

2. Metoda Najmniejszych Kwadratów

W tej rozdziale zajmiemy się modelami jednorównanowymi. Do estymacji tych modeli możemy zastosować szereg metod. Jedną z najszerzej stosowanych jest **Metoda Najmniejszych Kwadratów** (*MNK*). Estymatory *MNK* są to takie funkcje danych, które minimalizują sumę kwadratów wartości teoretycznych wynikających z modelu z wartościami rzeczywiście zaobserwowanymi.

Przy estymacji ważne jest ustalenie założeń potrzebnych, by estymatory posiadały pożądane własności. Te pożądane własności, to zwykle nieobciążoność i minimalna wariancja. W dalszej części tego rozdziału pokażemy, że dla pewnych założeń takie własności posiadają właśnie estymatory *MNK*. Pokażemy także, że dla nieco słabszych założeń estymatory *MNK* są obciążone w małych próbach ale asymptotycznie ich obciążenie maleje do zera. Omówimy także sytuacje, w których nie należy stosować estymatorów *MNK*, ponieważ są one nieefektywne lub asymptotycznie obciążone.

W trakcie wnioskowania statystycznego testuje się różnorodne hipotezy dotyczące parametrów oraz buduje się przedziały ufności dla uzyskanych estymatorów. Procedury takie można przeprowadzić pod warunkiem znajomości rozkładów testów i estymatorów. Z tego powodu dużo czasu poświęcimy analizie wyprowadzaniu tych rozkładów. Omówimy problem testowania prostych i złożonych hipotez statystycznych oraz testy diagnostyczne służące do badania, czy model spełnia założenia konieczne do tego, by estymatory miały porządane właściwości.

2.1. Model z wieloma zmiennymi

Wyprowadzenie wzorów na estymatory *MNK* dla przypadku wielu zmiennych najłatwiej przeprowadzić stosując zapis macierzowy. Model liniowy można wyrazić następującym wzorem:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

$$\text{gdzie } \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} x_{11} & \cdots & x_{K1} \\ \vdots & & \\ x_{1n} & \cdots & x_{Kn} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_K \end{bmatrix} \text{ i } \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}, \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$$

estymator parametru $\boldsymbol{\beta}$ oznaczamy przez \mathbf{b} . Reszty z modelu definiujemy jako

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{b}$$

Najprostszymi przypadkami modeli liniowych jest model, w którym występuje tylko jedna zmienna objaśniająca (poza stałą)

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i \quad (2.1)$$

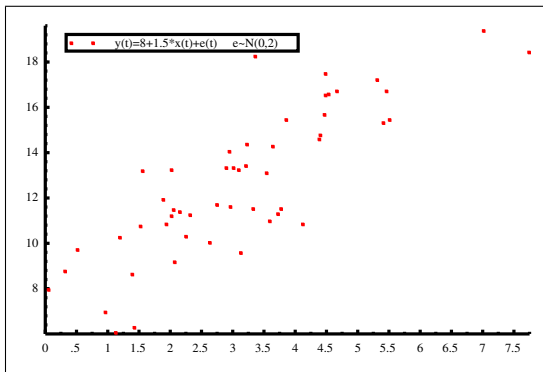
tę jedną zmienną może być np czas i wtedy mówimy o modelu trendu liniowego, który zapisać można jako

$$y_t = \beta_1 + \beta_2 t + \varepsilon_t \quad (2.2)$$

Idea Metody Najmniejszych Kwadratów polega na tym, że mając pewien zbiór danych zawierających y_i i \mathbf{X}_i dla $i = 1, \dots, n$ znajdujemy taki estymator \mathbf{b} parametru $\boldsymbol{\beta}$, który minimalizuje sumę kwadratów reszt. Przykładowo dla modelu (2.1) i zbioru danych zilustrowanym na rysunku 1 znalezione za pomocą *MNK* estymatory parametrów β_1 i β_2 wynoszą $b_1 = 7,5331$ i $b_2 = 1.6406$. Dopasowanie linii regresji ilustruje rysunek 2. Ważne jest przy tym rozróżnienie między parametrami a estymatorami (oszacowaniami) parametrów. Dane z rysunku 1 zostały sztucznie stworzone za pomocą generatora liczb losowych dla $\beta_1 = 8$ i $\beta_2 = 1,5$. Widzimy więc, że ogólnie prawdziwe parametry i ich estymatory są różniąc się od siebie. Dopasowanie za pomocą *MNK* trendu zilustrowane jest na rysunku 3

Linie regresji zilustrowane za pomocą rysunków 2 i 3 są szczególnymi przypadkami hiperpłaszczyzny regresji, którą w ogólnym wielowymiarowym przypadku, dopasowujemy za pomocą *MNK*. Jak już wspomnieliśmy estymator ten uzyskujemy minimalizując funkcję celu $S(\mathbf{b})$ równą sumie kwadratów reszt

$$S(\mathbf{b}) = \mathbf{e}'\mathbf{e} = (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b}) = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\mathbf{b} - \mathbf{b}'\mathbf{X}'\mathbf{y} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{y}'\mathbf{y} - 2\mathbf{b}'\mathbf{X}'\mathbf{y} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}$$



Rysunek 1: Obserwacje (x_i, y_i)

przy czym ostatnia równość wynika z faktu, że $\mathbf{y}'\mathbf{X}\mathbf{b}$ i $\mathbf{b}'\mathbf{X}'\mathbf{y}$ są skalarami a dla skalarów transpozycja nic nie zmienia. Warunki pierwszego rzędu dla

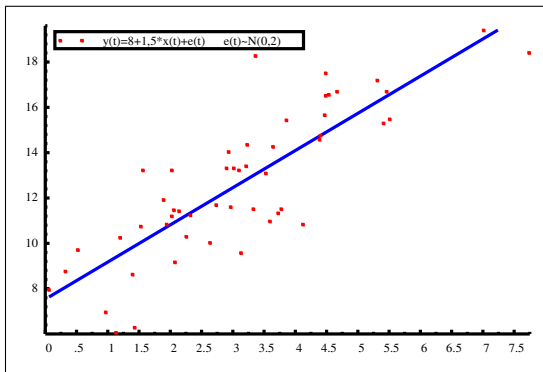
$$\min_{\mathbf{b}} S(\mathbf{b}) \tag{2.3}$$

uzyskujemy różniczkując $S(\mathbf{b})$ względem \mathbf{b}

$$\frac{\partial S(\mathbf{b})}{\partial \mathbf{b}} = \frac{\partial \mathbf{y}'\mathbf{y}}{\partial \mathbf{b}} - 2 \frac{\partial \mathbf{b}'\mathbf{X}'\mathbf{y}}{\partial \mathbf{b}} + \frac{\partial \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}}{\partial \mathbf{b}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\mathbf{b}$$

przyrównując uzyskaną pochodną do zera

$$-2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{0}$$



Rysunek 2: Krzywa regresji

po przeniesieniu $X'y$ na prawą stronę i podzieleniu obu stron przez 2 uzyskujemy układ równań zwany układem równań normalnych

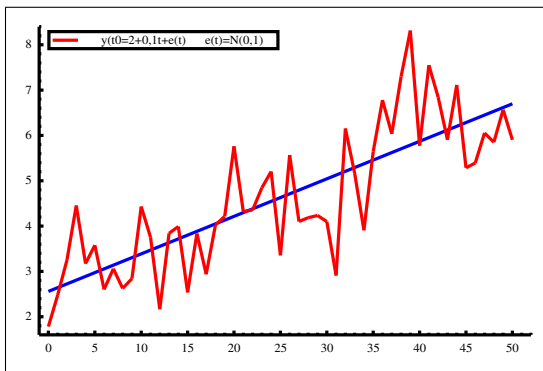
$$X' X b = X' y$$

Rozwiązując go dla b otrzymujemy postać analityczną estymatora MNK

$$b = (X' X)^{-1} X' y$$

Warunki drugiego rzędu dla minimalizacji (2.3) związane są z określonością hessianu (macierzy drugich pochodnych) funkcji $S(b)$ o postaci

$$\frac{\partial^2 S(b)}{\partial b \partial b'} = -2 \frac{\partial X' y}{\partial b'} + 2 \frac{\partial X' X b}{\partial b'} = 2 X' X$$



Rysunek 3: Model trendu liniowego

Ponieważ macierz $X'X$ jest zawsze dodatnio określona więc istotnie znaleźliśmy minimum funkcji $S(\mathbf{b})$

Przykład 2.1 CAPM

Rozważmy przypadek inwestora, który inwestując kieruje się jedynie wartością oczekiwaną i wariancją inwestycji. Załóżmy, że może on zainwestować w portfel, który opisany jest wektorem

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

gdzie x_i oznacza udział papieru wartościowego i w portfelu. Przychód z aktywów można opisać

wektorem

$$\mathbf{R} = \begin{bmatrix} R_1 \\ \vdots \\ R_n \end{bmatrix},$$

gdzie R_i oznacza stopę przychodu uzyskiwaną z papieru wartościowego i . Oczekiwane przychody z aktywów można oznaczyć jako $\overline{\mathbf{R}} = \mathbb{E}(\mathbf{R})$ a macierz wariancji kowariancji tych przychodów jako $\text{Var}(\mathbf{R})$.

Zakładamy, że inwestorzy chcą osiągnąć maksymalny zysk i charakteryzują się różnym stopniem awersji do ryzyka. Zakładamy, że istnieją pewne aktywa, z których inwestor może osiągnąć zysk R_0 bez ponoszenia żadnego ryzyka. Oczywiście jest, że inwestor będzie skłonny ponosić ryzyko jedynie wtedy, kiedy spodziewany przychód z inwestycji $\mathbf{x}'\overline{\mathbf{R}}$ będzie wyższy niż koszt alternatywny $(\sum_{i=1}^n x_i) R_0 = \mathbf{x}'\mathbf{l}R_0$, gdzie $\mathbf{l} = (1, \dots, 1)'$ Maksymalizacja funkcji użyteczności inwestora sprowadza się do maksymalizacji zysku $\mathbf{x}'(\overline{\mathbf{R}} - \mathbf{l}R_0)$ przy z góry założonej przez inwestora wariancji portfela $\sigma^2 = \text{Var}(\mathbf{x}'\mathbf{R}) = \mathbf{x}'\text{Var}(\mathbf{R})\mathbf{x}$. Problem ten można zapisać następująco

$$\begin{aligned} \max_{\mathbf{x}} \mathbf{x}'(\overline{\mathbf{R}} - \mathbf{l}R_0) \\ \text{s.t. } \mathbf{x}'\text{Var}(\mathbf{R})\mathbf{x} = \sigma^2 \end{aligned}$$

Rozwiązanie można znaleźć za pomocą różniczkując funkcję Lagrange'a:

$$L = \mathbf{x}'(\overline{\mathbf{R}} - \mathbf{l}R_0) + \lambda(\sigma^2 - \mathbf{x}'\text{Var}(\mathbf{R})\mathbf{x})$$

Warunki pierwszego rzędu mają wtedy postać:

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{x}} = \overline{\mathbf{R}} - \mathbf{l}R_0 - 2\lambda \text{Var}(\mathbf{R})\mathbf{x} &= \mathbf{0} \\ \frac{\partial L}{\partial \lambda} = \sigma^2 - \mathbf{x}'\text{Var}(\mathbf{R})\mathbf{x} &= \mathbf{0} \end{aligned} \tag{2.4}$$

