

4. Problemy związane z danymi

Jednym z niezwykle ważnych etapów badania ekonometrycznego jest weryfikacja poprawności i jakości danych. Bardzo często błędne wyniki badania ekonometrycznego są skutkiem niewielkiej liczby błędnych obserwacji. Ważnym etapem badania ekonometrycznego jest więc weryfikacja danych pod kątem ich poprawności i logicznej spójności. W danych pojawiać się też mogą nietypowe obserwacje, które silnie wpływają na wynik badania a są efektem zjawisk wykraczających poza te, które chcemy wyjaśnić naszym modelem. Często ten wstępny etap badania statystycznego polegający na weryfikacji i czyszczeniu bazy danych należy do najbardziej pracochłonnych a równocześnie najważniejszych dla uzyskania sensownych wyników.

Własności zbioru danych mogą także spowodować trudności w identyfikacji wpływu poszczególnych zmiennych objaśniających na analizowaną zmienną objaśnianą. Dzieje się tak wtedy, gdy zmienne objaśniające są silnie skorelowane. Powiedzmy, że x_1 i x_2 rosną "naogół" razem. Jeśli równocześnie rośnie y , to pojawia się pytanie, czy wzrost y jest związany ze wzrostem x_1 , czy też x_2 . W przypadku silnej korelacji x_1 i x_2 udzielenie odpowiedzi na tak postawione pytanie może okazać się bardzo trudne. Taką silną korelację między zmiennymi nazywamy współliniowością. Często niewiele można zrobić by rozwiązać problemy wynikające ze współliniowości zmiennych, niemniej samo wykrycie występowania współliniowości jest ważne dla interpretacji wyników badania.

Rozważania w tym rozdziale będą miały z konieczności miły charakter nieformalny, ponieważ naogół nie da się sformułować formalnego modelu, który mógłby posłużyć do statystycznej analizy problemów związanych z niedoskonałością zbioru danych. Podane reguły mają charakter zdroworozsądkowy, lub też są "ogólnie przyjętą praktyką".

4.1. Obserwacje nietypowe - statystyki diagnostyczne

Badając dane pod kątem istnienia w nich obserwacji błędnych lub nietypowych powinniśmy szczególną uwagę zwracać na obserwacje, które w silny sposób wpływają na wynik estymacji. W celu wyodrębnienia tego typu obserwacji zaproponowano kilka statystyk. Jedną z takich kluczowych takich statystyk jest *dźwignia* zdefiniowana jako:

$$h_i = (\mathbf{P}_X)_{ii} = \boldsymbol{\delta}'_i \mathbf{P}_X \boldsymbol{\delta}_i = \boldsymbol{\delta}'_i \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\delta}_i = \mathbf{x}_i (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}'_i$$

Statystyka ta posłuży nam do zdefiniowania kilku innych łatwiej interpretowalnych statystyk. Zajmiemy się najpierw własnościami h_i . Z formy macierzy \mathbf{P}_X wynika, że

$$\sum_{i=1}^n h_i = \text{tr}(\mathbf{P}_X) = \text{tr}[\mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}'] = \text{tr} \mathbf{I}_k = k$$

h_t jest formę kwadratową:

$$h_t = \boldsymbol{\delta}'_i \mathbf{C} \boldsymbol{\Lambda} \mathbf{C}' \boldsymbol{\delta}_i = \boldsymbol{\alpha}' \boldsymbol{\Lambda} \boldsymbol{\alpha} \geq 0$$

gdzie $\boldsymbol{\alpha} = \mathbf{C}' \boldsymbol{\delta}_i$ i $\boldsymbol{\alpha}' \boldsymbol{\Lambda} \boldsymbol{\alpha} \geq 0$ ponieważ \mathbf{P}_X nieujemnie określona. Skoro

$$\boldsymbol{\alpha}' \boldsymbol{\alpha} = \boldsymbol{\delta}'_i \mathbf{C} \mathbf{C}' \boldsymbol{\delta}_i = \boldsymbol{\delta}'_i \boldsymbol{\delta}_i = 1$$

więc

$$\|\boldsymbol{\Lambda} \boldsymbol{\alpha}\|^2 = \sum_{i=1}^n \lambda_i^2 \alpha_i^2 \leq \sum_{i=1}^n \alpha_i^2 = \|\boldsymbol{\alpha}\|^2 = 1$$

ponieważ dla macierzy idempotentnej $\lambda_i = 1$ lub $\lambda_i = 0$. Wynika z tego, że h_i musi spełniać:

$$0 \leq h_i \leq 1$$

Jesli w modelu jest stała wtedy z (2.12) wynika, że P_X można zdekomponować w następujący sposób

$$h_i = \delta_i' P_X \delta_i = \delta_i' [P_{X^*} + P_l] \delta_i = \delta_i' P_{X^*} \delta_i + \frac{1}{n} \quad (4.1)$$

gdzie P_{X^*} jest macierzą rzutów dla X^* zawierającego odchylenia kolumn macierzy X od swoich średnich, a P_l macierzą rzutów dla wektora jedynek l . Ponieważ P_{X^*} jest nieujemnie określone, więc $\delta_i' P_{X^*} \delta_i \geq 0$ i w tym ważnym przypadku

