL, jednak nie można a priori zakładać możliwości sprowadzenia zależności ekonomicznych do restrykcyjnej postaci liniowej.

Ad. 2°. Założenie drugie ma częściowo nową postać, można je przedstawić w odniesieniu do powyższego zapisu tak: **X** jest znaną macierzą nielosową (jak w KMNRL) lub X jest macierzą losową stochastycznie niezależną od wektora składników losowych. Stochastyczna niezależność jest ściśle zdefiniowana, jednak definicja nie będzie tu przedstawiona. Dopuszczenie losowości macierzy wartości

zmiennych objaśniających **X** uchyla krytykowane już założenie odpowiadające przeprowadzeniu kontrolowanego eksperymentu. Wartości zmiennych objaśniających są zwykle również wielkościami powstającymi w procesach ekonomicznych i przyjęcie ich losowości samo się narzuca. Rozszerzenie tego założenia nastąpi w przedstawionym dalej modelu (regresja z losowymi zmiennymi objaśniającymi), tam też skomentowany zostanie wymóg stochastycznej niezależności **X** od $\boldsymbol{\varepsilon}$, tu zatrzymujemy się na takiej postaci 2°. Można ją uzasadnić poprzez założenie, że źródła losowych wahań wartości zmiennych objaśniających **X** oraz składników losowych $\boldsymbol{\varepsilon}$ są inne.

Ad. 3°. Założenie to dotyczy rzędu macierzy $\mathbf{A}(\boldsymbol{\beta})$ ($T \times k$; $T > k$) – ma on być równy k, czyli liczbie kolumn, „dla prawie wszystkich $\boldsymbol{\beta}$". Tu należy szczególnie ostrożnie należy traktować słowo „prawie". Jak zwykle w ekonometrii, traci ono swoje potoczne luźne znaczenie, i nabiera znaczenia ściśle określonego. „Prawie" należy rozumieć „z wyjątkiem podzbiorów miary zero". A co to oznacza? To zależy od miary, ale intuicyjnie można powiedzieć tak: w tym wypadku $\mathrm{rz}[\mathbf{A}(\boldsymbol{\beta})] = k$ dla $\boldsymbol{\beta} \in \mathbf{B}$ z możliwym wyjątkiem zbioru zawierającego się w **B** i będącego podzbiorem przestrzeni o wymiarze mniejszym niż k. Intuicyjnie, gdyby zbiór **B** był kołem bez brzegu, to podzbiór **B** dla którego $\mathrm{rz}[\mathbf{A}(\boldsymbol{\beta})] \neq k$ mógłby być częścią wspólną pewnej hiperboli i **B**, i założenie pozostawałoby spełnione. Gdy dopuścimy losowość macierzy **X**, sprawa jeszcze się komplikuje: wymagamy, by „$\mathrm{rz}[\mathbf{A}(\boldsymbol{\beta})] = k$ dla prawie wszystkich $\boldsymbol{\beta}$ i prawie zawsze"; „prawie zawsze" oznacza „z prawdopodobieństwem równym jeden". Zdarzenie zachodzące z prawdopodobieństwem 1 nie jest równoznaczne ze „zdarzeniem pewnym". Zdarzenie pewne jest ściśle zdefiniowane, jest to zbiór zawierający wszystkie elementy zbioru zdarzeń elementarnych, w wypadku zmiennej losowej – zbiór zawierający wszystkie możliwe jej realizacje. Jeśli od tego zbioru odejmiemy niepuste podzbiory miary zero (w sensie miary prawdopodobieństwa), otrzymamy zbiór odpowiadający zdarzeniu prawie pewnemu, zachodzi ono z prawdopodobieństwem 1 (ponieważ „różni się" od zdarzenia pewnego o podzbiory miary 0, jednak nie jest z nim tożsame). Ciągła zmienna losowa przyjmuje z prawdopodobieństwem jeden wartości $x \neq x^*$, czyli przyjmuje wartości różne od x^* prawie zawsze. Pojęcia „z prawdopodobieństwem jeden" , „prawie zawsze" są odległe od potocznej intuicji, trzeba się jednak z nimi oswoić, bo w ekonometrii występują często (jak również ich odpowiedniki „prawie nigdy" „z prawdopodobieństwem 0" – ciągła zmienna losowa X przyjmuje wartość x^* z prawdopodobieństwem 0). Można się też domyślić, że założenie dotyczące rzędu ma m.in. zapewnić możliwość odwrócenia macierzy i przez to istnienie odpowiednich formuł.

Teraz wykażemy, że KMNRL jest w istocie szczególnym przypadkiem MNRN. Aby to zrobić, przyjmijmy odpowiednią postać funkcji f, i wyprowadźmy odpowiadające jej wektor **g** i macierz $\mathbf{A}(\boldsymbol{\beta})$. W wypadku liniowej postaci funkcji f liczba zmiennych objaśniających jest równa liczbie parametrów, czyli $p = k$.

$$\underset{(T\times 1)}{\mathbf{g}}(\mathbf{X};\boldsymbol{\beta}) = \begin{bmatrix} f(\mathbf{x}_1;\boldsymbol{\beta}) \\ f(\mathbf{x}_2;\boldsymbol{\beta}) \\ \cdots \\ f(\mathbf{x}_T;\boldsymbol{\beta}) \end{bmatrix} = \begin{bmatrix} \beta_1 x_{11} + & \beta_2 x_{12} + & \cdots & +\beta_k x_{1k} \\ \beta_1 x_{21} + & \beta_2 x_{22} + & \cdots & +\beta_k x_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ \beta_1 x_{T1} + & \beta_2 x_{T2} + & \cdots & +\beta_k x_{Tk} \end{bmatrix} = \mathbf{X}\boldsymbol{\beta}$$

$$\underset{(T\times k)}{\mathbf{A}(\boldsymbol{\beta})} = \begin{bmatrix} \dfrac{\partial f(\mathbf{x}_1;\boldsymbol{\beta})}{\partial \beta_1} & \dfrac{\partial f(\mathbf{x}_1;\boldsymbol{\beta})}{\partial \beta_2} & \cdots & \dfrac{\partial f(\mathbf{x}_1;\boldsymbol{\beta})}{\partial \beta_k} \\ \dfrac{\partial f(\mathbf{x}_2;\boldsymbol{\beta})}{\partial \beta_1} & \dfrac{\partial f(\mathbf{x}_2;\boldsymbol{\beta})}{\partial \beta_2} & \cdots & \dfrac{\partial f(\mathbf{x}_2;\boldsymbol{\beta})}{\partial \beta_k} \\ \cdots & \cdots & \cdots & \cdots \\ \dfrac{\partial f(\mathbf{x}_T;\boldsymbol{\beta})}{\partial \beta_1} & \dfrac{\partial f(\mathbf{x}_T;\boldsymbol{\beta})}{\partial \beta_2} & \cdots & \dfrac{\partial f(\mathbf{x}_T;\boldsymbol{\beta})}{\partial \beta_k} \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ x_{T1} & x_{T2} & \cdots & x_{Tp} \end{bmatrix} = \mathbf{X}$$

ponieważ:

$$f(\mathbf{x}_t;\boldsymbol{\beta}) = \beta_1 x_{t1} + \beta_2 x_{t2} + ... + \beta_k x_{tk} \qquad \text{oraz}$$

$$\frac{\partial f(\mathbf{x}_t;\boldsymbol{\beta})}{\partial \beta_i} = \frac{\partial(\beta_1 x_{t1} + ... + \beta_i x_{ti} + ... + \beta_k x_{tk})}{\partial \beta_i} = \frac{\partial(\beta_i x_{ti})}{\partial \beta_i} = x_{ti}$$

Po podstawieniu powyższych rezultatów do założeń MNRN ($\mathbf{A}(\boldsymbol{\beta}) = \mathbf{X}$ oraz $\mathbf{g} = \mathbf{X}\boldsymbol{\beta}$) otrzymujemy układ założeń KMNRL. W ten sposób pokazaliśmy, że KMNRL jest szczególnym przypadkiem MNRN i wobec tego własności MNRN będą występowały w KMNRL. By stwierdzić, jakie to będą własności, omówić musimy rezultaty przyjęcia założeń MNRN.