Mnożąc pierwsze równanie przez \mathbf{x}' i wykorzystując drugie warunek z drugiego równania otrzymujemy, że

$$\mathbf{x}' (\bar{\mathbf{R}} - \mathbf{l}R_0) - 2\lambda\sigma^2 = 0$$

Rozwiązując tę równość dla λ dostajemy

$$\lambda = \frac{\mathbf{x}' (\bar{\mathbf{R}} - \mathbf{l}R_0)}{2\sigma^2}$$

Z kolei rozwiązując drugie równanie dla \mathbf{x} otrzymujemy

$$\mathbf{x} = \frac{1}{2\lambda} \text{Var}^{-1}(\mathbf{R}) (\bar{\mathbf{R}} - \mathbf{l}R_0)$$

Zauważmy, że w ostatnim równaniu λ jest skalarzem a $\text{Var}^{-1}(\mathbf{R}) (\bar{\mathbf{R}} - \mathbf{l}R_0)$ wektorem. Wynika z tego, że proporcje między papierami wartościowymi w portfelu są zdeterminowane przez $\text{Var}(\mathbf{R})$, $\bar{\mathbf{R}}$ i R_0 , a ilość zainwestowanych środków zależy także od σ^2 . Jeśli oczekiwania inwestorów na temat $\text{Var}(\mathbf{R})$, $\bar{\mathbf{R}}$ i R_0 są homogeniczne to proporcje papierów wartościowych w portfelach inwestorów o różnym poziomie awersji do ryzyka będą takie same i równe proporcjom kapitalizacji rynkowej tych papierów. CAPM może być więc prawdziwy tylko wtedy, gdy portfel \mathbf{x}_e , zawierający papiery w takich proporcjach w jakich są one na rynku, spełnia równania (2.4). Znormalizujemy $\tilde{\mathbf{x}}$ tak, że $\sum_{i=1}^n \tilde{x}_i = \mathbf{x}'\mathbf{l} = 1$ i oznaczymy $R_e = \mathbf{x}'_e \bar{\mathbf{R}}$. Wartość R_e można interpretować jako indeks średniego przychodu z inwestycji na giełdzie¹. Z równania na λ otrzymujemy

$$\lambda = \frac{R_e - R_0}{2\sigma^2}$$

a z (2.4)

$$\bar{\mathbf{R}} = \mathbf{l}R_0 + \frac{\text{Var}(\mathbf{R}) \mathbf{x}_e}{\sigma^2} (R_e - R_0)$$

¹W Polsce takim indeksem jest WIG.

Zauważmy teraz, że $\sigma_e^2 = \mathbf{x}'_e \text{Var}(\mathbf{R}) \mathbf{x}_e = \text{Var}(\mathbf{x}'_e \mathbf{R}) = \text{Var}(R_e)$ a element i wektora $\text{Var}(\mathbf{R}) \mathbf{x}_e$ jest równy $\text{Cov}(R_i, \mathbf{x}'_e \mathbf{R}) = \text{Cov}(R_i, R_e)$. W rezultacie

$$\bar{R}_i = R_0 + \frac{\text{Cov}(R_i, R_e)}{\text{Var}(R_e)} (R_e - R_0)$$

Jeśli założymy, że R_0 oraz współczynnik $\text{Cov}(R_i, R_e)$, $\text{Var}(R_e)$ są stałe w czasie, to $\beta = \frac{\text{Cov}(R_i, R_e)}{\text{Var}(R_e)}$ można oszacować estymując równanie

$$R_{it} = \alpha + \beta R_{et} + \varepsilon_t$$

ponieważ

$$b = \frac{\frac{\sum_{t=1}^T (R_{it} - \bar{R}_i)(R_{et} - \bar{R}_e)}{T}}{\frac{\sum_{t=1}^T (R_{et} - \bar{R}_e)^2}{T}} \xrightarrow{p} \frac{\text{Cov}(R_i, R_e)}{\text{Var}(R_e)}$$

przy czym R_0 można oszacować korzystając z tego, że

$$\alpha = (1 - \beta) R_0$$

Literatura: Varian (1992), str. 371-375

2.2. Własności hiperpłaszczyzny regresji

Hiperpłaszczyzną regresji nazywamy płaszczyznę dopasowaną do danych z pomocą MNK. Płaszczyzna tej wyznaczona funkcją $\tilde{\mathbf{y}} = \tilde{\mathbf{X}} \mathbf{b}$, gdzie \mathbf{b} jest estymatorem MNK a $\tilde{\mathbf{X}} \in R^K$.

Twierdzenie 2.2 Hiperpłaszczyzna regresji wyznaczona za pomocą MNK posiada następujące własności

1. $\mathbf{X}' \mathbf{e} = \mathbf{0}$

$$2. \hat{\mathbf{y}}' \mathbf{e} = 0$$

Dodatkowo dla modelu ze stałą

$$3. \mathbf{l}' \mathbf{e} = 0$$

$$4. \bar{y} = \widehat{\bar{y}}$$

gdzie $\hat{\mathbf{y}} = \mathbf{X} \mathbf{b}$ i $\mathbf{l}' = (1, \dots, 1)$.

Dowód.

$$\mathbf{X}' \mathbf{y} - \mathbf{X}' \mathbf{X} \mathbf{b} = \mathbf{0} \Leftrightarrow \mathbf{X}' (\mathbf{y} - \mathbf{X} \mathbf{b}) = \mathbf{0} \Leftrightarrow \mathbf{X}' \mathbf{e} = \mathbf{0}$$

$$\hat{\mathbf{y}} = \mathbf{X} \mathbf{b}$$

$$\hat{\mathbf{y}}' \mathbf{e} = \mathbf{b}' \mathbf{X}' \mathbf{e} = 0$$

$$\mathbf{X} = \begin{bmatrix} 1 & x_{21} & x_{31} & \cdots & x_{K1} \\ 1 & x_{22} & x_{32} & \cdots & x_{K2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{2n} & x_{3n} & \cdots & x_{Kn} \end{bmatrix}$$

$$\mathbf{X}' \mathbf{e} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{21} & x_{22} & \cdots & x_{2n} \\ x_{31} & x_{32} & \cdots & x_{3n} \\ \vdots & \vdots & & \vdots \\ x_{K1} & x_{K2} & \cdots & x_{Kn} \end{bmatrix} \mathbf{e}$$

$$\mathbf{X}' \mathbf{e} = \mathbf{0}; \mathbf{l}' \mathbf{e} = 0$$

$$\mathbf{y} = \mathbf{X} \mathbf{b} + \mathbf{e} = \hat{\mathbf{y}} + \mathbf{e}$$

$$\mathbf{l}' \mathbf{y} = \mathbf{l}' \hat{\mathbf{y}} + \mathbf{l}' \mathbf{e} = \mathbf{l}' \hat{\mathbf{y}}$$

$$\frac{\mathbf{l}' \mathbf{y}}{n} = \bar{y} = \frac{\mathbf{l}' \hat{\mathbf{y}}}{n} = \widehat{\bar{y}}$$

■

Wniosek 2.3 Reszty są ortogonalne (prostopadłe) do \mathbf{X} . Próba regresji \mathbf{e} (niewyjaśnionej części \mathbf{y}) na \mathbf{X} da $\mathbf{b}_e = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e} = \mathbf{0}$.

Definicja 2.4 Sum kwadratów TSS , ESS i RSS

$$TSS = (\mathbf{y} - \bar{\mathbf{y}})'(\mathbf{y} - \bar{\mathbf{y}}) \quad \text{całkowita suma kwadratów (Total Sum of Squares)}$$

$$ESS = (\hat{\mathbf{y}} - \bar{\hat{\mathbf{y}}})'(\hat{\mathbf{y}} - \bar{\hat{\mathbf{y}}}) \quad \text{wyjaśniona suma kwadratów (Explained Sum of Squares)}$$

$$RSS = \mathbf{e}'\mathbf{e} \quad \text{suma kwadratów reszt (Residual Sum of Squares)}$$

gdzie $\bar{\mathbf{y}} = l\bar{y}$, $\bar{\hat{\mathbf{y}}} = l\bar{\hat{y}}$.

Lemat 2.5 W modelu ze stałą

$$TSS = ESS + RSS \quad (2.5)$$

Dowód. Z tego, że $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e} = \hat{\mathbf{y}} + \mathbf{e}$ oraz $\bar{\mathbf{y}} = \bar{\hat{\mathbf{y}}}$ wynika, że $\mathbf{y} - \bar{\mathbf{y}} = l\bar{y} - l\bar{\hat{y}} = \bar{\hat{\mathbf{y}}} + \mathbf{e}$ oraz, że

$$\mathbf{y} - \bar{\mathbf{y}} = (\hat{\mathbf{y}} - \bar{\hat{\mathbf{y}}}) + \mathbf{e}$$

$$\begin{aligned} (\mathbf{y} - \bar{\mathbf{y}})'(\mathbf{y} - \bar{\mathbf{y}}) &= [(\hat{\mathbf{y}} - \bar{\hat{\mathbf{y}}}) + \mathbf{e}]' [(\hat{\mathbf{y}} - \bar{\hat{\mathbf{y}}}) + \mathbf{e}] \\ &= (\hat{\mathbf{y}} - \bar{\hat{\mathbf{y}}})'(\hat{\mathbf{y}} - \bar{\hat{\mathbf{y}}}) + 2(\hat{\mathbf{y}} - \bar{\hat{\mathbf{y}}})'\mathbf{e} + \mathbf{e}'\mathbf{e} \\ &= (\hat{\mathbf{y}} - \bar{\hat{\mathbf{y}}})'(\hat{\mathbf{y}} - \bar{\hat{\mathbf{y}}}) + \mathbf{e}'\mathbf{e} \end{aligned}$$

ponieważ:

$$(\hat{\mathbf{y}} - \bar{\hat{\mathbf{y}}})'\mathbf{e} = \hat{\mathbf{y}}'\mathbf{e} + \bar{\hat{\mathbf{y}}}'\mathbf{e} = \hat{\mathbf{y}}'\mathbf{e} + (l\bar{\hat{y}})'\mathbf{e} = \hat{\mathbf{y}}'\mathbf{e} + \hat{\mathbf{y}}l'\mathbf{e} = 0$$

przy czym korzystamy z własności hiperpłaszczyzny regresji. ■

Twierdzenie 2.6 Współczynnik determinacji R^2 definiujemy jako

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

i dla modelu ze stałą spełnia on zawsze następujące nierówności

$$0 \leq R^2 \leq 1$$

Dowód.

$$TSS > 0, ESS > 0, RSS > 0: \frac{ESS}{TSS} \geq 0$$

$$TSS = ESS + RSS$$

$$RSS > 0 \Leftrightarrow TSS - ESS > 0 \Leftrightarrow TSS > ESS \Leftrightarrow \frac{ESS}{TSS} \leq 1$$

■

Uwaga 2.7 R^2 interpretujemy jako procent wyjaśnionej przez model zmienności zmiennej endogenicznej. Wnioski dotyczące możliwych wartości R^2 są prawdziwe jedynie dla modelu ze stałą. Gdy w modelu nie ma stałej R^2 może być ujemne. R^2 ma następujące własności:

1. jest równy jeden, jeśli $\mathbf{y} = \mathbf{X}\mathbf{b} \Leftarrow \mathbf{e} = \mathbf{0}$
2. jest równy zero, jeśli zmienna endogeniczna jest ortogonalna do zmiennych egzogenicznych $\Leftrightarrow \mathbf{y}'\mathbf{X} = \mathbf{0}$
3. pozostaje niezmienny, gdy zmieniamy skalę zmiennych, to jest zmienne egzogeniczne x_{ki} zastępujemy zmiennymi $x_{ki}^* = \theta_{1k} + \theta_{2k}x_{ki}$ a zmienną endogeniczną zmienną y_i zmienną $y_i^* = \gamma_1 + \gamma_2 y_i$, gdzie $\gamma_1, \gamma_2, \theta_{1k}, \theta_{2k}$ są stałymi, które zależą od indeksu zmiennej $k = 1, \dots, K$ ale nie zależą od indeksu obserwacji $i = 1, \dots, n$.