$$\frac{1}{n} \leq h_i \leq 1$$

Dźwignia jest w gruncie rzeczy miarą nietypowości obserwacji. Można to łatwiej zrozumieć wprowadzając pojęcie *odległości Mahalanobisa* dla obserwacji i

$$MD_i^2 = (\mathbf{x}_i - \bar{\mathbf{x}})' \Sigma_X^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}) = \delta_i' P_{X^*} \delta_i$$

gdzie $\Sigma_X = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}}) (\mathbf{x}_i - \bar{\mathbf{x}})'$. Odległość Mahalanobisa jest miarą odchylenia \mathbf{x}_i od jego średniej, zważone przez empiryczną macierz wariancji kowariancji \mathbf{x} . Z równania (4.1) wynika, że

$$h_i = \frac{1}{n-1} MD_i^2 + \frac{1}{n}$$

Dźwignia jest więc miarą, na którą wpływa nietypowość danej obserwacji oraz wielkość próby. Ponieważ dla dużych prób pojedyncze obserwacje mają mniejsze znaczenie dla końcowego wyniku estymacji więc dźwignia dla poszczególnych obserwacji w takich próbach będzie mniejsza. Literatura przedmiotu sugeruje, że bliższemu zbadaniu należy poddać obserwacje, dla których $h_i > \frac{2k}{n}$ i takie, dla których $MD_i > \chi_{k-1,0.95}^2$.

Pierwszą statystyką jaką zdefiniujemy przy pomocy dźwigni są wystandaryzowane reszty:

$$\hat{e}_i = \frac{e_i}{\sqrt{\text{Var}(e_i)}}$$

Wariancję e_i liczymy w następujący sposób

$$\begin{aligned}\text{Var}(\mathbf{e}) &= \text{Var}(\mathbf{M}_X \boldsymbol{\varepsilon}) = \mathbf{M}_X (\mathbf{I} \sigma^2) \mathbf{M}_X = \sigma^2 \mathbf{M}_X \\ \text{Var}(e_i) &= \text{Var}(\boldsymbol{\delta}'_i \mathbf{e}) = \sigma^2 \boldsymbol{\delta}'_i \mathbf{M}_X \boldsymbol{\delta}_i = \sigma^2 (1 - \boldsymbol{\delta}'_i \mathbf{P}_X \boldsymbol{\delta}_i) = \sigma^2 (1 - h_i)\end{aligned}$$

a więc

$$\hat{e}_i = \frac{e_i}{\sigma \sqrt{1 - h_i}}$$

zwykle zastępujemy σ przez estymator $\hat{\sigma}$ równy $\sqrt{s^2_{(i)}}$, gdzie $s^2_{(i)}$ jest estymatorem s^2 wyestymowanym na podstawie wszystkich obserwacji poza obserwacją i -tą. W literaturze przedmiotu sugeruje się, że te obserwacje, dla których $|\hat{e}_i| > 2$ wymagają bliższego zbadania.

Jednym z możliwych sposobów wykrycia obserwacji nietypowych jest zwrócenie uwagi na te obserwacje, które istotnie wpływają na wyestymowaną macierz wariancji kowariancji wektora losowego. Po to by stworzyć syntetyczną miarę zmiany macierzy wariancji kowariancji na skutek dodania obserwacji i możemy użyć wyznacznika tej macierzy. W *MNK* estymator macierzy wariancji kowariancji \mathbf{b} jest dany wzorem $\mathbf{S} = s^2 (\mathbf{X}' \mathbf{X})^{-1}$. Zaproponowano następującą miarę wpływu i -tej obserwacji na wyestymowaną macierz wariancji kowariancji:

$$COVRATIO_i = \frac{\det(\mathbf{S}_{(i)})}{\det(\mathbf{S})}$$

gdzie $\mathbf{S}_{(i)}$ jest estymatorem \mathbf{S} policzony na podstawie wszystkich obserwacji poza obserwacją i -tą. Literatura przedmiotu sugeruje, że należy zwrócić uwagę na obserwacje, dla których $|COVRATIO_{(i)} - \frac{3p}{N}|$

Jak już wcześniej wspomnieliśmy, najbardziej interesujące są dla nas te obserwacje, które istotnie wpływają na wynik estymacji. Wpływ na wynik estymacji obserwacji i może być mierzony zmianą wielkości estymatora \mathbf{b} po usunięciu tej obserwacji z próby albo też zmianą wielkości wartości dopasowanych $\hat{\mathbf{y}}$ po usunięciu tej obserwacji. Wyprowadzimy teraz wzór na obie te statystyki i

przekonamy się, że w obu przypadkach ważny wpływ na ich wilkość wywiera właśnie h_i . Oznaczmy jako $\mathbf{y}_{(i)}$, $\mathbf{X}_{(i)}$, $\mathbf{u}_{(i)}$ macierze danych i reszt, z których usunięto obserwację i . Z kolei $\mathbf{b}_{(i)}$ będzie oznaczał estymator *MNK* parametru β uzyskany na podstawie próby, z której usunięto obserwację i . Mamy więc następujące równanie:

$$\mathbf{y}_{(i)} = \mathbf{X}_{(i)} \mathbf{b}_{(i)} + \mathbf{u}_{(i)} \quad (4.2)$$