Przypomnijmy te założenia w nieco skróconej (i niepełnej) formie:

1.4.2 Model Normalnej Regresji Nieliniowej:

1° $\underset{(T\times 1)}{\mathbf{y}} = \underset{(T\times 1)}{\mathbf{g}}(\underset{(T\times p)}{\mathbf{X}}; \underset{(k\times 1)}{\boldsymbol{\beta}}) + \underset{(T\times 1)}{\boldsymbol{\varepsilon}}$, g grupuje znane funkcje klasy C^2 względem $\boldsymbol{\beta}$

2° $\mathbf{X}$ jest znaną macierzą nielosową lub $\mathbf{X} \perp\!\!\!\perp \boldsymbol{\varepsilon}$

3° $P\{ rz[\mathbf{A}(\boldsymbol{\beta})] = k \} = 1$ dla prawie wszystkich $\boldsymbol{\beta} \in \mathbf{B}$

4° $E(\boldsymbol{\varepsilon}) = \underset{(T\times 1)}{\mathbf{0}}$

5° $V(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I_T}$ $\qquad \sigma^2 > 0$

6° $\boldsymbol{\varepsilon}$ ma T-wymiarowy rozkład normalny

By otrzymać estymator wektora nieznanych parametrów $\boldsymbol{\beta}$ w powyższym układzie założeń, odwołamy się do jeszcze ogólniejszego modelu, jednak z pominięciem wyprowadzenia, tylko przez hasłowe powołanie. (dokładne wyprowadzenie w następnym rozdziale którego może wcale nie będzie)

1.4.3 Estymator Nieliniowej MNK

1.4.3.1 Geneza i własności

Ogólną metodą estymacji parametrycznych modeli statystycznych (takimi się zajmujemy), jest Metoda Największej Wiarygodności (MNW; Maximum Likelihood ML). Estymator uzyskany tą metodą w powyższym układzie założeń ma postać:

$$\hat{\boldsymbol{\beta}}_{NW} = \arg \min_{\boldsymbol{\beta}\in\mathbf{B}} [\mathbf{y} - \mathbf{g}(\mathbf{X};\boldsymbol{\beta})]' [\mathbf{y} - \mathbf{g}(\mathbf{X};\boldsymbol{\beta})]$$

Zauważmy, że jest to kryterium definiujące estymator MNW przy podanych założeniach; zastosowanie tego estymatora wymaga rozwiązania zagadnienia optymalizacyjnego (znalezienia minimum podanej funkcji), wobec tego dla uzyskania dającego się zastosować estymatora należy znaleźć sposób realizacji podanego kryterium. Należy tu podkreślić, że powyższy zapis zakłada istnienie odpowiedniego minimum (ono nie zawsze istnieje); przy tym założeniu (można je nawet określić jako założenie 7°) oraz założeniach 1°-6° powyższa równość jest prawdziwa. Prawa jej strona określa estymator tzw. nieliniowej MNK. Ponieważ do metody Największej Wiarygodności odwołujemy się tylko w celu wyprowadzenia estymatora nieliniowej MNK, będziemy oznaczali go jako $\hat{\boldsymbol{\beta}}$, rezygnując z subskryptu NW, za czym dodatkowo przemawia fakt, że

jest on uogólnieniem estymatora MNK (liniowej) tak oznaczanego, natomiast stosowanie i własności MNW będą tu tylko marginesowo poruszone i dokładniej opisane w innym miejscu.

Przy powyższych założeniach można dowieść, że dla $T\to\infty$:

$$\hat{\boldsymbol{\beta}} \overset{\bullet}{\sim} \mathbf{N}\left(\boldsymbol{\beta}, \sigma^2[\mathbf{A}(\boldsymbol{\beta})'\mathbf{A}(\boldsymbol{\beta})]^{-1}\right)$$

Powyższy wzór specyfikuje asymptotyczne (dla $T\to\infty$ czyli dla liczby obserwacji zmierzającej do nieskończoności) własności estymatora nieliniowej MNK. Mówi on, że $\hat{\boldsymbol{\beta}}$ ma dla $T\to\infty$ (kropka nad tyldą) rozkład normalny o wartości oczekiwanej równej wartości nieznanego parametru wektorowego i o macierzy kowariancji danej wzorem jak w nawiasie. Wobec tego można wnioskować tylko o asymptotycznej macierzy kowariancji $\hat{\boldsymbol{\beta}}$:

$$\mathbf{V}_{as}(\hat{\boldsymbol{\beta}}) = \sigma^2[\mathbf{A}(\boldsymbol{\beta})'\mathbf{A}(\boldsymbol{\beta})]^{-1}$$ (co wykorzystuje założenie 3°)

Zgodnym estymatorem asymptotycznej macierzy kowariancji $\hat{\boldsymbol{\beta}}$ jest:

$$\hat{\mathbf{V}}_{as}(\hat{\boldsymbol{\beta}}) = s^2[A(\hat{\boldsymbol{\beta}})'A(\hat{\boldsymbol{\beta}})]^{-1}$$; gdzie:

$$s^2 = \frac{1}{T-k}[y - g(X;\hat{\boldsymbol{\beta}})]'[y - g(X;\hat{\boldsymbol{\beta}})]$$

Estymator nieliniowej MNK jest zgodny $\hat{\boldsymbol{\beta}}$ i ma (opisany wyżej) asymptotyczny rozkład normalny. Nad cechami tymi warto zatrzymać się nieco dłużej.