Wniosek 2.8 W modelu ze stałą i jedną zmienną R^2 jest równy kwadratowi współczynnika korelacji

$$\hat{\rho}_{xy} = \frac{s_{yx}}{\sqrt{s_y s_x}} \text{ między } y \text{ i } x$$

Dowód.

$$\begin{aligned} R^2 &= \frac{(\hat{\mathbf{y}} - \bar{\mathbf{y}})' (\hat{\mathbf{y}} - \bar{\mathbf{y}})}{(\mathbf{y} - \bar{\mathbf{y}})' (\mathbf{y} - \bar{\mathbf{y}})} = \frac{\sum_{i=1}^n (b_0 + x_i b_1 - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \\ &= b_1^2 \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \left(\frac{s_{yx}}{s_x^2} \right)^2 \frac{s_x^2}{s_y^2} = \frac{s_{yx}^2}{s_y^2 s_x^2} = \hat{\rho}_{xy}^2 \end{aligned}$$

przy czym w wyprowadzeniach skorzystaliśmy z łatwego do udowodnienia (patrz zadania) faktu, że w modelu z jedną zmienną $b_0 = \bar{y} - b_1 \bar{x}$ i $b_1 = \frac{s_{yx}}{s_x^2}$. ■

2.3. Interpretacja geometryczna MNK

Oznaczmy przestrzeń liniową generowaną przez kolumny macierzy \mathbf{X} jako $\text{sp}(\mathbf{X})$. Jak wiemy z algebry liniowej dla każdego wektora \mathbf{y} o odpowiednich wymiarach $\mathbf{X}\mathbf{y} \in \text{sp}(\mathbf{X})$. Oznaczmy z kolei jako \mathbf{X}_\perp macierz złożoną z wektorów ortogonalnych do \mathbf{X} taką, że $[\mathbf{X}, \mathbf{X}_\perp]$ jest macierzą o pełnym rzędzie. Przestrzeń liniowa $\text{sp}(\mathbf{X}_\perp)$ będzie więc zawierać wektory ortogonalne do wektorów będących elementami przestrzeni $\text{sp}(\mathbf{X})$.

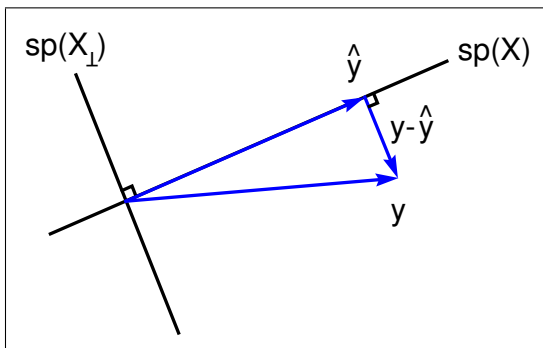
Zastanówmy się teraz jak można zinterpretować geometrycznie zadanie minimalizacyjne, która jest podstawą szacowania estymatorów za pomocą MNK . Szukamy takiego \mathbf{b} , które minimalizuje

$$\begin{aligned} \min_{\mathbf{b}} S(\mathbf{b}) &= \min_{\mathbf{b}} (\mathbf{y} - \mathbf{X}\mathbf{b})' (\mathbf{y} - \mathbf{X}\mathbf{b}) = \min_{\mathbf{b}} \sqrt{\|\mathbf{y} - \mathbf{X}\mathbf{b}\|^2} \\ &\Leftrightarrow \min_{\mathbf{b}} \|\mathbf{y} - \hat{\mathbf{y}}\| \quad \text{s.t. } \hat{\mathbf{y}} \in \text{sp}(\mathbf{X}) \end{aligned}$$

a więc szukamy takiego wektora $\hat{\mathbf{y}}$ należącego do przestrzeni generowanej przez kolumny macierzy \mathbf{X} , który jest najmniej odległy od wektora \mathbf{y} . Układ równań normalnych

$$\mathbf{X}'(\mathbf{y} - \mathbf{X}\mathbf{b}) = \mathbf{X}'(\mathbf{y} - \hat{\mathbf{y}}) = \mathbf{0}$$

implikuje, że wektor $\hat{\mathbf{y}}$, który stanowi rozwiązanie tego zadania ma tą cechę, że wektor różnic $\mathbf{y} - \hat{\mathbf{y}} \in \text{sp}(\mathbf{X}_\perp)$. Warunek ten jest oczywisty jeśli przeanalizujemy następujący rysunek 4. Widać



Rysunek 4: Geometria MNK

z niego, że rzeczywiście wektorem, który należąc do wyznaczonej przestrzeni liniowej wyznaczonej osią $\text{sp}(\mathbf{X})$, jest najmniej odległy od wektora \mathbf{y} jest wektor $\hat{\mathbf{y}}$, dla którego różnica $\hat{\mathbf{y}} - \mathbf{y}$ jest prostopadła do $\text{sp}(\mathbf{X})$.

Rzut prostopadły w przestrzeń generowaną \mathbf{X} , to takie przekształcenie liniowe $\mathbf{F}(\mathbf{y})$, dla którego:

1. $\mathbf{F}(\mathbf{y}) \in \text{sp}(\mathbf{X})$ (przekształca \mathbf{y} na wektor będący kombinacją liniową kolumn macierzy \mathbf{X})
2. $\mathbf{F}(\mathbf{X}) = \mathbf{X}$ (nie wpływa na kolumny macierzy \mathbf{X})
3. $\mathbf{y} - \mathbf{F}(\mathbf{y}) \in \text{sp}(\mathbf{X}_\perp)$ (rzutuje prostopadle)

Jak wiadomo z algebry każde przekształcenie liniowe można opisać za pomocą macierzy. Spróbujemy więc znaleźć macierz rzutu prostopadłego przestrzeń $\text{sp}(\mathbf{X})$

Twierdzenie 2.9 *Macierz*

$$P_X = X (X'X)^{-1} X'$$

definiuje rzut prostopadły w przestrzeń $\text{sp}(X)$.

Dowód. Punkt (1). Dla dowolnego wektora z mamy

$$P_X z = X \left[(X'X)^{-1} X'y \right] \in \text{sp}(X)$$

Punkt (3). Z definicji P_X

$$P_X X = X (X'X)^{-1} X'X = X$$

Punkt (3). Z definicji P_X i ortogonalności

$$X' (y - P_X y) = X' (I - P_X) y = (X - X) y = 0$$

■

Twierdzenie 2.10 *Macierz*

$$M_X = I - X (X'X)^{-1} X' = I - P_X$$

definiuje rzut prostopadłego w przestrzeń $\text{sp}(X_\perp)$

Dowód. Podobno jak w poprzednim punkcie. ■

Macierz P_X rzutuje wektor y na przestrzeń $\text{sp}(X)$ a macierz M_X na przestrzeń $\text{sp}(X_\perp)$. Ponadto z definicji tych macierzy tych wynika, że

$$I = P_X + M_X$$

i przekształcenia zdefiniowane przez macierze P_X i M_X można interpretować w kategoriach przekształcenia bazy wektora z , ze złożonej z wektorów jednostkowych do takiej, która składa się z kolumn macierzy $[X, X_\perp]$ przy czym

$$y = P_X y + M_X y = Xb + e = \hat{y} + e$$

z czego wynika, że w tej nowej bazie elementy wektora \mathbf{y} związane z wektorami bazowymi danymi przez kolumny macierzy \mathbf{X} będą dane przez wektor \mathbf{b} , $\hat{\mathbf{y}}$ jest rzutem prostopadłym \mathbf{y} w przestrzeń $\text{sp}(\mathbf{X})$ zaś wektor reszt \mathbf{e} jest rzutem prostopadłym \mathbf{y} w przestrzeń $\text{sp}(\mathbf{X}_\perp)$.

2.4. Twierdzenie Frischa-Waugh-Lavella (FWL)

Twierdzenie 2.11 *Estymator MNK parametru β_1 policzony dla modelu*

$$\mathbf{y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \varepsilon \quad (2.6)$$

ma tę samą wartość co estymator MNK parametru β_1^ policzony dla modelu*

$$\mathbf{y}^* = \mathbf{X}_1^*\beta_1^* + \varepsilon^*, \quad (2.7)$$

gdzie $\mathbf{y}^* = \mathbf{M}_2 \mathbf{y}$, $\mathbf{X}_1^* = \mathbf{M}_2 \mathbf{X}_1$ i $\mathbf{M}_2 = \mathbf{I} - \mathbf{X}_2 (\mathbf{X}_2' \mathbf{X}_2)^{-1} \mathbf{X}_2'$ są resztami z regresji \mathbf{y} na \mathbf{X}_2 i \mathbf{X}_1 na \mathbf{X}_2 . Ponadto reszty \mathbf{e} z modelu (2.6) są równe resztom \mathbf{e}^* z modelu (2.7).

Dowód. Estymator parametru β_1^* dla modelu (2.7) będzie z definicji równy

$$\mathbf{b}_1^* = (\mathbf{X}_1^{*'} \mathbf{X}_1^*)^{-1} \mathbf{X}_1^{*'} \mathbf{y}^* = (\mathbf{X}_1' \mathbf{M}_2 \mathbf{X}_1)^{-1} \mathbf{X}_1' \mathbf{M}_2 \mathbf{y} \quad (2.8)$$

przy czym druga równość wynika z tego, że \mathbf{M}_2 jest symetryczne i idempotentne. Reszty dla takiego równania regresji można policzyć z równania

$$\mathbf{M}_2 \mathbf{y} = \mathbf{M}_2 \mathbf{X}_1 \mathbf{b}_1^* + \mathbf{e}^* \quad (2.9)$$

Z kolei w przypadku modelu (2.6) reszty można policzyć z równania

$$\mathbf{y} = \mathbf{X}_1 \mathbf{b}_1 + \mathbf{X}_2 \mathbf{b}_2 + \mathbf{e} \quad (2.10)$$

przy czym \mathbf{b}_1 i \mathbf{b}_2 uzyskujemy za pomocą MNK. Z własności hiperpłaszczyzny regresji wiemy, że reszty muszą spełniać układ równań normalnych

$$\mathbf{X}' \mathbf{e} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_2 \end{bmatrix}' \mathbf{e} = \mathbf{0}$$

co implikuje $X_1'e = 0$ i $X_2'e = 0$. Zdefiniujmy

$$M_2 = I - X_2 (X_2'X_2)^{-1} X_2'$$

Mnożąc równanie (2.10) przez M_2 otrzymujemy

$$M_2\mathbf{y} = M_2X_1\mathbf{b}_1 + M_2X_2\mathbf{b}_2 + M_2\mathbf{e}$$

Ponieważ $M_2X_2 = 0$ i $M_2\mathbf{e} = \mathbf{e} - X_2 (X_2'X_2)^{-1} X_2'\mathbf{e} = \mathbf{e}$, więc

$$M_2\mathbf{y} = M_2X_1\mathbf{b}_1 + \mathbf{e} \tag{2.11}$$

Mnożąc z kolei to równanie prawostronnie przez X_1 uzyskujemy

$$X_1'M_2\mathbf{y} = X_1'M_2X_1\mathbf{b}_1 + X_1'\mathbf{e}$$

a skoro $X_1'\mathbf{e} = 0$ więc

$$\mathbf{b}_1 = (X_1'M_2X_1)^{-1} X_1'M_2\mathbf{y}$$

i tym samym dowiedliśmy, że $\mathbf{b}_1^* = \mathbf{b}_1$. To, że reszty z modelu (2.6) i modelu (2.7) muszą być takie same wynika teraz z równania (2.11). ■

Przykład 2.12 *Regresja na odchyleniach od średnich.* Zastanówmy się nad tym jak się ma się estymator MNK parametru β_1 policzony dla modelu

$$\mathbf{y} = \beta_0 + \mathbf{X}\beta_1 + \boldsymbol{\varepsilon}$$

i policzony dla modelu

$$\mathbf{y}^* = \mathbf{X}^*\beta_1^* + \boldsymbol{\varepsilon}^*$$

gdzie \mathbf{y}^* , \mathbf{X}^* są odchyleniami \mathbf{y} , \mathbf{X} od swoich średnich.