Rozważmy teraz następujące równanie regresji, w którym zawarto i -tą obserwację i dodatkowo zmienną zerojedynkową przyjmującą wartość 1 jedynie dla obserwacji i :

$$\begin{bmatrix} \mathbf{y}_{(i)} \\ y_i \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{(i)} & \mathbf{0} \\ \mathbf{x}_i & 1 \end{bmatrix} \begin{bmatrix} \mathbf{b}_{(i)} \\ \alpha \end{bmatrix} + \begin{bmatrix} \mathbf{e}_{(i)}^* \\ e_i^* \end{bmatrix} \quad (4.3)$$

Z własności hiperpłaszczyzny regresji wiemy, że kolumny macierzy danych muszą być ortogonalne do reszt:

$$\begin{aligned} \mathbf{X}'_{(i)} \mathbf{e}_{(i)}^* + \mathbf{x}_i e_i^* &= \mathbf{X}'_{(i)} \mathbf{e}_{(i)}^* = \mathbf{0} \\ e_i^* &= 0 \end{aligned}$$

Wynika z tego, że reszta i -ta reszta e_i^* w tej regresji jest równa zeru. Co więcej ponieważ układ równań normalnych $\mathbf{X}'_{(i)} \mathbf{e}_{(i)}^* = \mathbf{0}$ jest identyczny do tego dla regresji (4.2), więc estymator β uzyskany z regresji (4.3) jest taki sam jak estymator uzyskany z regresji (4.2): $\mathbf{b}_{(i)}^* = \mathbf{b}_{(i)}$. Co więcej jeśli zdefiniujemy wektor $\mathbf{e}_{(i)}^* = \left(\mathbf{e}_{(i)}^*, e_i^* \right)'$, to $\mathbf{X}' \mathbf{e}^* = \mathbf{0}$ i $\delta'_i \mathbf{e}^* = 0$. Uzyskany wynik świadczy o tym, że wprowadzając zmienną zerojedynkową wyeliminowaliśmy wpływ obserwacji i na wynik regresji. Współczynnik α stojący przy tej zmiennej można zinterpretować zauważając, że ponieważ

$$e_i^* = 0 = y_i - \mathbf{x}_i \mathbf{b}_{(i)} - \alpha$$

a więc α , jest równa błędowi dopasowania dla i -tej obserwacji jeśli nie została ona użyta w regresji:

$$u_i = y_i - \mathbf{x}_i \mathbf{b}_{(i)} = \alpha$$

Przeanalizujemy teraz różnicę między wynikami uzyskanymi z regresji na całej próbie:

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

i regresji, w której zawarto zmienną zero jedynkową dla obserwacji i .

$$\mathbf{e}^* - \mathbf{e} = (\mathbf{y} - \mathbf{X}\mathbf{b}_{(i)} - \delta_i\alpha) - (\mathbf{y} - \mathbf{X}\mathbf{b}) = \mathbf{X}(\mathbf{b} - \mathbf{b}_{(i)}) - \delta_i\alpha \quad (4.4)$$

Mnożąc obie strony tego równania przez $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ i wykorzystując własności hiperpłaszczyzny regresji otrzymujemy:

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{e}^* - \mathbf{e}) = 0 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{b} - \mathbf{b}_{(i)}) - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\delta_i\alpha$$

W rezultacie różnica między estymatorem *MNK* parametru β dla pełnej próby i próby z usuniętą obserwacją i -wynosi:

$$\mathbf{b} - \mathbf{b}_{(i)} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\delta_i\alpha \quad (4.5)$$

Mnożąc obie strony równania (4.4) i wykorzystując, to że $\mathbf{M}_X\mathbf{X} = 0$ otrzymujemy:

$$\mathbf{M}_X(\mathbf{e}^* - \mathbf{e}) = \mathbf{M}_X[\mathbf{X}(\mathbf{b} - \mathbf{b}_{(i)}) - \delta_i\alpha] = -\mathbf{M}_X\delta_i\alpha$$

Po przemnożeniu obu stron przez δ_i i wykorzystaniu faktu, że $\delta_i'e^* = 0$ oraz definicji h_i mamy:

$$e_i = -\delta_i'\mathbf{M}_X\delta_i\alpha = -\delta_i'(\mathbf{I} - \mathbf{P}_X)\delta_i\alpha = (1 - h_i)\alpha,$$

co pozwala policzyć nam parametr α :

$$\alpha = \frac{e_i}{1 - h_i}$$

Wstawiając uzyskany wynik do równania (4.5) otrzymujemy wreszcie końcowy wynik:

$$\mathbf{b} - \mathbf{b}_{(i)} = \frac{e_i}{1 - h_i}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \quad (4.6)$$