Własności asymptotyczne zwane też własnościami dużej próby dotyczą hipotetycznej sytuacji dysponowania dowolnie dużą liczbą obserwacji ($T\to\infty$). W rzeczywistości mamy zawsze do czynienia z własnościami małej próby, tj. skończoną i często niewielką liczbą obserwacji. Posługując się estymatorem o poznanych własnościach asymptotycznych popełniamy pewien błąd, jego wielkość zależy od tego, jak szybko ze wzrostem liczby obserwacji estymator zaczyna przejawiać swoje asymptotyczne własności. Błąd jest w tej sytuacji nie do uniknięcia, można zakładać, że będzie on niewielki, można próbować ocenić trafność tego założenia w konkretnych okolicznościach za pomocą symulacji komputerowych, wreszcie przyjmować różne poprawki; jednak gdy odwołujemy się do własności asymptotycznych, musimy odnieść się też do problemu własności małopróbkowych. Dotychczas w modelach KMRL i KMNRL opisywane własności miały charakter dokładny (czyli bezpośrednio małopróbkowy), dzięki czemu problem oceny rozbieżności pomiędzy znanymi własnościami asymptotycznymi a własnościami w małej próbie nie występował. Odwoływanie się do własności dużej próby zwykle spowodowane jest faktem, że własności asymptotyczne zazwyczaj łatwiej określić niż dokładne; często dysponujemy jedynie asymptotyczną teorią: tak jest w przypadku MNRN, gdzie ogólne własności dokładne są nieokreślone.

Zgodność estymatora jest własnością asymptotyczną mówiącą, że *prawdopodobieństwo popełnienia dowolnego ustalonego błędu maleje do zera ze wzrostem liczby obserwacji do nieskończoności*. W pewnym sensie zgodność daje informację negatywną – zgodność **nie** gwarantuje popełnienia niewielkiego błędu w małej próbie – mówi tylko, że prawdopodobieństwo popełnienia dowolnego ustalonego błędu będzie maleć. Jednak brak zgodności stanowi bardzo ewidentny sygnał – jeśli nawet dla bardzo wielu obserwacji popełniamy nieznany błąd, to stosowanie takiego estymatora jest bardzo dyskusyjne. Niezgodność jest jednym z najcięższych zarzutów jakie można postawić estymatorowi.

Aby przybliżyć ich znaczenie i interpretację zgodności i asymptotycznej normalności $\hat{\boldsymbol{\beta}}$ objaśnijmy ich genezę. Kryterium które wyznacza estymator nieliniowej MNK otrzymaliśmy na gruncie Metody Największej Wiarygodności. Estymatory MNW mają przy dość ogólnych założeniach m.in. następujące własności: są zgodne i mają asymptotyczny rozkład normalny. Zarówno zgodność jak i asymptotyczna normalność są

własnościami dużej próby, własnościami asymptotycznymi, które mają pojawić się przy liczbie obserwacji zmierzającej do nieskończoności.

Model MNRN (przy założeniach 1° - 6° + 7°) spełnia warunki znacznie ogólniejszego modelu opartego na MNW. Wobec tego wszystkie własności MNW przechodzą na MNRN – stąd o estymatorze $\hat{\boldsymbol{\beta}}$ możemy powiedzieć, że jest zgodny i ma asymptotyczny rozkład normalny. Model KMNRL jako szczególny przypadek MNRN również zachowa wymienione własności MNW. W KMNRL estymator (liniowej) MNK $\hat{\boldsymbol{\beta}}$ jest zgodny i ma asymptotyczny rozkład normalny; jednakże założenia KMNRL są dużo mocniejsze od założeń MNW, dlatego w KMNRL można pokazać mocniejsze własności $\hat{\boldsymbol{\beta}}$ – znany jest jego rozkład dokładny, wobec czego nie ma potrzeby posługiwać się rozkładem asymptotycznym.

Widzimy tu, że przechodząc od modeli mniej ogólnych do modeli o większym zakresie stosowalności (od KMNRL do MNRN), przechodzimy jednocześnie do coraz słabszych własności: od własności dokładnych KMNRL do asymptotycznych MNRN; przejście od modeli liniowych do nieliniowych odbyło się za cenę rezygnacji z teorii dokładnej na rzecz asymptotycznej.

Widać to dobrze na przykładzie macierzy kowariancji estymatora MNK. W MNRN możemy szacować tylko przybliżoną, asymptotyczną macierz kowariancji, dlatego uzyskane z jej oszacowania oceny błędów średnich szacunku będą miały tylko przybliżony charakter, inaczej niż KMNRL.

1.4.3.2 Realizacja: Algorytm Gaussa-Newtona

Wyżej opisane zostały geneza i własności estymatora nieliniowej MNK, zostało podane kryterium jego otrzymania, pozostaje jeszcze znalezienie sposobu pozwalającego zrealizować to kryterium. Znalezienie minimum skomplikowanej funkcji jest trudne do zrealizowania wyłącznie analitycznymi metodami. Zwykle wykorzystywane będą do tego celu techniki numeryczne. Wiąże się to z komplikacjami dla badacza: opanować trzeba rozwiązywanie problemów numerycznych, których złożoność i czasochłonność rośnie w szybkim tempie ze wzrostem liczby parametrów. Uwzględnienie posiadanej wiedzy (poprzez odpowiednie dobranie punktów startowych lub wykorzystanie analitycznych postaci pochodnych cząstkowych) pozwala na bardziej efektywne wykorzystanie mocy obliczeniowej. Proponowanym tu sposobem jego realizacji kryterium definiującego $\hat{\boldsymbol{\beta}}$ (i uzyskania postaci estymatora dającej się zastosować) będzie algorytm Gaussa-Newtona. Jest to numeryczna procedura iteracyjna polegająca na powtarzalnym stosowaniu MNK dla liniowej aproksymacji wyjściowego modelu.

Jest to bardzo prosty przypadek problemu charakterystycznego dla współczesnych zastosowań ekonometrii. Skomplikowane procedury numeryczne (np. wielowymiarowe całkowanie numeryczne, symulacje Monte Carlo, problemy maksymalizacji funkcji-kryterium na wielowymiarowej przestrzeni parametrów) stanowią obecnie niezbędne ogniwo w łańcuchu prowadzącym do uzyskania oszacowań czy prognoz. Przedstawiony tu algorytm jest procedurą iteracyjną – oznacza to, że wymaga zadania wartości początkowych i wykonuje się potencjalnie nieskończoną ilość razy – musi zostać zatrzymany z zewnątrz. Odpowiadają temu zagadnienia ustalenia punktów startowych i kryterium zatrzymania. Każdy kolejny krok algorytmu sprowadza się do skorygowania ocen parametrów uzyskanych w poprzednim kroku. Na podstawie ocen kroku poprzedniego (j-1) w kroku bieżącym (j) ustalana jest poprawka, która po uwzględnieniu (dodaniu) daje wartość ocen parametrów kroku j. Jeśli kolejne wartości poprawek zmierzają do zera - wobec czego oceny parametrów stabilizują się (przy założonym poziomie dokładności) - osiągnięta została zbieżność algorytmu.