Rozwiązanie 2.13 Zauważmy, że

$$M_l = I - l(l'l)^{-1}l' = I - n^{-1}ll'$$

Reszty z regresji \mathbf{y} na wektorze jedynek $\mathbf{l} = [1, \dots, 1]'$ są równe

$$\mathbf{y}^* = M_{\mathbf{l}} \mathbf{y} = \mathbf{y} - \bar{y}$$

podobnie

$$\mathbf{x}_k^* = M_{\mathbf{l}} \mathbf{x}_k = \mathbf{x}_k - \bar{x}_k$$

Do drugiego modelu możemy więc zastosować twierdzenie FWL, przy czym $\mathbf{X}_2 = \mathbf{l}$ a $\mathbf{X}_1 = \mathbf{X}$ i wynioskować, że estymator $\mathbf{b}_1 = \mathbf{b}_1^*$.

Z twierdzenia FWL wynika, że regresję można przeprowadzić w dwóch krokach. Najpierw przeprowadzić regresję \mathbf{y} i \mathbf{X}_1 . Uzyskane w ten sposób reszty dla \mathbf{y} i \mathbf{X}_1 opisują zmienność, której nie da się wyjaśnić za pomocą \mathbf{X}_2 . Możemy teraz przeprowadzić regresję reszt z regresji \mathbf{y} na resztach z regresji \mathbf{X}_1 a uzyskany w ten sposób estymator \mathbf{b}_1 będzie równy estymatorowi uzyskanemu z regresji \mathbf{y} na \mathbf{X}_1 i \mathbf{X}_2 . Wynika z tego ważny wniosek dotyczący interpretacji elementów b_k wektora \mathbf{b} . Opisują one wpływ \mathbf{x}_k na \mathbf{y} po uwzględnieniu wpływu wszystkich pozostałych zmiennych.

Wniosek 2.14 Jeśli przeprowadzimy regresję \mathbf{y} na \mathbf{X}_2 a następnie \mathbf{y} na $\mathbf{X}_1^* = M_2 \mathbf{X}_1$, to

$$\hat{\mathbf{y}} = \hat{\mathbf{y}}^* + \hat{\mathbf{y}}_2$$

gdzie $\mathbf{y} = \mathbf{X}\mathbf{b}$, $\hat{\mathbf{y}}^* = \mathbf{X}_1^* \mathbf{b}_1$, $\hat{\mathbf{y}}_2 = \mathbf{X}_1^* \mathbf{b}_2^+$ i $\mathbf{b}_2^+ = (\mathbf{X}_2' \mathbf{X}_2)^{-1} \mathbf{X}_2' \mathbf{y}$ jest estymatorem MNK z regresji \mathbf{y} na \mathbf{X}_2 , $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2]$. Wynika z tego, następującą macierz \mathbf{P}_X :

$$\mathbf{P}_X = \mathbf{P}_{\mathbf{X}_1^*} + \mathbf{P}_2 \quad (2.12)$$

gdzie

$$\mathbf{P}_{\mathbf{X}_1^*} = \mathbf{X}_1^* (\mathbf{X}_1^{*'} \mathbf{X}_1^*)^{-1} \mathbf{X}_1^{*'}$$

$$\mathbf{P}_2 = \mathbf{X}_2 (\mathbf{X}_2' \mathbf{X}_2)^{-1} \mathbf{X}_2'$$

Dowód. Zauważmy, że pokazaliśmy, że dla każdego \mathbf{y} mamy, że $\mathbf{e} = \mathbf{y}^* - \mathbf{X}_1^* \mathbf{b}_1^*$. Podstawiając $\mathbf{y}^* = \mathbf{y} - \mathbf{X} \mathbf{b}_2^+ = \mathbf{y} - \hat{\mathbf{y}}_2$, $\hat{\mathbf{y}}^* = \mathbf{X}_1^* \mathbf{b}_1^*$ i $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$ do tego wzoru otrzymujemy udowodniany wynik. Drugą część twierdzenia możemy dowieść w sposób następujący: ponieważ macierze są kwadratowe i dla każdego \mathbf{y} mamy, że

$$\mathbf{P}_X \mathbf{y} = \hat{\mathbf{y}} = \hat{\mathbf{y}}^* + \hat{\mathbf{y}}_2 = \mathbf{P}_{\mathbf{X}_1^*} \mathbf{y} + \mathbf{P}_2 \mathbf{y}$$

więc musi zachodzić równość (2.12). ■

W sensie geometryczny twierdzenie to mówi, że rzut prostopadły w przestrzeń $sp(\mathbf{X})$ można zdekomponować na rzut w przestrzeń $sp(\mathbf{X}_2)$ i rzut w tę część przestrzeni $sp(\mathbf{X}_1)$, która jest ortogonalna do przestrzeni $sp(\mathbf{X}_2)$.

Wniosek 2.15 W równaniu regresji (2.7) RSS jest równa sumie kwadratów reszt w regresji

$$\mathbf{y} = \mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{X}_2 \boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$$

a TSS jest równa sumie kwadratów reszt w regresji

$$\mathbf{y} = \mathbf{X}_2 \boldsymbol{\gamma}_2 + \mathbf{u}$$

R^2 w równaniu regresji (2.7) ma wzór

$$R^2 = \frac{\hat{\mathbf{u}}' \hat{\mathbf{u}} - \mathbf{e}' \mathbf{e}}{\mathbf{e}' \mathbf{e}}$$

i

$$\hat{\mathbf{u}}' \hat{\mathbf{u}} - \mathbf{e}' \mathbf{e} = \mathbf{b}_1' \mathbf{X}_1' \mathbf{M}_2 \mathbf{X}_1 \mathbf{b}_1 \geq 0 \quad (2.13)$$

Dowód. Z twierdzenia FWL wiemy, że z regresji (2.7) są równe \mathbf{e} , a więc $RSS = \mathbf{e}' \mathbf{e}$, skoro z własności hiperpłaszczyzny regresji wynika, że $\tilde{\mathbf{e}} = \frac{\mathbf{1}' \mathbf{e}}{n} = 0$. Podobnie, z definicji macierzy \mathbf{M}_2 mamy, że $\mathbf{y}^* = \mathbf{M}_2 \mathbf{y} = \hat{\mathbf{u}}$ i stąd $TSS = (\hat{\mathbf{e}} - \tilde{\mathbf{e}})' (\hat{\mathbf{e}} - \tilde{\mathbf{e}}) = \hat{\mathbf{e}}' \hat{\mathbf{e}}$. Wzór na równanie R^2

uzyskujemy postawiając uzyskane wzory na TSS i RSS do definicji tej statystyki. Z równania (2.11) wnioskujemy, że $\hat{\mathbf{u}} = \mathbf{M}_2 \mathbf{X}_1 \mathbf{b}_1 + \mathbf{e}$ a więc

$$\begin{aligned}\hat{\mathbf{u}}' \hat{\mathbf{u}} &= (\mathbf{M}_2 \mathbf{X}_1 \mathbf{b}_1 + \mathbf{e})' (\mathbf{M}_2 \mathbf{X}_1 \mathbf{b}_1 + \mathbf{e}) \\ &= \mathbf{b}_1' \mathbf{X}_1' \mathbf{M}_2 \mathbf{X}_1 \mathbf{b}_1 + 2\mathbf{e}' \mathbf{M}_2 \mathbf{X}_1 \mathbf{b}_1 + \mathbf{e}' \mathbf{e} \\ &= \mathbf{b}_1' \mathbf{X}_1' \mathbf{M}_2 \mathbf{X}_1 \mathbf{b}_1 + \mathbf{e}' \mathbf{e}\end{aligned}$$

przy czym skorzystaliśmy z tego, że $\mathbf{M}_2 \mathbf{e} = \mathbf{e}$ i $\mathbf{X}_1' \mathbf{e} = \mathbf{0}$. Wartość $\mathbf{b}_1' \mathbf{X}_1' \mathbf{M}_2 \mathbf{X}_1 \mathbf{b}_1$ jest zawsze dodatnia, ponieważ macierz $\mathbf{X}_1' \mathbf{M}_2 \mathbf{X}_1$ jest dodatnio określona. ■

Wzór (2.13) implikuje, że usuwając z modelu zmienne powodujemy zawsze wzrost sumy kwadratów reszt. Ponieważ TSS w obu przypadkach jest takie same, więc

$$RSS_2 \geq RSS_1 =: R_1^2 \leq R_2^2.$$

Statystyka R^2 maleje więc zawsze jeśli z modelu usuwamy zmienne. Nie ma przy tym żadnego znaczenia, czy zmienne te są w rzeczywistości istotne dla kształtowania się zmiennej zależnej a wynika jedynie z własności algebraicznych MNK . Z tego też powodu statystykę R^2 można stosować jedynie jako statystykę opisową. Nie może ona jednak służyć np. do ustalanie, czy zmienne w modelu są istotne satystycznie.

Uwaga 2.16 *Dodając do modelu kolejne zmienne egzogeniczne otrzymujemy model o wyższym R^2 nawet wtedy, gdy zmienne te są nieistotne i ich dodanie podwyższa wariancję estymatora \mathbf{b} . Z tego powodu R^2 nie może być jedyną miarą adekwatności modelu. Lepszą miarą jakości dopasowania wydaje się skorygowane R^2*

$$\bar{R}^2 = 1 - \frac{n-1}{n-K} (1 - R^2)$$

Modyfikacja ta jest tak zaprojektowana, że \bar{R}^2 spada, gdy wartość bezwzględna statystyki t dla usuniętej zmiennej jest większa od 1.

Uwaga 2.17 Znaczenie poszczególnych zmiennych w wyjaśnianiu \mathbf{y} można mierzyć cząstkowym R_j^2 zdefiniowanym jako R^2 w modelu $\tilde{y}_i = \beta \tilde{x}_{ji}$, gdzie są \tilde{y} i \tilde{x}_j są resztami z regresji y i x_j na wszystkich zmiennych poza x_j . Cząstkowy współczynnik określoności R_j^2 mierzy procent zmienności y , którą wyjaśnia tylko zmienność x_j . Ponieważ suma reszt jest równa zero, więc z twierdzenia 2.8 otrzymujemy, że R_j^2 jest równy

$$R_j^2 = \frac{\sum_{i=1}^n \tilde{y}_i \tilde{x}_{ji}}{\sum_{i=1}^n \tilde{y}_i^2 \sum_{i=1}^n \tilde{x}_{ji}^2} = \frac{(\mathbf{y} \mathbf{M}_j \mathbf{x}_j)^2}{(\mathbf{x}'_j \mathbf{M}_j \mathbf{x}_j) (\mathbf{y}' \mathbf{M}_j \mathbf{y})}$$

gdzie $\mathbf{M}_j = \mathbf{I} - \mathbf{X}_j (\mathbf{X}_j' \mathbf{X}_j)^{-1} \mathbf{X}_j'$ i \mathbf{X}_j zawiera wszystkie zmienne należące do \mathbf{X} poza zmienną x_j . Zauważmy, że $R_j^2 \leq R^2$ i $\sum_{j=1}^K R_j^2 \neq R^2$.