Im większa dźwignia tym większy wpływ \mathbf{x}_i na wielkość estymatora parametru β . Ważna jest także wielkość reszty dla obserwacji i . Szczególną uwagę powinniśmy zwracać na słabo dopasowane obserwacje o dużej dźwigni. Wzór (4.6) będzie wskazywał na istotny wpływ obserwacji i na wyniki estymacji jeśli różnica między elementami \mathbf{b} i $\mathbf{b}_{(i)}$ znacznie przekroczy wielkość odchylenia standardowego dla tych elementów. Łatwiej będzie nam wykryć takie obserwacje standaryzując elementy wektora różnic przez odchylenia standardowe elementów \mathbf{b} . W ten sposób uzyskamy statystykę *DFBETAS*:

$$DFBETAS_{k(i)} = \frac{\mathbf{b}_k - \mathbf{b}_{k(i)}}{\sqrt{\text{Var}(\mathbf{b}_k)}} = \frac{e_i}{1 - h_i} \frac{\delta'_k (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'_i}{\sigma \sqrt{(\mathbf{X}'\mathbf{X})_{ii}^{-1}}} = \frac{\hat{e}_i}{\sqrt{1 - h_i}} \frac{\delta'_k (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'_i}{\sqrt{(\mathbf{X}'\mathbf{X})_{ii}^{-1}}}$$

W literaturze przedmiotu podaje się, że warte zbadania są obserwacje, dla których $|DFBETAS_{k(i)}| > \frac{2}{\sqrt{N}}$.

Różnica między wartością dopasowaną \hat{y}_i policzoną na podstawie pełnej próby i wartością dopasowaną $\hat{y}_{(i)}$ uzyskaną z użyciem estymatora policzonego dla próby z usuniętą obserwacją i wynosi:

$$\hat{y}_i - \hat{y}_{(i)} = \mathbf{x}_i \mathbf{b} - \mathbf{x}_i \mathbf{b}_{(i)} = \frac{e_i}{1 - h_i} \mathbf{x}_i (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_i = e_i \frac{h_i}{1 - h_i}$$

Aby ocenić wielkość tego odchylenia lepiej posłużyć się jego wielkością wystandaryzowaną przez odchylenie standardowe \hat{y}_i równe $\sqrt{\text{Var}(\hat{y}_i)}$. Ponieważ

$$\text{Var}(\hat{\mathbf{y}}) = \text{Var}(\mathbf{X}\mathbf{b}) = \sigma^2 \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' = \sigma^2 \mathbf{P}_X,$$

więc $\text{Var}(\hat{y}_i) = \text{Var}(\delta'_i \hat{\mathbf{y}}) = \delta'_i \mathbf{P}_X \delta_i = \sigma^2 h_i$. Wystandaryzowana miara oznaczana jest jako *DFITS* i ma postać

$$DFITS_{(i)} = \frac{\hat{y}_i - \hat{y}_{(i)}}{\sqrt{\text{Var}(\hat{y}_i)}} = \frac{e_i \frac{h_i}{1 - h_i}}{\sqrt{\sigma^2 h_i}} = \hat{e}_i \sqrt{\frac{h_i}{1 - h_i}}$$

W literaturze można znaleźć sugestie, że warte zbadania są obserwacje, dla których $DFITS > 2\sqrt{\frac{k}{N}}$

Dla całego wektora wartości dopasowanych analogiczna różnica jest równa:

$$\hat{\mathbf{y}} - \hat{\mathbf{y}}_{(i)} = \mathbf{X}\mathbf{b} - \mathbf{X}\mathbf{b}_{(i)} = \frac{e_i}{1 - h_i} \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_i$$

Możemy teraz stworzyć statystykę, która odzwierciedlać będzie całkowity wpływ obserwacji i na wartości dopasowane:

$$\begin{aligned} (\hat{\mathbf{y}} - \hat{\mathbf{y}}_{(i)})' (\hat{\mathbf{y}} - \hat{\mathbf{y}}_{(i)}) &= \frac{e_i^2}{(1 - h_i)^2} \mathbf{x}_i' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_i \\ &= \frac{e_i^2 h_i}{(1 - h_i)^2} \end{aligned}$$

Jeśli znormalizujemy tak uzyskany wynik przez $p\sigma^2$ to uzyskamy tak zwaną *odległość Cooka* dla obserwacji i

$$CD_i = \frac{(\hat{\mathbf{y}} - \hat{\mathbf{y}}_{(i)})' (\hat{\mathbf{y}} - \hat{\mathbf{y}}_{(i)})}{p\sigma^2} = \frac{e_i^2 h_i}{p\sigma^2 (1 - h_i)^2} = \frac{\hat{e}_i^2}{p} \frac{h_i}{1 - h_i}$$

Literatura przedmiotu sugeruje, że naogół należy zbadać te obserwacje, dla których $CD_i > 0.5$ a zawsze należy zwrócić uwagę na te, dla których $CD_i > 1$.

4.2. Obserwacje nietypowe - analiza graficzna

Analiza graficzna wyników regresji należy do najważniejszych narzędzi diagnostycznych. Za jej pomocą najłatwiej wykryć nietypowe obserwacje oraz inne problemy związane z danymi.