Jak już powiedziano, algorytm GN jest polega na iteracyjnym stosowaniu MNK dla liniowej aproksymacji wyjściowego modelu. Aby to pokazać, rozważmy rozwinięcie $\mathbf{g}$ w szereg Taylora (ucięty po pierwszych pochodnych) wokół pewnego *ustalonego* punktu $\boldsymbol{\beta}^{(j-1)}$ (górny indeks oznacza numer kroku):

$$\mathbf{y} = \mathbf{g}(\mathbf{X};\boldsymbol{\beta}) = \mathbf{g}(\mathbf{X};\boldsymbol{\beta}^{(j-1)}) + \frac{1}{1!}\mathbf{A}(\boldsymbol{\beta}^{(j-1)})(\boldsymbol{\beta}-\boldsymbol{\beta}^{(j-1)}) + \ldots \approx \mathbf{g}(\mathbf{X};\boldsymbol{\beta}^{(j-1)}) + \mathbf{A}(\boldsymbol{\beta}^{(j-1)})(\boldsymbol{\beta}-\boldsymbol{\beta}^{(j-1)})$$

wobec tego:

$$\mathbf{y} - \mathbf{g}(\mathbf{X};\boldsymbol{\beta}^{(j-1)}) \approx \mathbf{A}(\boldsymbol{\beta}^{(j-1)})(\boldsymbol{\beta} - \boldsymbol{\beta}^{(j-1)})$$

w powyższym równaniu po lewej stronie widać różnicę obserwacji na zmiennej objaśnianej i wartości teoretycznych odpowiadających ocenom w kroku j-1; po prawej macierz wartości pochodnych cząstkowych wektora funkcji **g** w punkcie odpowiadającym ocenom kroku j-1 pomnożoną przez różnicę między prawdziwymi wartościami parametrów a ich ocenami. Ten ostatni czynnik jest poprawką o którą należy skorygować oceny kroku j-1 aby otrzymać dokładne wartości parametrów. Wartość oszacowania tej właśnie poprawki jest rezultatem każdego kroku algorytmu. Powyższa równość jest oczywiście tylko przybliżona, można jednak na jej podstawie oszacować wartość poprawki. W tym celu przyjmijmy:

$$\mathbf{z}^{(j-1)} = \mathbf{y} - \mathbf{g}(\mathbf{X};\boldsymbol{\beta}^{(j-1)}); \qquad \boldsymbol{\alpha}^{(j-1)} = (\boldsymbol{\beta} - \boldsymbol{\beta}^{(j-1)})$$

gdzie $\boldsymbol{\alpha}^j$ jest wartością poprawki która ma być szacowana w kroku j. Zapiszmy poprzednie równanie:

$$\mathbf{z}^{(j-1)} \approx \mathbf{A}(\boldsymbol{\beta}^{(j-1)})\,\boldsymbol{\alpha}^{(j-1)} + \boldsymbol{\varepsilon}$$

i do tak zmodyfikowanego równania zastosujmy estymator liniowej MNK:

$$\hat{\boldsymbol{\alpha}}^{(j-1)} = [\mathbf{A}(\boldsymbol{\beta}^{(j-1)})'\,\mathbf{A}(\boldsymbol{\beta}^{(j-1)})]^{-1}\,\mathbf{A}(\boldsymbol{\beta}^{(j-1)})'\,\mathbf{z}^{(j-1)}$$

w ten sposób uzyskujemy w każdym kroku oszacowanie poprawki; wartości ocen parametrów kroku następnego $\boldsymbol{\beta}^{(j)}$ definiujemy jako:

$$\boldsymbol{\beta}^{(j)} = \boldsymbol{\beta}^{(j-1)} + \hat{\boldsymbol{\alpha}}^{(j-1)}$$

Powyższy opis określa algorytm Gaussa Newtona następująco:

$$\boldsymbol{\beta}^{(j)} = \boldsymbol{\beta}^{(j-1)} + [\mathbf{A}(\boldsymbol{\beta}^{(j-1)})'\,\mathbf{A}(\boldsymbol{\beta}^{(j-1)})]^{-1}\,\mathbf{A}(\boldsymbol{\beta}^{(j-1)})'(\mathbf{y} - \mathbf{g}(\mathbf{X};\boldsymbol{\beta}^{(j-1)}))$$

Wymaga on zadania punktów startowych $\boldsymbol{\beta}^{(0)}$ oraz zatrzymania z zewnątrz – musimy założyć, jak bliską zeru wartość poprawki uznamy za pozwalającą przyjąć, że zbieżność algorytmu została osiągnięta (poprawka jest równa zeru z zadaną dokładnością).

Algorytm GN ma zapewnić znalezienie minimum opisanej powyżej funkcji-kryterium, oznaczmy ją:

$$S(\boldsymbol{\beta}) = [\mathbf{y} - \mathbf{g}(\mathbf{X};\boldsymbol{\beta})]'\,[\mathbf{y} - \mathbf{g}(\mathbf{X};\boldsymbol{\beta})]$$

Aby pokazać związek biegu algorytmu z minimalizacją S($\boldsymbol{\beta}$), rozważmy warunek konieczny istnienia minimum tej funkcji (zakładamy, że jest C^2 względem β (a może to już wynika z założeń o **g**)):

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \underset{(k\times 1)}{\mathbf{0}} \Leftrightarrow \mathbf{A}(\boldsymbol{\beta})'(\mathbf{y} - \mathbf{g}(\mathbf{X};\boldsymbol{\beta})) = \underset{(k\times 1)}{\mathbf{0}}$$

Wyprowadzenie tego warunku następuje nietrudnym bezpośrednim rachunkiem. Wobec tego:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}\in\mathbf{B}} S(\boldsymbol{\beta}) \Rightarrow \mathbf{A}(\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{g}(\mathbf{X};\hat{\boldsymbol{\beta}})) = \underset{(k\times 1)}{\mathbf{0}}$$

Zauważmy, że jego prawa strona określa sytuację, gdy poprawka $\boldsymbol{\beta}^{(j)}$ jest dokładnie równa zeru, co zapewnione jest przez zbieżność algorytmu z dowolną założoną dokładnością. Widać, że osiągnięcie zbieżności algorytmu jest równoznaczne ze spełnieniem koniecznego warunku znalezienia ekstremum S($\boldsymbol{\beta}$). Spełnienie pozostałych warunków i odrzucenie przypadku maksimum nie będzie formalnie rozważane.

Powyżej pokazano, że KMNRL spełnia założenia MNRN ($\mathbf{A}(\boldsymbol{\beta}) = \mathbf{X}$ oraz $\mathbf{g} = \mathbf{X}\boldsymbol{\beta}$). zważmy teraz, jaka postać estymatora $\hat{\boldsymbol{\beta}}$ wynika z przyjęcia w MNRN warunków KMNRL i zastosowania algorytmu GN; przyjmijmy dowolne punkty startowe $\boldsymbol{\beta}^{(0)}$:

$$\boldsymbol{\beta}^{(1)} = \boldsymbol{\beta}^{(0)} + [\mathbf{A}(\boldsymbol{\beta}^{(0)})'\mathbf{A}(\boldsymbol{\beta}^{(0)})]^{-1}\mathbf{A}(\boldsymbol{\beta}^{(0)})'(\mathbf{y} - \mathbf{g}(\mathbf{X};\boldsymbol{\beta}^{(0)})) = \boldsymbol{\beta}^{(0)} + [\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{(0)}) =$$
$$= \boldsymbol{\beta}^{(0)} + [\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{y} - [\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta}^{(0)} = \boldsymbol{\beta}^{(0)} + [\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{y} - \mathbf{I}\boldsymbol{\beta}^{(0)} = [\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{y} = \hat{\boldsymbol{\beta}}$$

W szczególnym przypadku MNRN, kiedy spełnione są założenia KMNRL, algorytm Gaussa-Newtona w pierwszym kroku daje estymator liniowej MNK, co ostatecznie pokazuje, że KMNRL jest szczególnym przypadkiem MNRN. Przy tym nie wprost pokazana została następująca własność estymatora liniowej MNK: jako szczególny przypadek estymatora $\hat{\boldsymbol{\beta}}$, minimalizuje od funkcję $S(\boldsymbol{\beta})$, która jest sumą kwadratów reszt MNK:

$$\hat{\boldsymbol{\beta}} = [\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{y} = \arg\min_{\boldsymbol{\beta}\in\mathbf{B}} S(\boldsymbol{\beta}) = \arg\min_{\boldsymbol{\beta}\in\mathbf{B}} [(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})]$$

co opisuje „operacyjne” własności analitycznego wzoru i dopełnia opis KMNRL.