Wniosek 2.18 W równaniu regresji

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

w którym macierz $\mathbf{X} = [\mathbf{X}_1 \quad \mathbf{X}_2]$ i odpowiednio $\boldsymbol{\beta} = [\beta'_1 \quad \beta'_2]'$ estymator MNK

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$$

można także zapisać jako

$$b_1 = (\mathbf{X}'_1 \mathbf{M}_2 \mathbf{X}_1)^{-1} \mathbf{X}'_1 \mathbf{M}_2 \mathbf{y}$$

$$b_2 = (\mathbf{X}'_2 \mathbf{M}_1 \mathbf{X}_2)^{-1} \mathbf{X}'_2 \mathbf{M}_1 \mathbf{y}$$

gdzie

$$\mathbf{M}_1 = \mathbf{I} - \mathbf{X}_1 (\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1$$

$$\mathbf{M}_2 = \mathbf{I} - \mathbf{X}_2 (\mathbf{X}'_2 \mathbf{X}_2)^{-1} \mathbf{X}'_2$$

Dowód. Wzór dla b_1 wynika ze wzoru (2.8) i tego, że $b_1 = b_1^*$. Ponieważ podział na zmienne X_1 i X_2 jest arbitralny, więc w identyczny sposób dowodzimy równania dla b_2 . ■

2.5. Własności statystyczne *MNK*

Założenia **Klasycznego Modelu Regresji Liniowej** (*KMRL*) można streścić w następujących punktach:

Założenie 2.19 Związek między y a X opisany jest równaniem $y = X\beta + \varepsilon$

Założenie 2.20 Macierz zmiennych egzogenicznych X jest nielosowa

Założenie 2.21 $E(\varepsilon_i) = 0$ wartość oczekiwana błędu losowego jest równa zero

Założenie 2.22 $E(\varepsilon_i \varepsilon_j) = 0$ błędy losowe są nieskorelowane

Założenie 2.23 $\text{Var}(\varepsilon_i) = \sigma^2$ wariancja błędu losowego pozostaje stała w czasie

Założenie 2.24 $\varepsilon \sim N(0, \sigma^2 I)$ błąd losowy ma rozkład normalny

Założenia 2.21, 2.22, 2.23 można opisać też zapisać jako $E(\varepsilon\varepsilon') = \sigma^2 I$. Założenie 2.24 wykorzystuje się jedynie w niektórych wyprowadzeniach.

Twierdzenie 2.25 Własności estymatora *MNK*

1. Estymator *MNK* jest nieobciążony
2. Wariancja estymatora *MNK* wynosi $\Sigma = \sigma^2 (X'X)^{-1}$
3. Nieobciążony estymator wariancji błędu losowego ma postać

$$s^2 = \frac{e'e}{n - K} = \frac{\sum_{i=1}^n e_i^2}{n - K}$$

4. Nieobciążonym estymatorem wariancji b jest estymator $S = s^2 (X'X)^{-1}$

Dowód. (1)

$$\begin{aligned} E(\mathbf{b}) &= E\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}\right) = E\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})\right) \\ &= E\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta}\right) + E\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\right) = \\ &E(\boldsymbol{\beta}) + (\mathbf{X}'\mathbf{X})^{-1}E(\boldsymbol{\varepsilon}) = \boldsymbol{\beta} \end{aligned}$$

(2)

$$\begin{aligned} \text{Var}(\mathbf{b}) &= \text{Var}\left(\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\right) = E\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right] = \\ &(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

(3)

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{b} = \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y} = \mathbf{M}\mathbf{y}$$

$$\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

$$\mathbf{M}' = \mathbf{M}$$

$$\mathbf{M}\mathbf{M} = \mathbf{M}$$

$$\begin{aligned} \mathbf{M}\mathbf{M} &= \left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right] \left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right] \\ &= \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ &= \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{M} \end{aligned}$$

$$\mathbf{M}\mathbf{X} = \mathbf{X} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{0}$$

$$\mathbf{e} = \mathbf{M}\mathbf{y} = \mathbf{M}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \mathbf{M}\boldsymbol{\varepsilon}$$

$$\mathbf{e}'\mathbf{e} = \boldsymbol{\varepsilon}'\mathbf{M}'\mathbf{M}\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}$$

Ostatnie wyrażenie jest skalarem. Wynika z tego, że

$$e'e = \text{tr}(e'e) = \text{tr}(M\varepsilon\varepsilon')$$

$$E(e'e) = E(\text{tr}(M\varepsilon\varepsilon')) = \text{tr}(ME(\varepsilon\varepsilon')) = \text{tr}(M\sigma^2I) = \sigma^2 \text{tr}(M)$$

$$\begin{aligned} \text{tr}(M) &= \text{tr}(I_{n \times n}) - \text{tr}\left(X(X'X)^{-1}X'\right) = \text{tr}(I_{n \times n}) - \text{tr}\left((X'X)^{-1}X'X\right) \\ &= \text{tr}(I_{n \times n}) - \text{tr}I_{K \times K} = n - K \end{aligned}$$

$$E(e'e) = \sigma^2 \text{tr}(M) = \sigma^2(n - K)$$

a więc

$$E\left(\frac{e'e}{n - K}\right) = \sigma^2$$

(4)

$$E\left(s^2(X'X)^{-1}\right) = E(s^2)(X'X)^{-1} = \sigma^2(X'X)^{-1} = \Sigma$$

■

Literatura: Steward (1991) str. 25-43, Green (1997) str. 221-257, Chow (1995) str. 58-60, Goldberger (1972) str. 209-216 i 220-222, Theil (1979) str. 56-64 i 121-137.

2.6. Twierdzenie Gaussa-Markowa

Twierdzenie 2.26 (Gaussa-Markowa) b_{MNLK} jest najlepszym liniowym i nieobciążonym estymatorem parametru β w KMRL (Klasycznym Modelu Regresji Liniowej).

Estymator liniowy, o którym jest mowa w twierdzeniu, jest to estymator, który można przedstawić jako Cy gdzie C jest dowolną $K \times n$ elementową macierzą współczynników nielosowych.

Lemat 2.27 Dla wszystkich nieobciążonych estymatorów liniowych w KMRL $CX = I$.

Dowód.

$$\begin{aligned}\tilde{\beta} &= C\mathbf{y} = C(\mathbf{X}\beta + \varepsilon) \\ E(\tilde{\beta}) &= E(C\mathbf{X}\beta + C\varepsilon) = C\mathbf{X}\beta + CE(\varepsilon)\end{aligned}$$

i jest równy β wtedy i tylko wtedy gdy $C\mathbf{X} = \mathbf{I}$ ■

Dowód. (twierdzenia Gaussa-Markowa)

$$\text{Var}(\tilde{\beta}) = \text{Var}(C\mathbf{X} + C\varepsilon) = \text{Var}(C\varepsilon) = C \text{Var}(\varepsilon) C' = C\sigma^2 \mathbf{I} C' = \sigma^2 C C'$$

Zdefiniujmy takie D , że

$$D = C' - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

a więc

$$\begin{aligned}C &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' + D' \\ C\mathbf{X} = \mathbf{I} &\Leftarrow: (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{X} + D'\mathbf{X} = \mathbf{I} \Leftarrow: D'\mathbf{X} = \mathbf{0} \\ C C' &= [(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' + D'] [(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' + D']' \\ &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} + D'D = (\mathbf{X}'\mathbf{X})^{-1} + D'D \geq (\mathbf{X}'\mathbf{X})^{-1}\end{aligned}$$

Ostatnia nierówność wynika z tego, że macierz postaci $D'D$ jest dodatnio określona. W rezultacie

$$\text{Var}(\tilde{\beta}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} + \sigma^2 D'D \geq \text{Var}(\mathbf{b}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$$

Co oznacza, że estymator MNK jest rzeczywiście najlepszy (ma najmniejszą wariancję) w klasie estymatorów liniowych i nieobciążonych. ■

Uwaga 2.28 Zauważmy, że w dowodzie twierdzenia Gaussa-Markowa nie korzystamy założenia 2.24 o normalności rozkładu ε . Wynika z tego, że estymator MNK jest najlepszym liniowym i nieobciążonym estymatorem także dla ε pochodzących z innych rozkładów.

Literatura: Steward (1991) str. 43, Green (1997) str. 257-261, Chow (1995) str. 60, Goldberger (1972) str. 217-219, Theil (1979) str. 138.

2.7. Estymator nieobciążony o minimalnej wariancji dla KMRL

Załóżmy, że $\varepsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$, wtedy $\mathbf{y} \sim N(\mathbf{X}\beta, \sigma^2 \mathbf{I})$

$$p_N^n(\mathbf{y} | \mathbf{X}\beta, \sigma^2 \mathbf{I}) = (2\pi\sigma^2)^{-\frac{1}{2}n} \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)\right]$$

Spróbujmy znaleźć minimalną statystkę dostateczną dla $p_N^n(\mathbf{y} | \mathbf{X}\beta, \sigma^2 \mathbf{I})$. Iloraz prawdopodobieństw dla różnych \mathbf{y}_1 i \mathbf{y}_2 ale ustalonych \mathbf{X} , β , σ^2 wynosi

$$\frac{p_N^n(\mathbf{y}_1 | \mathbf{X}\beta, \sigma^2 \mathbf{I})}{p_N^n(\mathbf{y}_2 | \mathbf{X}\beta, \sigma^2 \mathbf{I})} = \exp\left\{-\frac{1}{2\sigma^2}[(\mathbf{y}_1 - \mathbf{X}\beta)' \mathbf{X} \mathbf{X} (\mathbf{y}_1 - \mathbf{X}\beta) - (\mathbf{y}_2 - \mathbf{X}\beta)' \mathbf{X} \mathbf{X} (\mathbf{y}_2 - \mathbf{X}\beta)]\right\}$$

Zauważmy, że

$$\begin{aligned}(\mathbf{y}_1 - \mathbf{X}_1 \beta)' (\mathbf{y}_1 - \mathbf{X}_1 \beta) &= [\mathbf{e}_1 + \mathbf{X} (\mathbf{b}_1 - \beta)]' [\mathbf{e}_1 + \mathbf{X} (\mathbf{b}_1 - \beta)] \\ &= \mathbf{e}_1' \mathbf{e}_1 + (\mathbf{b}_1 - \beta)' \mathbf{X}' \mathbf{X} (\mathbf{b}_1 - \beta) \\ &= \mathbf{e}_1' \mathbf{e}_1 + \mathbf{b}_1' \mathbf{X}' \mathbf{X} \mathbf{b}_1 + 2\mathbf{b}_1' \mathbf{X} \beta + \beta' \mathbf{X}' \mathbf{X} \beta\end{aligned}$$

a więc

$$\frac{p_N^n(\mathbf{y}_1 | \mathbf{X}\beta, \sigma^2 \mathbf{I})}{p_N^n(\mathbf{y}_2 | \mathbf{X}\beta, \sigma^2 \mathbf{I})} = \exp\left\{\mathbf{e}_1' \mathbf{e}_1 - \mathbf{e}_2' \mathbf{e}_2 + \mathbf{b}_1' \mathbf{X}' \mathbf{X} \mathbf{b}_1 - \mathbf{b}_2' \mathbf{X}' \mathbf{X} \mathbf{b}_2 + 2(\mathbf{b}_1 - \mathbf{b}_2)' \mathbf{X} \beta\right\}$$

z czego wnioskujemy, że minimalną statystyką dostateczną dla $p_N^n(\mathbf{y} | \mathbf{X}\beta, \sigma^2 \mathbf{I})$ są \mathbf{b} i $\mathbf{e}'\mathbf{e}$.

Ponieważ \mathbf{b} jest estymatorem nieobciążonym i jest funkcją minimalnej statystyki dostatecznej więc tym samym musi być estymatorem nieobciążonym o minimalnej wariancji (ENMW).