4.3. Dokładna współliniowość

Dokładna współliniowość pojawia się, gdy kolumny macierzy \mathbf{X} są liniowo zależne. W takim przypadku $\text{Rank}(\mathbf{X}) < K + 1$ a macierz $\mathbf{X}'\mathbf{X}$ jest macierzą osobliwą. Jej odwrócenie jest niewykonalne a tym samym niemożliwe jest policzenie estymatora MNK postaci $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. Problem z policzeniem estymatora MNK bierzą z niemożnością *identyfikacji* elementów wektora \mathbf{b} . Powiedzmy, że mamy model

$$y_t = \beta_0 + \beta_1 x_{1i} + \dots + \beta_K x_{Ki} + \varepsilon_i$$

i $1, x_1, \dots, x_K$ są liniowo zależne, co oznacza, że istnieje $(\alpha_0, \alpha_1, \dots, \alpha_K) \neq 0$ takie, że $\alpha_0 + \alpha_1 x_{1i} + \dots + \alpha_K x_{Ki} = 0$ dla każdego $i = 1 \dots N$. Analizowany model można więc przekształcić do równoważnego model:

$$\begin{aligned} y_t &= (\beta_0 + \beta_1 x_{1i} + \dots + \beta_K x_{Ki} + \varepsilon_i) + (\alpha_0 + \alpha_1 x_{1i} + \dots + \alpha_2 x_{Ki}) \\ &= (\beta_0 + \alpha_0) + (\beta_1 + \alpha_1) x_{1i} + \dots + (\beta_K + \alpha_K) x_{Ki} + \varepsilon_i \\ &= \gamma_0 + \gamma_1 x_{1i} + \dots + \gamma_K x_{Ki} + \varepsilon_i \end{aligned}$$

i $\beta_0 \neq \gamma_0, \beta_1 \neq \gamma_1, \dots, \beta_K \neq \gamma_K$. Jeśli jednak dwa modele o różnych parametrach są równoważne to nie ma najmniejszej szansy na zidentyfikowanie, który z nich jest prawdziwym modelem! Z problemem identyfikacji w ogólniejszym kontekście zetkniemy się przy okazji analizowania modeli wielorówaniowych.

Przykład 4.1 Powiedzmy, że mamy model na logarytmach, który objaśnia wielkość dochodu narodowego Y_t wielkością wydatków na edukację E_t , ilością ludności P_t oraz wydatkami na edukację per capita Z_t . W trakcie estymacji pojawił się problem z dokładną współliniowością. Dlaczego?

Odpowiedź: W modelu tym wystąpi współliniowość, ponieważ $\ln(Z_t) = \ln\left(\frac{E_t}{P_t}\right) = \ln(E_t) - \ln(P_t)$. Wynika z tego, że zmienne objaśniające w modelu $\ln(Z_t), \ln(E_t), \ln(P_t)$ są liniowo zależne.

Przykład 4.2 Załóżmy, że estymujemy pewien model na 20-to letnim szeregu czasowym. W modelu znajduje się 25 zmiennych objaśniających. Model ten także spowodował załamanie się programu do estymacji. Dlaczego?

Odpowiedź: Macierz \mathbf{X} ma w tym przypadku 25 kolumn i 20 wierszy. Maksymalna liczba liniowo niezależnych kolumn wynosi w tym przypadku 20. Oznacza to, że kolumny macierzy \mathbf{X} a tym samym i zmienne w modelu są liniowo zależne (współliniowe).

Dokładna współliniowość jest więc spowodowana błędnym zdefiniowaniem modelu lub zbyt dużą, w stosunku do wielkości próby, liczbą zmiennych objaśniających.

4.4. Niedokładna współliniowość

O niedokładnej współliniowości mówimy, gdy korelacje między zmiennymi egzogenicznymi są niezerowe. Jest ona raczej regułą niż wyjątkiem i w przeciwieństwie do dokładnej współliniowości jest cechą danych a nie modelu. Powoduje ona podobne problemy co dokładna współliniowość, utrudniając zidentyfikowanie wielkości parametrów przy zmiennych egzogenicznych.

Podzielmy zmienne na egzogeniczne na dwie grupy: $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2]$. Możemy teraz sformułować dwie regresje

$$\mathbf{y} = \mathbf{X}_1 \mathbf{b} + \mathbf{e}_1$$

$$\mathbf{y} = \mathbf{X}_1 \mathbf{b}_1 + \mathbf{X}_2 \mathbf{b}_2 + \mathbf{e}_2$$

Niech $\mathbf{M}_1 = \mathbf{I} - \mathbf{X}_1 (\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1'$, przy czym \mathbf{M}_1 jest oczywiście idempotentna. Mnożąc drugą regresję przez \mathbf{M}_1 uzyskujemy

$$\mathbf{M}_1 \mathbf{y} = \mathbf{M}_1 \mathbf{X}_2 \mathbf{b}_2 + \mathbf{e}_2,$$

ponieważ $M_1 X_1 = 0$. Ponieważ $X'_1 e_2 = 0$ więc $X'_1 e_2 = 0$, $X'_2 e_2 = 0$ tak, że $M_1 e_2 = e_2$. Mnożąc uzyskane równanie przez X_2 otrzymujemy

$$X'_2 M_1 y = X'_2 M_1 X_2 b_2$$

co prowadzi do wzoru na b_2

$$b_2 = (X'_2 M_1 X_2)^{-1} X'_2 M_1 y,$$

przy czym zauważmy, że $\hat{u} = M_1 X_2$ jest równe resztom z regresji X_2 na X_1 . Zamiast liczyć estymator b_2 na podstawie pełnej regresji, to jest na podstawie wzoru $b = (X'X)^{-1} X'y$, gdzie $b = (b'_1, b'_2)'$, można go policzyć w dwóch krokach. W pierwszym kroku przeprowadzamy regresję X_2 na X_1 , a w drugim regresję y na resztach \hat{u} z regresji X_2 na X_1 , otrzymując $b_2 = (\hat{u}'\hat{u})^{-1} \hat{u}'y$.