1.4.4 Wnioskowanie statystyczne w MNRN

Faktem zasadniczym dla własności wnioskowania statystycznego w MNRN jest asymptotyczny charakter teorii leżącej u podstaw tego modelu. W przeciwieństwie do KMNRL, własności małopróbkowe w MNRN nie są określone. Wnioskowanie odwołujące się do własności asymptotycznych ma charakter wyłącznie przybliżony – obarczone jest błędem wynikającym z rozbieżności między małopróbkowymi danymi a teorią dużej próby; teoria nie gwarantuje, że błąd ten będzie niewielki, mówi tylko, że ze wzrostem liczby obserwacji do nieskończoności prawdopodobieństwo popełnienia dowolnego ustalonego błędu będzie maleć do zera.

W konkretnych przypadkach własności małopróbkowe estymatora nieliniowej MNK można badać symulacyjnie. Badania symulacyjne mogą sugerować przyjęcie pewnych poprawek „na małą próbę”, jednak należy podkreślić, że tego typu postępowanie, aczkolwiek stosowane, ma charakter poprawek ad hoc, nie wynikających z teorii.

Przybliżony charakter wnioskowania znajduje swoje odbicie w fakcie, że szacowaniu podlegać może tylko asymptotyczna macierz kowariancji estymatora $\hat{\boldsymbol{\beta}}$. Nawet gdyby była ona *znana* (nie trzeba byłoby stosować estymatora), nie dawałaby dokładnej informacji o rozproszeniu $\hat{\boldsymbol{\beta}}$, a tylko przybliżoną, gdyż odpowiada sytuacji $T\rightarrow\infty$. Pierwiastki z elementów przekątniowych oszacowania tej macierzy to przybliżone błędy średnie szacunku.

Podobnie estymacja przedziałowa i testowanie hipotez mają wyłącznie charakter przybliżony. Do budowy przedziałów ufności dla jednego parametru regresji lub testowania hipotez o pojedynczym parametrze możemy wykorzystać procedury z KMNRL. Asymptotyczny charakter teorii powoduje, że rozkład t-Studenta przechodzi w rozkład normalny (rozpatrujemy T-k stopni swobody gdzie $T\rightarrow\infty$), wobec czego odpowiednie wartości krytyczne pobierać należałoby z danych dotyczących rozkładu normalnego. Jednak takie postępowanie, choć mające oparcie w teorii (asymptotycznej), nie odnosi się do problemu małej próby. Z drugiej strony można przyjmować ad hoc ateoretyczną poprawkę, której zasadność znajduje potwierdzenie w symulacjach. Taką poprawką jest stosowanie rozkładu t-Studenta i przyjęcie za T rzeczywistej liczby obserwacji, co odpowiada dokładnie procedurze KMNRL. Symulacje sugerują zasadność takiego postępowania (Ratkovsky), jednak nie ma ono pełnego uzasadnienia w teorii. Przedziały ufności otrzymane drugą metodą będą odpowiednio dłuższe, odpowiednio do tego różnić się będą wyniki testów. W szczególności oba podejścia prowadzić mogą do różnych konkluzji, co uzasadnia określenie „przybliżony charakter wnioskowania”. Przybliżony charakter przedziałów ufności można też zinterpretować następująco: uzyskany przedział jest realizacją „mniej więcej α-procentowego przedziałem ufności”.

Podsumowując, model MNRN stanowi uogólnienie deterministycznej części KMNRL, dopuszczenie nieliniowości podnosi użyteczność modelu, jednak ceną za te nowe możliwości jest przejście od własności dokładnych do asymptotycznych. Stosunek modeli KMNRL-MNRN stanowi prostą i podstawową ilustrację rozwoju modeli ekonometrycznych.

Geneza MNRN pokazuje, jak na całość modelu składają się różne „warstwy":

1. Leżący u podstawy jednorównaniowy model w kontekście Metody Największej Wiarygodności, podstawowej metody szacowania modeli parametrycznych, pozwalającej uzyskać estymatory zgodne i asymptotycznie normalne.
2. Założenia deterministyczne i stochastyczne MNRN (1° - 6° + 7°) w ramach których estymator Metody Największej Wiarygodności istnieje, posiada określone własności (daną wzorem asymptotyczną macierz kowariancji) oraz daje się przedstawić jako argument minimalizujący pewną funkcję-kryterium.
3. Algorytm Gaussa-Markowa pozwalający zrealizować teoretyczne kryterium i zastosować estymator nieliniowej MNK.

Dalsza generalizacja będzie przebiegać w dwóch kierunkach: głębiej rozważony zostanie zasygnalizowany już problem losowości zmiennych objaśniających, oraz uogólniona zostanie struktura stochastyczna MNRN (i KMNRL).

1.5 Regresja z losowymi zmiennymi objaśniającymi

Przy przedstawianiu KMRL zasygnalizowany został problem uznania losowości zmiennych objaśniających. Przyjęte wtedy założenie o nielosowości tych zmiennych pochodzi – jak i cały model – z doświadczalnictwa przyrodniczego, i odpowiada opisywaniu eksperymentu kontrolowanego. W wypadku procesów ekonomicznych założenie kontrolowanego eksperymentu jest nierealistyczne. Wartości zmiennych objaśniających powstają zazwyczaj w gospodarce (choć poza modelowanym jej wycinkiem), toteż przyjęcie, że przybierają podlegające kontroli stałe wartości nie wydaje się realistyczne. Dopuszczenie losowości zmiennych objaśniających zmierza do uzgodnienia struktury modelu z rozumieniem modelowanego procesu. Kolejnym krokiem w tym kierunku mogłoby być uznanie specjalnego, historycznego charakteru danych ekonomicznych – przyjęcie, że dysponujemy pojedynczą realizacją danych i rozpatrywanie możliwości wielokrotnej realizacji jest bezpodstawne – odpowiada to przejściu od klasycznego, częstościowego rozumienia prawdopodobieństwa do interpretacji bayesowskiej.