2.8. Estymator kombinacji liniowej parametrów

Wniosek 2.29 (z twierdzenia Gaussa-Markowa) Najlepszym liniowym i nieobciążonym estymatorem MNK kombinacji liniowej β o postaci $\delta' \beta$ jest esymator $\delta' \mathbf{b}$. Nieobciążonym estymatorem wariancji tego estymatora, równej $\sigma^2 \delta' (\mathbf{X}' \mathbf{X})^{-1} \delta$, jest estymator $s^2 \delta' (\mathbf{X}' \mathbf{X})^{-1} \delta$

Dowód.

$$E(\delta' \mathbf{b} - \delta' \beta) = \delta' E(\mathbf{b} - \beta) = \mathbf{0}$$

$$\begin{aligned} \text{Var}(\delta' \mathbf{b}) &= E(\delta' \mathbf{b} - \delta' \beta)(\delta' \mathbf{b} - \delta' \beta)' \\ &= E[\delta'(\mathbf{b} - \beta)(\mathbf{b} - \beta)' \delta] = \delta' E(\mathbf{b} - \beta)(\mathbf{b} - \beta)' \delta \\ &= \sigma^2 \delta' (\mathbf{X}' \mathbf{X})^{-1} \delta \end{aligned}$$

$$\begin{aligned} \text{Var}(\delta' \mathbf{C} \mathbf{y}) &= E[\delta'(\mathbf{C} \mathbf{y} - \beta)(\mathbf{C} \mathbf{y} - \beta)' \delta] \\ &= \sigma^2 \delta' (\mathbf{X}' \mathbf{X})^{-1} \delta + \sigma^2 \delta' \mathbf{D}' \mathbf{D} \delta \geq \sigma^2 \delta' (\mathbf{X}' \mathbf{X})^{-1} \delta = \text{Var}(\delta' \mathbf{b}) \end{aligned}$$

■

Wniosek 2.30 Najlepszym liniowym i nieobciążonym estymatorem β_i , to jest i -tego elementu wektora β jest b_i . Nieobciążonym estymatorem wariancji tego wektora równej $\sigma^2 (\mathbf{X}' \mathbf{X})_{ii}^{-1}$ jest estymator $S_{ii} = s^2 (\mathbf{X}' \mathbf{X})_{ii}^{-1}$, gdzie $(\mathbf{X}' \mathbf{X})_{ii}^{-1}$ oznaczamy i -ty element diagonalny odwrotności macierzy $\mathbf{X}' \mathbf{X}$.

Dowód. Wniosek udowadniamy natychmiast przyjmując, że elementy wektora δ są zdefiniowane następująco: $\delta_j = \begin{cases} 1 & \text{dla } j = i \\ 0 & \text{dla } j \neq i \end{cases}$. ■

Wniosek 2.31 Jeśli w momencie T znamy wartości jakie przyjmie x_{T+1} to najlepszą liniową i nieobciążoną prognozą wartości y_{T+1} jest $\hat{y}_{T+1} = \mathbf{x}_{T+1} \mathbf{b}$. Wariancja tej prognozy wynosi $\mathbf{x}_{T+1} \sigma^2 (\mathbf{X}'$

σ^2 . Mamy dwa źródła błędów prognozy: błędy estymacji i błąd losowy. Nieobciążonym estymatorem wariancji błędu prognozy jest $\mathbf{x}_{T+1} s^2 (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}'_{T+1} + s^2$

Dowód. Prognoza jest nieobciążona, ponieważ wartość oczekiwana błędu prognozy $e_{f,T+1} = y_{T+1} - \hat{y}_{T+1}$ wynosi $E(e_{f,T+1}) = E(y_{T+1} - \hat{y}_{T+1}) = \mathbf{x}_{T+1} \boldsymbol{\beta} - \mathbf{x}_{T+1} \boldsymbol{\beta}$. Z kolei wariancja tej błędu prognozy wynosi:

$$\begin{aligned} \text{Var}(e_{f,T+1}) &= E[(y_{T+1} - \hat{y}_{T+1})(y_{T+1} - \hat{y}_{T+1})'] = E[\mathbf{x}_{T+1}(\boldsymbol{\beta} - \mathbf{b})(\boldsymbol{\beta} - \mathbf{b})' \mathbf{x}'_{T+1}] \\ &\quad + 2E[\mathbf{x}_{T+1}(\boldsymbol{\beta} - \mathbf{b})\varepsilon_{t+1}] + E[\varepsilon_{t+1}^2] \\ &= \mathbf{x}_{T+1} \sigma^2 (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}'_{T+1} + \sigma^2 \end{aligned}$$

przy czym skorzystaliśmy z tego, że błąd estymacji $\boldsymbol{\beta} - \mathbf{b}$ zależy ε a ponieważ ε i ε_{T+1} są niezależne więc $E[(\boldsymbol{\beta} - \mathbf{b})\varepsilon_{t+1}] = 0$. ■

Literatura: Goldberger (1972) str. 222-227, Theil (1979) str. 139-142.

2.9. Przekształcenia zmiennych i MNK

Metoda Najmniejszych Kwadratów stosowana jest do estymacji modeli liniowych. Jedną z istotnych kwestii, którą powinniśmy rozważyć jest, pytanie co właściwie rozumiemy przez model liniowy. Okazuje się, że wiele z modeli nieliniowych można przekształcić do postaci modelu liniowego względem przekształconych zmiennych.

Definicja 2.32 Model $y_i = f(\mathbf{x}_i, \boldsymbol{\theta})$ można sprowadzić do modelu liniowego względem przekształconych zmiennych jeśli istnieją takie funkcje $g(\mathbf{x}_i, \boldsymbol{\theta})$, $\mathbf{h}(\mathbf{x}_i, \boldsymbol{\theta})$ i $\boldsymbol{\beta}(\boldsymbol{\theta})$, że

$$g[f(\mathbf{x}_i, \boldsymbol{\theta})] = \boldsymbol{\beta}(\boldsymbol{\theta}) \mathbf{h}(\mathbf{x}_i) \text{ dla } i = 1, \dots, N$$

i funkcja $\boldsymbol{\beta}(\boldsymbol{\theta})$ jest wzajemnie jednoznaczna.

Omówimy teraz najważniejsze modele, które można sprowadzić do liniowości.

• Model logarytmiczno-liniowy

Model logarytmiczno liniowy ma następującą formę

$$Y_i = AX_{1i}^{\beta_1} X_{2i}^{\beta_2} \dots X_{ki}^{\beta_k} e^{\varepsilon_i}$$

Logarytmując obie strony uzyskujemy

$$\ln Y_i = \ln A + \beta_1 \ln X_{1i} + \ln \beta_2 X_{2i} + \dots + \ln \beta_k X_{ki} + \varepsilon_i$$

dla przekształconych zmiennych $y_i = \ln Y_i$, $x_{1i} = \ln(X_{1i})$, $x_{2i} = \ln(X_{2i})$, ..., $x_{ki} = \ln(X_{ki})$ oraz przekształconego parametru $\beta_0 = \ln A$ uzyskujemy model liniowy względem przekształconych parametrów:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i$$

Ważną dla interpretacji modelu logarytmiczno liniowego cechą jest fakt, że parametry w tym modelu mogą być interpretowane jako elastyczności. Dla uproszczenia oznaczmy $\tilde{Y}_i = E(Y_i | \mathbf{X}_i)$.

$$\frac{\partial E(Y_i | \mathbf{X}_i)}{\partial X_{si}} = \frac{\partial \tilde{Y}_i}{\partial X_{si}} = \beta_s AX_{1i}^{\beta_1} \dots X_{si}^{\beta_s - 1} \dots X_{ki}^{\beta_k} = \beta_s \frac{\partial \tilde{Y}_i}{X_{si}}$$

$$\beta_s = \frac{\partial \tilde{Y}_i}{\partial X_{si}} \bigg/ \frac{\tilde{Y}_i}{X_{si}} = \frac{d\tilde{Y}_i}{\tilde{Y}_i} \bigg/ \frac{dX_{si}}{X_{si}}$$

W przypadku modelu logarytmiczno liniowego elastyczności są więc stałe i równe współczynnikom przy zmiennych niezależnych. Wynik ten można uzyskać w prostszy sposób, korzystając z tego, że

$$\frac{\partial \ln Y}{\partial \ln X} = \frac{\partial Y}{\partial X} \frac{\partial \ln Y}{\partial X} \left(\frac{\partial \ln X}{\partial X} \right)^{-1} = \frac{\partial Y}{\partial X} \frac{X}{Y} = \frac{dY}{Y} \bigg/ \frac{dX}{X} \quad (2.14)$$

Wynika z niego, że istotnie jeśli model jest staję się liniowy po przekształceniu zmiennych na ich logarytmy, to elastyczności muszą być stałe.

W modelu logarytmiczno normalnym szacowanym dla szeregów czasowych często pojawia się także trend. W takim przypadku model ma najczęściej postać:

$$Y_t = AX_{1t}^{\beta_1} X_{2t}^{\beta_2} \dots X_{kt}^{\beta_k} e^{\gamma t} e^{\varepsilon_i}$$

do więc po przekształceniu

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_k x_{kt} + \gamma t + \varepsilon_i$$

W takim przypadku γ można interpretować, jako średni procentowy wzrost y_t w jednostce czasu:

$$\begin{aligned} \frac{\partial \tilde{Y}_t}{\partial t} &= \gamma E \left(AX_{1t}^{\beta_1} X_{2t}^{\beta_2} \dots X_{kt}^{\beta_k} e^{\gamma t} e^{\varepsilon_i} \right) = \gamma \tilde{Y}_t \\ \gamma &= \frac{\partial \tilde{Y}_t}{\partial t} \bigg/ \tilde{Y}_t = \frac{d\tilde{Y}_t}{\tilde{Y}_t} \bigg/ dt \end{aligned}$$

Model logarytmiczno liniowy jest szczególnie przydatny do szacowania zależności, w przypadku których można sądzić, że elastyczności zmiennej zależnej względem zmiennych niezależnych pozostają w przybliżeniu stałe w zakresie wielkości zmiennych, które znalazły się w naszej próbie.

• Model semilogarytmiczny

Model powstaje po przekształceniu modelu

$$y_t = e^{\beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_k x_{kt} + \gamma t + \varepsilon_i}$$

do modelu

$$\ln y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_k x_{kt} + \gamma t + \varepsilon_i$$

Model jest często używany w przypadku modelowania wzrostu. Podobnie jak w modelu logarytmiczno liniowym γ można interpretować jako średni procentowy wzrost y_t . Analogicznie

$$\frac{\partial \tilde{y}_t}{\partial x_{st}} = \beta_s E \left(e^{\beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_k x_{kt} + \gamma t + \varepsilon_i} \right) = \beta_s \tilde{y}_t$$

$$\beta_s = \frac{\partial \tilde{y}_t}{\partial x_{st}} \bigg/ \tilde{y}_t = \frac{d \tilde{y}_t}{\tilde{y}_t} \bigg/ dx_{st}$$

a więc β_s można interpretować jako współczynnik, który mówi o ile procentowo wzrośnie y_t jeśli x_{st} wzrośnie o jednostkę.