Rozważmy przypadek, gdy X_2 składa się tylko z jednej zmiennej x_k (jednej kolumny). Przypuśćmy dalej, że w danych występuje współliniowość, co oznacza, że x_k jest blisko skorelowany z X_1 . W tym przypadku reszty \hat{u}_k z regresji x_k na X_1 są małe, a co za tym idzie mała będzie suma ich kwadratów RSS_k równa $RSS_k = \hat{u}'_k \hat{u}_k = x'_k M'_1 M_1 x_k = x'_k M_1 x_k$. Wariancja $b_k = (\hat{u}'_k \hat{u}_k)^{-1} \hat{u}'_k y$ będzie równa

$$\text{Var}(b_k) = \sigma^2 (x'_k M'_1 M_1 x_k)^{-1} = \sigma^2 (\hat{u}'_k \hat{u}_k)^{-1} = \frac{\sigma^2}{TSS_k (1 - R_k^2)},$$

gdzie $TSS_k = \sum (x_i - \bar{x})^2$ jest równe całkowitej sumie kwadratów w regresji x_k na X_1 a R_k^2 jest współczynnikiem determinacji w regresji x_k na X_1 . W praktyce σ^2 jest nieznanym parametrem szacowanym za pomocą estymatora $s^2 = \frac{e'e}{N-K} = \frac{TSS(1-R^2)}{N-K}$. W rezultacie wyestymowana wariancja estymatora b_k wynosi:

$$S_{kk} = \frac{1}{N - K} \frac{TSS}{TSS_k} \frac{(1 - R^2)}{(1 - R_k^2)}, \quad (4.7)$$

gdzie R_k^2 jest współczynnikiem determinacji w regresji x_k na \mathbf{X}_1 .

Statystyka t dla hipotezy, że $\beta_k = 0$ równa, jak się dowiemy w następnym rozdziale, $t = \frac{b_k}{\sqrt{S_{kk}}}$ będzie niska, ponieważ jej mianownik będzie duży. Zmienne współliniowe będą miały wysokie błędy standardowe oraz niskie statystyki t niezależnie od tego, czy są to zmienne istotne, czy też nieistotne w modelu.

Wynik ten można wytłumaczyć w bardziej intuicyjny sposób na przykładzie modelu z dwoma zmiennymi:

$$y_i = \beta_1 + \beta_2 x_{1i} + \beta_3 x_{2i} + \varepsilon_i$$

w którym x_1 i x_2 są współliniowe tak, że $x_{2i} = a_1 + a_2 x_{1i} + \eta_i$ i η_i jest pewną zmienną losową o niskiej wariancji. Jeśli podstawimy wzór na x_{2i} do tego modelu, to uzyskamy model

$$y_i = \beta_1 + \beta_2 x_{1i} + \beta_3 (a_1 + a_2 x_{1i} + \eta_i) + \varepsilon_i = \beta_1^* + \beta_2^* x_{1i} + \varepsilon_i + \beta_3 \eta_i,$$

dla $\beta_1^* = \beta_1 + a_1 \beta_3$ i $\beta_2^* = \beta_2 + a_2 \beta_3$. Jeśli wariancja $\beta_3 \eta_i$ jest niska to jakość dopasowania modelu¹ z jedną zmienną objaśniającą będzie niewiele mniejsza niż jakość dopasowania modelu z dwoma zmiennymi. Wniosek ten nie zależy od tego, jaką postać ma prawdziwy model a jedynie od cech danych.

Ponieważ statystyki testowe opierają się na porównywaniu jakości dopasowania modelu z większą i mniejszą ilością zmiennych, więc trudno będzie odrzucić hipotezę, że x_{2i} jest nieistotne.

¹Mierzona na przykład statystyką R^2

4.5. Wykrywanie współliniowości

Do wykrywania współliniowości możemy posłużyć się kilkoma statystykami. Pierwszą z nich jest R_k^2 , współczynnik determinacji dla regresji zmiennej x_k na pozostałych zmiennych w modelu. Im wyższy ten współczynnik tym większa współliniowość i, jak można zauważyć na podstawie wzoru (4.7), wyższa wariancja estymatora b_k . Wariancja b_k będzie wprost proporcjonalna do współczynnika VIF (Variance Inflation Ratio)

$$VIF_k = \frac{1}{1 - R_k^2}$$

Stosuje się też niekiedy regułę *ad hoc*, która mówi, że w modelu występuje współliniowość jeśli *którykolwiek* z R_k^2 jest wyższy niż R^2 w estymowanym modelu.