Założenie o nielosowości zmiennych objaśniających, choć nierealistyczne, jest używane, ponieważ przy jego przyjęciu można przedstawić podstawowe modele i procedury postępowania. Przy uchylaniu tego założenia ich własności mogą pozostać podobne, zmienić się nieznacznie lub zmienić się całkowicie. Przejaw tego wystąpił już przy przedstawianiu założeń MNRN, gdzie dopuszczona była losowość macierzy $\mathbf{X}$ pod warunkiem jej stochastycznej niezależności od wektora $\boldsymbol{\varepsilon}$. Wskazuje to, że konsekwencje losowości macierzy $\mathbf{X}$ będą zależne od dodatkowych założeń opisujących tę losowość. Dotychczas przedstawiane były własności estymatora MNK i związanych z nim procedur wnioskowania w różnych układach założeń. Obecnie rozważony zostanie wpływ wywierany te własności przez dopuszczenie losowości macierzy $\mathbf{X}$. W tym celu wyróżnione zostaną trzy przypadki:

I. Zmienne objaśniające są losowe, ale niezależne stochastycznie od składników losowych.

II. Zmienne objaśniające są losowe, zależne od uprzednich składników losowych, ale niezależne stochastycznie od bieżących i przyszłych składników losowych. (dotyczy danych ze sztywnym porządkiem obserwacji.)

III. Zmienne objaśniające są losowe i zależne od bieżącego składnika losowego.

Własności estymatora MNK i procedur wnioskowania statystycznego są inne w każdym z tych przypadków. By je określić, należy przypisać rozważany model do jednej z powyższych klas; dla zilustrowania kryteriów przypisania poniżej zarysowanie zostaną proste przykłady modeli odpowiadających wymienionym przypadkom.

1.5.1 Zmienne objaśniające losowe i niezależne od składników losowych

Ad. I. Przykładem sytuacji, gdy zmienne objaśniające są losowe, jednak niezależne od składników losowych $\boldsymbol{\varepsilon}$, jest model estymacji funkcji produkcji na podstawie danych z firm maksymalizujących zysk, przedstawiony przez Arnolda Zellnera, Jana Kmentę i Jacques'a Drčze'a.

1.5.1.1 Przykład: model Zellner Kmenta Dreze.

Rozpatrywana jest w nim funkcja produkcji Cobba i Douglasa, która po zlogarytmowaniu ma postać liniową względem parametrów. Obserwacje odpowiadały poszczególnym firmom, natomiast składnik losowy $\boldsymbol{\varepsilon}$ zlogarytmowanej funkcji produkcji odzwierciedlać miał oddziaływanie „sił natury" – systematyczny błąd popełniany przez wszystkie firmy na skutek oddziaływań będących poza ich kontrolą. Zmienne objaśniające w tym modelu – wielkości nakładów czynników produkcji tj. pracy i kapitału, nie były jednak „z góry zadane", lecz ustalane przez firmy. Dla określenia losowości tych zmiennych (co jest tutaj zasadniczym przedmiotem zainteresowania) kluczowe jest opisanie własności procesu który generuje ich wartości – czyli opisany musi zostać sposób ustalania przez firmy wielkości nakładów czynników produkcji. Założono, że firmy posiadają doskonałą znajomość cen (odpowiednio produktu: p, oraz nakładów: w i r), oraz, że firmy podejmując decyzję o zaangażowaniu czynników produkcji kierują się maksymalizacją zysku. Jednak zysk zależy od wielkości produkcji, na którą wpływ ma oddziaływanie „sił natury" – systematyczny błąd $\boldsymbol{\varepsilon}$, wobec czego jako wielkość losowa zysk nie może podlegać maksymalizacji. Firmy maksymalizują więc „spodziewany zysk" – wartość oczekiwaną zysku. Ponieważ wartość oczekiwana zysku zależy już od znanej wartości oczekiwanej $\boldsymbol{\varepsilon}$ (równej $\mathbf{0}$), oraz od znanych wielkości (parametrów technologii oraz cen), może podlegać maksymalizacji, wobec czego wielkości nakładów to wartości rozwiązujące zagadnienie optymalizacyjne. Gdyby firmy doskonale maksymalizowały „spodziewany zysk", wielkość zależną od znanych parametrów, ustalone przez nie wielkości nakładów pracy i kapitału byłyby nielosowe – byłyby znanymi funkcjami znanych zmiennych. Autorzy modelu przyjęli jednak, że firmy dokonują „niedoskonałej optymalizacji" – na ich decyzje mają wpływ indywidualne „błędy w zarządzaniu" – odchylenia charakterystyczne dla każdej firmy, które powodują, że angażowane nakłady czynników produkcji są ustalane na poziomie odbiegającym od poziomu idealnego – różnią się od niego o błąd $\mathbf{u}$. Wobec tego zmienne objaśniające, jako funkcje $\mathbf{u}$, są zmiennymi losowymi. Dla klasyfikacji modelu istotny jest rodzaj zależności pomiędzy $\boldsymbol{\varepsilon}$ a $\mathbf{u}$. Pierwszy z nich opisuje błąd systematyczny, wynikający ze wspólnego dla wszystkich firm oddziaływania „sił natury". Drugi natomiast odpowiada „błędom indywidualnym", wynikającym z odmiennych charakterystyk poszczególnych przedsiębiorstw. Wobec tego, że założone źródła losowości $\boldsymbol{\varepsilon}$ i $\mathbf{u}$ są odmienne, autorzy przyjęli niezależność stochastyczną zmiennych objaśniających (będących funkcjami $\mathbf{u}$) i składników losowych regresji ($\boldsymbol{\varepsilon}$) opisującej funkcję produkcji. W ten sposób zgodnie z intuicją przedstawiono zmienne objaśniające jako wielkości losowe, jednak na mocy teorii niedoskonałej optymalizacji ustalono, że źródła ich losowości są odmienne od źródeł losowości $\boldsymbol{\varepsilon}$, co pozwala przyjąć ich stochastyczną niezależność i odpowiada przypadkowi 1.

Przyjęcie liniowej postaci zależności deterministycznej i stochastycznej niezależności składników losowych i zmiennych objaśniających prowadzi do układu założeń regresji liniowej z losowymi zmiennymi objaśniającymi. Model ten jest istotnie różny od KMRL, własności estymatora $\hat{\boldsymbol{\beta}}$ tam określone nie zachowują ważności ze względu na nową jego postać: $\hat{\boldsymbol{\beta}}$ nie jest już znaną funkcją, ponieważ macierz $\mathbf{X}$ jest losowa.

1.5.1.2 Model Regresji Liniowej z Losowymi Zmiennymi Objaśniającymi

1°	$\underset{(T\times 1)}{\mathbf{y}} = \underset{(T\times k)}{\mathbf{X}}\ \underset{(k\times 1)}{\boldsymbol{\beta}} + \underset{(T\times 1)}{\boldsymbol{\varepsilon}}$
2A°	$\mathbf{X} \perp\!\!\!\perp \boldsymbol{\varepsilon}$
3°	$P\{\ rz[\mathbf{A}(\boldsymbol{\beta})] = k\ \} = 1$
4°	$E(\boldsymbol{\varepsilon}) = \underset{(T\times 1)}{\mathbf{0}}$
5°	$V(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I_T}$ $\qquad \sigma^2>0$

1.5.1.3 Własności estymatora MNK

Własności estymatora MNK, $\hat{\boldsymbol{\beta}}$ w powyższym układzie założeń wymagają określenia na nowo (w porównaniu z KMRL), z uwzględnieniem losowości **X**. Dla zbadania nieobciążoności $\hat{\boldsymbol{\beta}}$ musimy określić jego wartość oczekiwaną; przydatne tu będzie następujące twierdzenie:

Tw. Wartość oczekiwana iloczynu niezależnych zmiennych losowych jest równa iloczynowi ich wartości oczekiwanych (jeśli istnieją).