• Model z interakcjami

Model liniowy w swojej najprostszej postaci może okazać się niewystarczający jeśli siła oddziaływania jednej zmiennej zależy od innych zmiennych. Na przykład w modelu wyjaśniającym wielkość wydatków konsumpcyjnych reakcja popytu na zmianę cen zależy nie tylko od cen danego dobra ale i od wielkości spożycia i cen dla innych dóbr. Po to by uwzględnić w modelu tego typu efekty możemy do modelu wprowadzić poza zmiennymi także ich iloczyny:

$$y_i = \beta_1 x_{1i} + \dots + \beta_k x_{ki} + \sum_{r=1}^K \sum_{s=1}^r \gamma_{rs} x_r x_s + \varepsilon_i$$

Wielkość reakcji \tilde{y}_i na x_i dane jest wzorem

$$\frac{\partial \tilde{y}_i}{\partial x_{si}} = \beta_s + \sum_{r=1}^K \gamma_{ri} x_{ri}$$

i jest zależna od wielkości pozostałych zmiennych niezależnych. Model z interakcjami może służyć do przybliżenia nieliniowej między zmiennymi. Jak wiemy z twierdzenia Taylora (??), każdą dwukrotnie różniczkowalną funkcję można przybliżyć za pomocą funkcji kwadratowej.

- **Model translogarytmiczny**

Specjalnym typem modelu z interakcjami jest model translogarytmiczny, który jest modelem logarytmiczno liniowym z dodanymi interakcjami:

$$y_i = Ax_{1i}^{\beta_1} x_{2i}^{\beta_2} \dots x_{ki}^{\beta_k} \exp \left(\sum_{r=1}^K \sum_{s=1}^r \gamma_{rs} \ln x_{ri} \ln x_{si} \right) e^{\varepsilon_i}$$
$$\ln y_i = \beta_1 \ln x_{1i} + \dots + \beta_k \ln x_{ki} + \sum_{r=1}^K \sum_{s=1}^r \gamma_{rs} \ln x_{ri} \ln x_{si} + \varepsilon_i$$

W przeciwieństwie do modelu logarytmiczno liniowego w modelu translogarytmicznym, elastyczności nie są stałe

$$\frac{\partial \ln \tilde{y}_i}{\partial \ln x_{si}} = \beta_s + \sum_{r=1}^K \gamma_{rs} \ln x_{ri}$$

- **Model logistyczny**

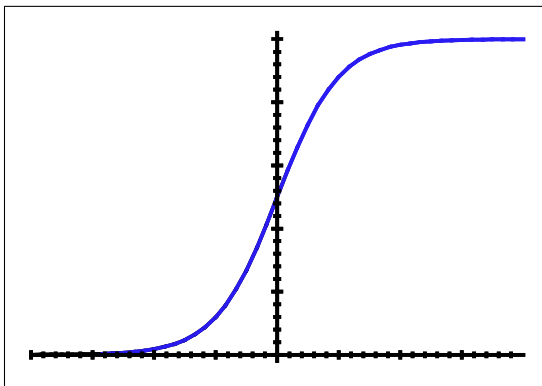
Model logistyczny stosuje się do opisu zjawisk, które charakteryzują się szybkim niskim wzrostem dla małych wartości zmiennej zależnej, szybkim dla pośrednich jej wartości i znowu niskim dla wysokich wartości zmiennej. Dla opisu tego typu zależności nadaje się funkcję logistyczną postaci

$$f(x) = \frac{1}{1 + e^{-\beta x}}$$

Funkcja $f(x)$ jest zawsze dodatnia. Jej pochodna

$$f'(x) = \frac{\beta e^{-\beta x}}{(1 + e^{-\beta x})^2}$$

jest zawsze dodatnia i ponadto $\lim_{x \rightarrow -\infty} = \lim_{x \rightarrow \infty} = 0$. Wykres tego rodzaju funkcji znajduje się na rysunku (5)



Rysunek 5: Krzywa logistyczna

Model ekonometryczny który zawiera więcej niż jedną zmienną i w którym uwzględniamy wpływ zaburzeń losowych wyglądać będzie następująco:

$$y_i = \frac{1}{1 + e^{-x_i\beta - \varepsilon_i}}$$

Po przekształceniu

$$\frac{1 - y_i}{y_i} = e^{-x_i\beta - \varepsilon_i}$$

i zlogarytmowaniu stronami, uzyskujemy model liniowy względem przekształconych zmiennych:

$$y_t^* = \ln \frac{y_t}{1 - y_t} = \mathbf{x}_t \boldsymbol{\beta} - \varepsilon_t$$

• Model przedziałami liniowy

W niektórych przypadkach okazuje się, że najłatwiej przybliżyć jest nieliniową zależność za pomocą funkcji, która jest przedziałami liniowa

$$y_i = \begin{cases} \alpha_1 + \beta_1 x_i & \text{dla } x_i \leq x_1^* \\ \alpha_2 + \beta_2 x_i & \text{dla } x_1^* < x_i \leq x_2^* \\ \vdots & \vdots \\ \alpha_s + \beta_s x_i & \text{dla } x_i > x_s^* \end{cases}$$

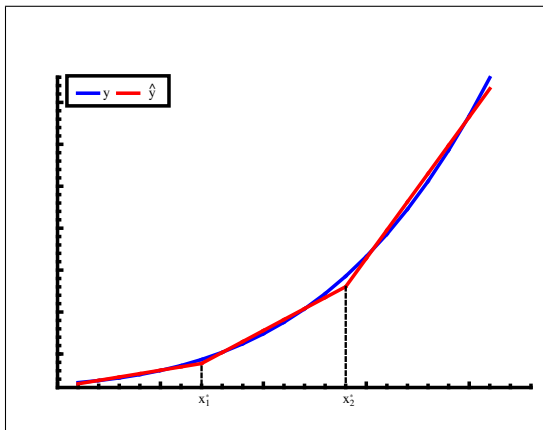
Najczęściej chcemy jednak, by funkcja ta była także funkcją ciągłą. Z tego powodu najczęściej używanym formą takiego modelu jest:

$$y_i = \begin{cases} \alpha + \beta_1 x_i & \text{dla } x_i \leq x_1^* \\ \alpha + \beta_1 x_1^* + \beta_2 (x_i - x_1^*) & \text{dla } x_1^* < x_i \leq x_2^* \\ \vdots & \vdots \\ \alpha + \beta_1 x_1^* + \sum_{j=1}^{s-1} \beta_r (x_j^* - x_{j-1}^*) + \beta_s x_i & \text{dla } x_i > x_s^* \end{cases}$$

Przykładowe dopasowanie tego typu funkcji do pewnej nieliniowej funkcji regresji pokazane jest na rysunku (6). Z reguły liczby x_1^*, \dots, x_s^* nazywane węzłami, są ustalane arbitralnie i mogą być traktowane jako nielosowe. Jeśli zdefiniujemy teraz zmienne zerojedynkowe² $d_j = I(x_i > x_j^*)$

$$y_i = \alpha + \beta_1 x_i + \sum_{j=2}^s \beta_j d_j (x_i - x_j^*)$$

²Zmienne zerojedynkowe omawiamy szczegółowo w podrozdziale 2.10.



Rysunek 6: Regresja odcinkami liniowa

Taka forma modelu na przekształconych zmiennych jest wygodniejsza do estymacji, ponieważ automatycznie uwzględnia ograniczenia wynikające z założonej ciągłości funkcji.

- **Przekształcenie Boxa-Coxa**

Jednym z częściej stosowanych w ekonometrii przekształceń zmiennych jest przekształcenie Boxa-Coxa:

$$x^{(\lambda)} = g(x, \lambda) = \frac{x^\lambda - 1}{\lambda}.$$

Główną teoretyczną zaletą tego przekształcenia jest to, że $g(x, 1) = x$ a $\lim_{\lambda \rightarrow 1} g(x, \lambda) = \ln x$. W rezultacie model postaci

$$y_i^{(\lambda)} = \beta_1 x_1^{(\lambda)} + \dots + \beta_k x_k^{(\lambda)} + \varepsilon_i$$

sprowadza się do modelu liniowego dla $\lambda = 1$ a do modelu logarytmiczno liniowego dla $\lambda = 1$. Model ten ma postać modelu liniowego dla dowolnego znanego λ . Jeśli jednak λ jest nieznanne, to model ten jest modelem nieliniowym.

2.10. Regresory zero-jedynkowe

Regresory zero-jedynkowe są to zmienne egzogeniczne, które opisują zjawiska w kategoriach jakościowych a nie ilościowych.

Przykład 2.33 *Typową zmienną zero-jedynkową jest zmienna oznaczająca płeć respondenta:*

$$x_{\sigma,i} = \begin{cases} 0 & \text{kobieta} \\ 1 & \text{mężczyzna} \end{cases}$$

$$x_{\varphi,i} = \begin{cases} 1 & \text{kobieta} \\ 0 & \text{mężczyzna} \end{cases}$$

Tak więc prostym modelem uzależniającym wzrost respondenta od jego płci

$$y_i = \beta_{\sigma} x_{\sigma,i} + \beta_{\varphi} x_{\varphi,i} + \varepsilon_i$$

Ponieważ $E(y_i | x_{\sigma,i} = 1) = \beta_{\sigma} \times 1 + \beta_{\varphi} \times 0 + E(\varepsilon_i) = \beta_{\sigma}$, więc β_{σ} określa wartość oczekiwaną wzrostu mężczyzny a β_{φ} kobiety. Hipotezę, że oczekiwany wzrost mężczyzny jest różny od oczekiwanego wzrostu kobiety można sformułować jako $H_0 : \beta_{\sigma} \neq \beta_{\varphi}$. Przy innym zdefiniowaniu zmiennej hipoteza ma nieco prostszą formę. Zauważmy, że

$$y_i = \beta_{\sigma} (x_{\sigma,i} + x_{\varphi,i}) + (\beta_{\varphi} - \beta_{\sigma}) x_{\varphi,i} + \varepsilon_i = \beta_{\sigma} + \beta_{\varphi}^* x_{\varphi,i} + \varepsilon_i$$

ponieważ z definicji $x_{\sigma,i} + x_{\varphi,i} = 1$, gdzie

$$\beta_{\varphi}^* = \beta_{\varphi} - \beta_{\sigma}.$$

Hipotezę o równym oczekiwanym wzroście można teraz zdefiniować jako hipotezę $H_0 : \beta_{\varphi}^* = 0$ i przetestować za pomocą statystyki t . W tak zdefiniowanym modelu, β_{φ}^* ma interpretację różnicy między oczekiwanym wzrostem kobiet i oczekiwaną wartością zmiennej bazowej, którą jest wzrost mężczyzn.

Nasze wnioski można uogólnić na przypadek, w którym zmienna jakościowa może przyjąć więcej możliwych poziomów. Załóżmy, że zmienna x_i może przyjmować jeden z S możliwych poziomów $x_i \in \{1, 2, \dots, S\}$. Wtedy możemy zdefiniować s zmiennych zero- jedynkowych, takich że

$$x_{s,i} = \begin{cases} 1 & x_{s,i} = s \\ 0 & x_{s,i} \neq s \end{cases} \quad \text{dla } s = 1, 2, \dots, S,$$

gdzie $\sum_{s=1}^S x_{s,i} = 1$ dla $i = 1, 2, \dots, n$. Model ma teraz postać

$$y_i = \mathbf{z}_i \boldsymbol{\gamma} + \sum_{s=1}^S x_{s,i} \beta_s + \varepsilon_i,$$

gdzie \mathbf{z}_i zawiera inne zmienne istotne dla zachowania modelu przy czym \mathbf{z}_i nie może zawierać stałej. Jeśli przeddefiniujemy zmienne w podobny sposób co poprzednio przyjmując pewien poziom k za bazowy, to model będzie miał postać

$$y_i = \mathbf{z}_i \boldsymbol{\gamma} + \beta_k + \sum_{\substack{s=1 \\ s \neq k}}^S x_{s,i} \beta_s^* + \varepsilon_i \quad (2.15)$$

gdzie $\beta_s^* = \beta_s - \beta_k$ interpretujemy jako różnice między wartością oczekiwaną y_i jeśli $x_i = s$ a wartością oczekiwaną y_i gdy $x_i = k$. Na koniec zauważmy, że model, występują zmienne zero-

jedynek nie może zawierać jednocześnie wszystkich poziomów zmiennej i stałej. Jeśli sformułowalibyśmy model

$$y_i = z_i \gamma + \alpha + \sum_{s=1}^S x_{s,i} \beta_s + \varepsilon_i$$

to między wektorem jedynek związanym ze stałą i zmiennymi zero-jedynekowymi wystąpiłaby dokładna współliniowość, ponieważ $\sum_{s=1}^S x_{s,i} = 1$. W rezultacie, kiedy rozbijamy zmienną jakościową na zmienne zero-jedynekowe to albo mamy w modelu osobną zmienną dla każdego poziomu i wtedy w modelu nie może być stałej, albo jeden z poziomów wybieramy jako bazowy, i wtedy dla tego poziomu nie ma zmiennej zero-jedynekowej ale w modelu jest stała.