Inną miarą współliniowości może być warunek rzędu. W tym przypadku porównujemy największą i najmniejszą wartość własną macierzy $\mathbf{X}'\mathbf{X}$. Wartości własne tej macierzy są większe lub równe zero. W przypadku, kiedy macierz jest bliska współliniowości część jej wartości własnych jest bliska zero. Z kolei, gdy kolumny \mathbf{X} są ortogonalne, macierz $\mathbf{X}'\mathbf{X} = \tilde{\Lambda}$ i $\tilde{\lambda}_i = \mathbf{x}'_i\mathbf{x}_i$. Widać z tego, że wielkość wartości własnych zależy od skali \mathbf{x}_i . Normalizujemy więc \mathbf{X}_i dzieląc jej kolumny przez $(\mathbf{x}'_i\mathbf{x}_i)^{\frac{1}{2}}$ i tworzymy miarę współliniowości będącą pierwiastkiem stosunku największej do najmniejszej wartości własnej tak powstałej macierzy znormalizowanej

$$\gamma = \left(\frac{\lambda_{\max}}{\lambda_{\min}} \right)^{\frac{1}{2}}$$

Z poprzednich rozważań wynika, że dla \mathbf{X} ortogonalnego $\gamma = 1$ zaś dla macierzy \mathbf{X} bliskiej współliniowości γ jest duże. W literaturze można znaleźć sugestię, że $\gamma > 20$ wskazuje na współliniowość. Statystkę γ nazywamy liczbą warunkową (*conditional value*).

4.6. Usuwanie współliniowości

• Estymator regresji grzbietowej

Jednym z proponowanych sposobów radzenia sobie ze współliniowością jest zastosowanie estymatora regresji grzbietowej (*ridge regression estimator*). Estymator ten ma postać

$$\mathbf{b}_r = (\mathbf{X}'\mathbf{X} + r\mathbf{D})^{-1} \mathbf{X}'\mathbf{y}$$

gdzie \mathbf{D} jest macierzą diagonalną z elementami diagonalnymi macierzy $\mathbf{X}'\mathbf{X}$ na przekątnej, zaś r jest pewną arbitralnie dobraną stałą większą od zera. Estymator taki jest obciążony, ponieważ

$$E(\mathbf{b}_r) = E\left[(\mathbf{X}'\mathbf{X} + r\mathbf{D})^{-1} \mathbf{X}'\mathbf{y}\right] = (\mathbf{X}'\mathbf{X} + r\mathbf{D})^{-1} \mathbf{X}' E(\mathbf{y}) = (\mathbf{X}'\mathbf{X} + r\mathbf{D})^{-1} \mathbf{X}'\mathbf{X}\beta \neq \beta$$

jednak jego wariancja

$$\text{Var}(\mathbf{b}_r) = \sigma^2 (\mathbf{X}'\mathbf{X} + r\mathbf{D})^{-1} \mathbf{X}'\mathbf{X} (\mathbf{X}'\mathbf{X} + r\mathbf{D})^{-1} \quad (4.8)$$

jest mniejsza od wariancji estymatora *MNK*. Estymator regresji grzbietowej nie jest zbyt popularny w ekonometrii ze względu na jego obciążenie i arbitralność jego formy. Pokazano jednak, że można estymator ten można wyprowadzić jako estymator *bayesowski* parametru β przy pewnym rozkładzie *a priori* parametru $\beta \sim N\left(0, \frac{\sigma^2}{r} \mathbf{D}\right)$.

• Metoda czynników głównych

Innym sposobem radzenia sobie ze współliniowością jest zastosowanie metody czynników głównych (*principal components*). Metoda ta może służyć do wyeliminowania zarówno dokładnej współliniowości spowodowanej zbyt dużą ilością zmiennych w stosunku do ilości obserwacji jak i niedokładnej współliniowości. Celem badania ekonometrycznego jest znalezienie związków między zmienną zależną a zmiennymi niezależnymi. Informację na temat tych związków uzyskujemy analizując zależności między zaobserwowaną zmiennością zmiennej zależnej a zaobserwowaną zmiennością

zmiennych niezależnych. Jeśli źródeł zmienności zmiennych niezależnych jest mniej niż zmiennych, to zmienne te będą mniej lub bardziej współliniowe. W takim przypadku logiczny wydaje się pomysł, by zamiast oryginalnych zmiennych użyć funkcji tych zmiennych, które przybliżałyby pierwotne źródła zmienności. Celem metody czynników głównych jest znalezienie takich kombinacji liniowych oryginalnych zmiennych ezogenicznych, które najlepiej przybliżają zmienność tych zmiennych. Na następnym etapie wybieramy spośród tych kombinacji pewną (małą) ich liczbę, która wyjaśnia większość zmienności zmiennych niezależnych i zastępujemy tymi kombinacjami oryginalne zmienne.