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \overset{1^\circ}{=} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}$$

$$E(\hat{\boldsymbol{\beta}}) = E[\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}] = \boldsymbol{\beta} + E\,[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}],$$

co na mocy powyższego twierdzenia i założenia $2A^\circ$ daje:

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} + E\,[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] * E[\boldsymbol{\varepsilon}] \overset{4^\circ}{=} \boldsymbol{\beta} + \mathbf{0};$$

co kończy dowód nieobciążoności $\hat{\boldsymbol{\beta}}$ (ponownie nie korzysta on z 5°) – estymator MNK jest nieobciążony mimo losowości zmiennych objaśniających (przy $2A^\circ$).

Ponadto:

$$s^2 = \frac{1}{T-k}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}), \qquad \text{i} \qquad E(s^2) = \sigma^2$$

Nieobciążonym estymatorem macierzy kowariancji estymatora MNK jest:

$$\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}) = s^2(\mathbf{X}'\mathbf{X})^{-1}, \qquad E\left(\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}})\right) = \mathbf{V}(\hat{\boldsymbol{\beta}}) = \sigma^2 E(\mathbf{X}'\mathbf{X})^{-1}$$

Wobec tego praktyczny aspekt procedur stosowanych w KMRL pozostaje niezmieniony, błędy średnie szacunku $\hat{\boldsymbol{\beta}}$ zachowują ważność, oblicza się je korzystając z tych samych wzorów; różnice dotyczą teoretycznych własności $\hat{\boldsymbol{\beta}}$ - nie jest on już znaną funkcją, przestaje być liniowy co powoduje, że nie obowiązuje twierdzenie Gaussa i Markowa. Oznacza to, że $\hat{\boldsymbol{\beta}}$ traci optymalne własności – przestaje być estymatorem liniowym, traci więc własność efektywności. Opisane własności mają przy tym charakter małopróbkowy – przy powyższych założeniach.

Dodanie do omawianego schematu założenia o wielowymiarowym rozkładzie normalnym dla **ε:**

6°	**ε** ma T-wymiarowy rozkład normalny

daje możliwości testowania hipotez i estymacji przedziałowej analogicznie jak w KMNRL, jednak model ten NIE JEST tożsamy z modelem KMNRL – jest jego generalizacją; przy założeniach 1° - 6°:

$$\frac{\hat{\beta}_i - \beta_i}{D(\hat{\beta}_i)} \sim St(T-k)$$

Wobec czego procedury testowania i estymacji przedziałowej analogiczne jak w KMNRL zachowują ważność. Przyjęcie losowości zmiennych objaśniających w postaci 2A° - czyli przypadek 1. prowadzi do sytuacji, gdzie wszystkie sposoby postępowania zachowują swoją postać, zmieniają się niektóre ich teoretyczne własności, co oczywiście dotyczy przypadku opisanego modelu (Zellner, Kmenta, Dreze).

1.5.2 Zmienne objaśniające losowe i zależne tylko od uprzednich składników losowych

Ad. II. Przypadek kolejny dotyczy sytuacji, gdy zmienne objaśniające są losowe, zależne od uprzednich składników losowych, ale niezależne stochastycznie od bieżących i przyszłych składników losowych. Może on zachodzić tylko wtedy, gdy sensowne jest określenie „uprzedni" czy „przyszły", tj. w wypadku danych o sztywnym porządku obserwacji, co w praktyce dotyczy szeregów czasowych. Podstawowym narzędziem statystycznym pozwalającym modelować ekonomiczne szeregi czasowe jest teoria procesów stochastycznych. Przykładowo pozwala ona skonstruować model, w którym bieżąca wartość zmiennej jest objaśniana przez jej przeszłe wartości – wobec czego uprzednie realizacje zmiennej są wartościami zmiennych objaśniających dla części deterministycznej jej bieżącej wartości.

1.5.2.1 Przykład: proces AR(m) dla zmiennej objaśnianej z niezależnym składnikiem losowym

Ilustruje to proces autoregresyjny rzędu m dla zmiennej objaśniającej z nałożonym niezależnym zakłóceniem losowym:

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \ldots + \beta_m y_{t-m} + \beta_0 + \varepsilon_t, \quad \text{gdzie } \varepsilon_t \perp\!\!\!\perp \varepsilon_s \text{ dla } t \neq s;$$

Bieżąca wartość y jest wobec tego funkcją całej przeszłości, na którą nałożone jest swoiste, niezależne zakłócenie losowe (niezależne od przeszłości zakłócenie, swoiste dla danego okresu nazywane jest w ekonometrii „innowacją"). Takie określenie procesu generującego wartości y_t zapewnia spełnienie założeń punktu II, ponieważ y_t jest na mocy konstrukcji funkcją wszystkich poprzednich wartości y, a wobec tego i wszystkich poprzednich wartości ε (widać to po dokonaniu kolejnych podstawień za y_{t-1} itd.), jednak jest niezależny od bieżącego zakłócenia losowego ε_t i od zakłóceń przyszłych.

W takim przypadku do oszacowania (m+1) parametrów β można zastosować MNK dostosowując jego postać do rozpatrywanego zagadnienia w następujący sposób:

Szereg czasowy N wartości zmiennej y numerujemy liczbami całkowitymi od (1-m) przez 0 do (N-m), po czym T = N – m ostatnich wartości tworzy wektor **y**. Wydzielenie m pierwszych wartości (tzw. warunków początkowych) uszczupla pulę obserwacji, jest jednak konieczne dla dostarczenia wartości zmiennych objaśniających dla obserwacji o numerze 1. Warunki początkowe, czyli (y_{1-m} ... y_0) są odmiennie traktowane od innych obserwacji – wyniki estymacji są warunkowane względem ich wartości. Odpowiednio do postaci zależności definiującej y_t tworzona jest też macierz **X**:

$$\underset{(T\times 1)}{\mathbf{y}} = \begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ y_T \end{bmatrix}, \qquad \underset{(T\times(m+1))}{\mathbf{X}} = \begin{bmatrix} \mathit{y_o} & \mathit{y_{-1}} & \cdots & \mathit{y_{1-m}} & 1 \\ y_1 & \mathit{y_o} & \cdots & \mathit{y_{2-m}} & 1 \\ \cdots & \cdots & \cdots & \cdots & 1 \\ y_{T-1} & y_{T-2} & \cdots & y_{T-m} & 1 \end{bmatrix}$$

(kursywą zaznaczono elementy wektora wartości początkowych)

Estymator MNK ma tu postać : $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. Jego własności w tym przykładzie i ogólnie w przypadku II są następujące:

1.5.2.2 Własności estymatora MNK

Estymator MNK jest zgodny, przy czym:

$$\hat{\mathbf{V}}_{as}(\hat{\boldsymbol{\beta}}) = s^2(\mathbf{X}'\mathbf{X})^{-1}$$

jest zgodnym estymatorem asymptotycznej macierzy kowariancji $\hat{\boldsymbol{\beta}}$. Oznacza to, że stosując estymator MNK w przypadku, gdy zmienne objaśniające są losowe i zależne od przeszłych składników losowych a niezależne od bieżącego i przyszłych składników losowych, nie popełnia się błędu systematycznego. Własności estymatora liniowej MNK są asymptotyczne, w małej próbie można traktować je jako przybliżone. Dodanie założenia:

6°	$\boldsymbol{\varepsilon}$ ma T-wymiarowy rozkład normalny

pozwala stosować procedury estymacji przedziałowej oraz testowanie hipotez – wnioskowanie ma charakter analogiczny jak w MNRN, mimo liniowej postaci modelu.