Specyficznym przykładem zmiennych zero-jedynekowych są zmienne sezonowe. Tworzymy je gdy w szeregu czasowym występuje zależność między zachowaniem porą roku a poziomem zmiennej objaśnianej (n.p. w przypadku wydatków konsumpcyjne wiadomo, że są one średnio wyższe w grudniu niż w pozostałych miesiącach). Pora roku może być opisana zmienną jakościową, której ilość poziomów zależeć będzie od częstotliwości obserwacji. Przykładowo dla danych kwartalnych $S = 4$, gdzie np. $x_{4,i} = 1$ oznacza, że obserwacja i dotyczy pierwszego kwartału. Sposób postępowania z tymi zmiennymi jest identyczny jak w przypadku standardowych zmiennych zero-jedynekowych.

Literatura: Steward (1991) str. 97-101, Green (1997) str. 379-390, Goldberger (1972) str. 290-294.

• Analiza wariancji (ANOVA)

Analiza wariancji (**An**alysis of **V**ariance) ma za zadanie rozstrzygnięcie pytania, czy dzieląc próbe na klasy możemy wyjaśnić część zmienności obserwacji. Modele ANOVA można traktować jako szczególny przypadek modelu liniowego, w którym występujące zerojedynekowe zmienne objaśniające można interpretować jako definiujące podział populacji na klasy. Klasyfikację taką będziemy

uważać za poprawną jeśli obserwacje wewnątrz klas (grup) są do siebie podobne i różnią się istotnie od obserwacji w innych klasach.

Przykład 2.34 *Naszym zadaniem jest wyjaśnienie nierówności dochodów pracowniczych. Jedną z możliwych hipotez jest, że ważną przyczyną nierówności dochodów jest zróżnicowanie poziomu wykształcenia między pracownikami. Jeśli dochody osób o tym samym wykształceniu (np wyższym) są do siebie podobne i różnią się istotnie od dochodów osób o innych poziomach wykształcenia (np niepełnym podstawowym lub średnim) to klasyfikacja tą możemy uważać za poprawną.*

Zajmiemy się tutaj najprostrzym przypadkiem, kiedy analizujemy zachowanie jednej zmiennej. W tej sytuacji podobieństwo obserwacji wewnątrz i między klasami można analizować analizując odchylenia obserwacji na tej zmiennej od średniej. Załóżmy, że mamy K klas i wewnątrz każdej klasy j mamy n_j obserwacji. Całkowita ilość obserwacji wynosi $n = \sum_{j=1}^K n_j$. Obserwację i -tą w j -tej klasie będziemy oznaczać jako y_{ij} . Oznaczmy jako

- $\bar{y}_{\bullet\bullet} = n^{-1} \sum_{j=1}^K \sum_{i=1}^{n_j} y_{ij}$ średnią w populacji
- $\bar{y}_{\bullet j} = n_j^{-1} \sum_{i=1}^{n_j} y_{ij}$ średnią w klasie j

Całkowitą zmienność obserwacji w populacji definiujemy jako sumę kwadratów odchyłeń zmiennej od średniej w populacji. Zauważmy, że

$$\underbrace{\sum_{j=1}^K \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_{\bullet\bullet})^2}_{\text{Całkowita zmienność}} = \underbrace{\sum_{j=1}^K n_j (\bar{y}_{\bullet j} - \bar{y}_{\bullet\bullet})^2}_{\text{Zmienność międzygrupowa}} + \underbrace{\sum_{j=1}^K \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_{\bullet j})^2}_{\text{Zmienność wewnątrzgrupowa}} \quad (2.16)$$

Dowód.

$$\begin{aligned}\sum_{j=1}^K \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_{\bullet\bullet})^2 &= \sum_{j=1}^K \sum_{i=1}^{n_j} [(\bar{y}_{\bullet j} - \bar{y}_{\bullet\bullet}) + (y_{ij} - \bar{y}_{\bullet j})]^2 \\ &= \sum_{j=1}^K n_j (\bar{y}_{\bullet j} - \bar{y}_{\bullet\bullet})^2 + \sum_{j=1}^K \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_{\bullet j})^2 \\ &\quad + \sum_{j=1}^K \sum_{i=1}^{n_j} (\bar{y}_{\bullet j} - \bar{y}_{\bullet\bullet}) (y_{ij} - \bar{y}_{\bullet j})\end{aligned}$$

ale

$$\sum_{j=0}^K \sum_{i=1}^{n_j} (\bar{y}_{\bullet j} - \bar{y}_{\bullet\bullet}) (y_{ij} - \bar{y}_{\bullet j}) = \sum_{j=0}^K \left[(\bar{y}_{\bullet j} - \bar{y}_{\bullet\bullet}) \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_{\bullet j}) \right] = 0$$

ponieważ $\sum_{i=1}^{n_j} (y_{ij} - \bar{y}_{\bullet j}) = n_j \bar{y}_{\bullet j} - n_j \bar{y}_{\bullet j} = 0$. ■

Klasyfikacja obserwacji będzie tłumaczyć zmienność obserwacji jeśli istotna część zmienności całkowitej będzie wyjaśniona przez zmienność międzygrupową. Po to, by ustalić kiedy istotna statystycznie część zmienności wyjaśniona jest za pomocą klasyfikacji należy sformułować model statystyczny, który wyjaśnia w jaki sposób generowane są obserwacje. Załóżmy, że w grupie j obserwacja i generowana w następujący sposób

$$y_{ij} = \beta_j + \varepsilon_{ij}, \quad j = 1, \dots, K$$

gdzie $\varepsilon_{ij} \sim N(0, \sigma^2)$. Zauważmy, że w tym miejscu robimy arbitralne założenie, że wariancja błędu losowego wewnątrz różnych grup jest taka sama. W świetle tego co powiedziano wcześniej model ten można zapisać także jako model liniowy ze zmiennymi zerojedynkowymi:

$$y_i = \sum_{j=1}^K \beta_j x_{ij} + \varepsilon_i, \quad i = 1, \dots, n$$

przy czym $x_i = 1$, gdy obserwacja i należy do j -tej klasy a $x_i = 0$ dla pozostałych przypadków. Przy takim zapisie widać, że wzór (2.16) jest szczególnym przypadkiem wzoru (2.5) przy czym TSS jest równa całkowitej zmienności ESS zmienności międzygrupowej a RSS zmienności wewnątrzgrupowej. Jeśli klasyfikacji nie ma żadnej mocy objaśniającej to prawdziwa jest hipoteza $H_0 : \beta_1 = \dots = \beta_K$ lub prz zapisie (2.15) $H_0 : \beta_1 = \dots = \beta_i = \dots = \beta_K = 0$, gdzie $i \neq k$. Hipotezę tą można przetestować za pomocą statystyki sformułowanej we wniosku 3.10.

Analiza wariancji ulega znacznemu skomplikowaniu jeśli klasyfikacja dokonywana jest według więcej niż jednego kryterium. Należy wtedy przeanalizować wpływ nie tylko każdego z grupowań z osobna ale także interakcje między grupowaniami.

Przykład 2.35 *Po to, by wyjaśnić różnicownie dochodów bierzemy pod uwagę nie tylko wykształcenie pracowników jak w przykładzie 2.34 ale także ich płeć. W przypadku, kiedy między grupowaniami mogą występować interakcje powinniśmy zdefiniować osobną zmienną zerojedynkową dla każdej kombinacji charakterystyk. W poniższej tabeli pokazano jak można zakodować zmienne w omawianym przykładzie.*

podstawowe	średnie	wyższe	
z_{1i}	z_{3i}	z_{5i}	kobieta
z_{2i}	z_{4i}	z_{6i}	mężczyzna

Przy takim kodowaniu model będzie miał postać

$$y_i = \beta_1 z_{1i} + \beta_2 z_{2i} + \beta_3 z_{3i} + \beta_4 z_{4i} + \beta_5 z_{5i} + \beta_6 z_{6i} + \varepsilon_i \quad (2.17)$$

Hiptezę o tym, że między wpływem wykształcenia i płci nie zachodzą interakcje, to dodatkowy dochód wynikający z uzyskania lepszego wykształcenia będzie taki sam dla kobiet i mężczyzn. Hipoteza o braku interakcji między efektem dochodowym płci i wykształcenia można sformułować jako hipotezę złożoną postaci:

$$H_0 : \begin{cases} \beta_1 - \beta_3 = \beta_2 - \beta_4 \\ \beta_3 - \beta_5 = \beta_4 - \beta_6 \end{cases}$$

Oznacza ona, że średnia różnica między zarobkami kobiety o wykształceniu podstawowym i kobiety o wykształceniu średnim powinna być taka sama, jak różnica między średnimi zarobkami mężczyzny o wykształceniu podstawowym i mężczyzny o wykształceniu średnim oraz podobnie w przypadku różnic między dochodami osób o wykształceniu średnim i wyższym. Jeśli jest to prawdą to mówimy, że wpływ płci i wykształcenia na zarobki jest addytywny.

Przypadek występowania interakcji można przeanalizować także w kontekście wcześniejszych rozważań na temat modeli z interakcjami. Jeśli zdefiniujemy tak jak wcześniej zmienne $x_{\sigma,i}$, $x_{\sigma,i}$, oraz zmienne x_{P_i} , x_{S_i} , x_{W_i} oznaczające, że respondent uzyskał odpowiednio wykształcenie podstawowe, średnie i wyższe, to równoważny do poprzedniej formy modelu z interakcjami będzie model postaci

$$y_i = \beta_1 + \beta_{\sigma}x_{\sigma,i} + \beta_Sx_{S_i} + \beta_Wx_{W_i} + \beta_{\sigma S}x_{S_i}x_{\sigma,i} + \beta_{\sigma W}x_{W_i}x_{\sigma,i} \quad (2.18)$$

a poprzednio sformułowana hipoteza będzie miała teraz postać

$$H_0 : \begin{cases} \beta_4^* = 0 \\ \beta_5^* = 0 \end{cases}$$

przy czym interpretacja współczynników jest następująca: β_1 oznacza średni dochód kobiety o wykształceniu podstawowym (poziom bazowy), β_{σ} jest różnicą między tym dochodem a dochodem mężczyzny o wykształceniu podstawowym, β_S jest różnicą między dochodem bazowym i średnim dochodem kobiety o wykształceniu średnim, β_W jest jest analogiczną różnicą w przypadku kobiety o wykształceniu wyższym, zaś $\beta_{\sigma S}$ i $\beta_{\sigma W}$ są różnicami między przyrostami dochodu związanymi ze zdobyciem danego poziomu wykształcenia między kobietami i mężczyznami.

Poza zerojedynkowymi zmiennymi objaśniającymi w modelach takich mogą występować też zmienne ciągłe.