1.5.3 Zmienne objaśniające losowe i zależne od bieżących składników losowych

Ad. III. Przypadek III dotyczy sytuacji, gdy zmienne objaśniające są zależne od bieżącego składnika losowego.

1.5.3.1 Własności estymatora MNK

W tym wypadku **estymator MNK nie jest zgodny**. Jego stosowanie nawet jeśli liczba obserwacji zmierza do nieskończoności obarczone jest nieznanym, systematycznym błędem. Własności w małej próbie są nieokreślone – można je wyłącznie badać symulacyjnie w konkretnych przypadkach.

Wyklucza to stosowanie estymatora MNK w modelach odpowiadających omawianej sytuacji. Konieczne jest odwołanie się do innych metod estymacji.

Poniżej przedstawione zostaną dwa przykłady sytuacji zależności zmiennych objaśniających od bieżących składników losowych: model autoregresyjny z zależnymi składnikami losowymi, oraz model wielorównaniowy o równaniach łącznie współzależnych.

1.5.3.2 Przykład: Model autoregresyjny z zależnymi składnikami losowymi.

Model ten jest rozszerzeniem przedstawionego w powyższym punkcie modelu autoregresyjnego dla zmiennej objaśnianej; odmiennie niż poprzednio zakłada się, że składniki losowe ε nie są innowacjami – czyli nie odpowiadają swoistym, niezależnym zakłóceniom danego okresu, lecz zależą od swojej przeszłości – mają „pamięć". Modele tego typu są szczególnie istotne w modelowaniu szeregów czasowych – pamięć składnika losowego może odzwierciedlać specyficzną dynamikę procesów zwłaszcza makroekonomicznych.

Składnik losowy opisany procesem AR(1) ma następującą postać:

AR(1): $\varepsilon_t = \varphi\varepsilon_{t-1} + \xi_t$, gdzie $\xi_t \perp\!\!\!\perp \xi_s$ dla $t \neq s$ oraz $\varphi \in (-1,1)$;

Jest to zmodyfikowana wersja procesu AR(1) dla ε z parametrem φ. Parametr – wyraz wolny nie występuje (jest równy zero) ze względu na wymaganie zerowej wartości oczekiwanej składnika losowego; ograniczenie wartości φ jest podyktowane koniecznością zapewnienia stacjonarności procesu. Proces niestacjonarny ma zupełnie inne własności, co nie jest tu rozpatrywane. Tak określony składnik losowy „pamięta" frakcję φ poprzedniego zakłócenia, na co nałożona jest innowacja ξ. Ponieważ wcześniejsze zakłócenia podobnie zależą od swoich poprzedników, ε_t zależy od całej swojej przeszłości tj. od wszystkich poprzednich realizacji ξ oraz od warunków początkowych ε_0. Analogicznie można rozważać rząd procesu większy niż 1.

Podobne zjawisko występuje w wypadku składnika losowego opisanego procesem średnich ruchomych MA(m), tutaj MA(1):

MA(1): $\varepsilon_t = \eta\xi_{t-1} + \xi_t$, gdzie $\xi_t \perp\!\!\!\perp \xi_s$ dla $t \neq s$

Jeśli przyjmiemy

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \ldots + \beta_m y_{t-m} + \beta_0 + \varepsilon_t,$$

a składnik losowy ε jest opisany którymś z powyższych procesów, to widać, że zmienna objaśniająca y_{t-1} zależy od ε_{t-1}, (poprzez równanie objaśniające powstawanie y_{t-1}), a poprzez ε_{t-1} także od ξ_{t-1} – innowacji okresu t-1. Jednak ε_t zależy również od ξ_{t-1} na mocy konstrukcji. Zmienna objaśniająca i bieżący składnik losowy są zależne od ξ_{t-1}, wobec czego mamy do czynienia z przypadkiem III, estymator MNK jest niezgodny i problem ten wymaga stosowania innych technik estymacji. W związku z tym przy rozważaniu stosowania MNK dla badania zjawisk dynamicznych opisanych procesami autoregresyjnymi powstaje konieczność testowania składnika losowego pod kątem ewentualnej jego autokorelacji. Tego rodzaju test zostanie omówiony poniżej łącznie z metodą estymacji pozwalającą uwzględnić to zjawisko. Kolejny przykład sytuacji skutkującej niezgodnością estymatora MNK dotyczy modeli wielorównaniowych.

1.5.3.3 Przykład: Modele wielorównaniowe o równaniach łącznie współzależnych

Problem zależności zmiennej objaśniającej i bieżącego składnika losowego może też wystąpić w niezmiernie prostych nawet modelach wielorównaniowych. Ma on miejsce w układach charakteryzujących się występowaniem sprzężeń zwrotnych.

Rozważany będzie dla ilustracji prosty keynesowski z ducha model gospodarki – o bardzo uproszczonej postaci. Opisuje on współkształtowanie się dwóch wielkości: konsumpcji C i dochodu Y przy ustalonej z góry wielkości inwestycji I:

$$C_t = \alpha + \beta Y_t + \varepsilon_t$$
$$Y_t = C_t + I_t$$

Drugie równanie to zależność bilansowa – pozbawiona składnika losowego. Badanym zagadnieniem jest estymacja parametru β opisującego krańcową skłonność do konsumpcji. „Statystycznie naiwne" podejście mogłoby sugerować oszacowanie parametrów pierwszego równania za pomocą MNK, jednak:

$$C_t = \alpha + \beta(C_t + I_t) + \varepsilon_t, \quad \text{więc:}$$
$$C_t(1 - \beta) = \alpha + \beta I_t + \varepsilon_t$$

$$C_t = \alpha/(1 - \beta) + [\beta(1 - \beta)]\, I_t + \varepsilon_t$$
$$Y_t = \alpha/(1 - \beta) + [1 + \beta(1 - \beta)]\, I_t + 1/(1 - \beta)\, \varepsilon_t$$

jest to podejście abstrahujące całkowicie od ekonomicznego sensu tego modelu: opisuje on bowiem łączne kształtowanie konsumpcji i dochodu; powstanie wielkości warunkujących równowagę badanego systemu. Ostatnie dwa równania pokazują, że obie badane zmienne są silnie, funkcyjnie zależne od bieżącego składnika losowego, co powoduje niezgodność estymatora MNK.

Problematyka modeli wielorównaniowych jest tu wyłącznie sygnalizowana w kontekście własności estymatora MNK w przypadku losowych zmiennych objaśniających; dokładniej jest ona przedstawiona w rozdziale 6 „Wprowadzenia do Ekonometrii w przykładach i zadaniach" autorstwa Profesora.