Kombinację liniową z_i zmiennych niezależnych o współczynnikach danych wektorem c_i można zapisać jako

$$z_i = x_1 c_{i1} + \dots + x_K c_{iK}, \quad \text{dla } i = 1, \dots, L \quad (4.9)$$

$$\mathbf{Z} = \mathbf{X}\mathbf{C}$$

Założmy, że zdefiniowanych czynników głównych jest L . Wiemy z własności R^2 , że jakość dopasowania \mathbf{X} do \mathbf{Z} nie zależy nie zmienia się \mathbf{X} przekształcimy liniowo. Znormalizujemy więc \mathbf{C} , tak by

$$\mathbf{Z}'\mathbf{Z} = \mathbf{I} \quad (4.10)$$

Naszym celem jest znalezienie takiego wektora \mathbf{Z} , który umożliwiałby możliwie najlepsze przybliżenie kolumn macierzy \mathbf{X} , to jest znalezienie takiego \mathbf{C} , który minimalizowałby sumę kwadratów reszt regresji \mathbf{X} na \mathbf{Z} .

$$\mathbf{X} = \mathbf{Z}\mathbf{B} + \mathbf{E}$$

$$S(\mathbf{B}, \mathbf{C}) = \text{tr} [(\mathbf{X} - \mathbf{Z}\mathbf{B})'(\mathbf{X} - \mathbf{Z}\mathbf{B})] = \text{tr} [(\mathbf{X} - \mathbf{X}\mathbf{C}\mathbf{B})'(\mathbf{X} - \mathbf{X}\mathbf{C}\mathbf{B})] = \text{tr} \mathbf{E}'\mathbf{E}, \quad (4.11)$$

i minimalizacja będzie realizowana przy warunku pobocznym (4.10)

$$\min_{\mathbf{B}, \mathbf{C}} S(\mathbf{B}, \mathbf{C}) \quad \text{s.t. } \mathbf{Z}'\mathbf{Z} = \mathbf{I}.$$

Sformułujmy funkcję Lagrange'a dla tego problemu:

$$\ell(\mathbf{B}, \mathbf{C}) = \text{tr } \mathbf{E}'\mathbf{E} + \text{tr} [(\mathbf{I} - \mathbf{Z}'\mathbf{Z}) \mathbf{\Lambda}]$$

Naszym zadaniem jest teraz znalezienie $\min_{\mathbf{B}, \mathbf{C}} \ell(\mathbf{B}, \mathbf{C})$. Ponieważ warunek poboczny nie jest związany \mathbf{B} , więc warunki pierwszego rzędu dla \mathbf{B} mają postać klasycznego układu równań normalnych

$$\begin{aligned} \frac{\partial \ell(\mathbf{B}, \mathbf{C})}{\partial \text{vec}(\mathbf{B})} &= \frac{\partial \text{tr} [(\mathbf{X} - \mathbf{Z}\mathbf{B})'(\mathbf{X} - \mathbf{Z}\mathbf{B})]}{\partial \text{vec}(\mathbf{B})} \\ &= \frac{\partial \text{tr}(\mathbf{X}'\mathbf{X})}{\partial \text{vec}(\mathbf{B})} - 2 \frac{\partial \text{tr}(\mathbf{B}'\mathbf{Z}'\mathbf{X})}{\partial \text{vec}(\mathbf{B})} + \frac{\partial \text{tr}(\mathbf{B}\mathbf{Z}'\mathbf{Z}\mathbf{B})}{\partial \text{vec}(\mathbf{B})} \\ &= -2 \frac{\partial \text{vec}(\mathbf{B})'}{\partial \text{vec}(\mathbf{B})} \text{vec}(\mathbf{Z}'\mathbf{X}) + \frac{\partial [\text{vec}(\mathbf{B})'(\mathbf{I} \otimes \mathbf{Z}'\mathbf{Z}) \text{vec}(\mathbf{B})]}{\partial \text{vec}(\mathbf{B})} \\ &= -2 \text{vec}(\mathbf{Z}'\mathbf{X}) + 2(\mathbf{I} \otimes \mathbf{Z}'\mathbf{Z}) \text{vec}(\mathbf{B}) = 2 \text{vec}(-\mathbf{Z}'\mathbf{X} + \mathbf{Z}'\mathbf{Z}\mathbf{B}) = 0 \end{aligned}$$

Ostatnia równość jedynie, gdy $\mathbf{Z}'\mathbf{X} = \mathbf{Z}'\mathbf{Z}\mathbf{B}$, więc

$$\mathbf{B} = (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{X} = \mathbf{Z}'\mathbf{X}$$

a wykorzystując normalizację (4.10) otrzymujemy, że $\mathbf{B} = \mathbf{Z}'\mathbf{X}$. Możemy teraz zapisać reszty jako

$$\mathbf{E} = \mathbf{X} - \mathbf{Z}\mathbf{B} = \mathbf{M}_{\mathbf{Z}}\mathbf{X},$$

gdzie macierz $\mathbf{M}_{\mathbf{Z}} = \mathbf{I} - \mathbf{Z}\mathbf{Z}'$ jest macierzą idempotentną.

$$\begin{aligned} S(\mathbf{C}) &= \text{tr } \mathbf{E}'\mathbf{E} = \text{tr}(\mathbf{X}'\mathbf{M}_{\mathbf{Z}}\mathbf{X}) = \text{tr}(\mathbf{X}'\mathbf{X}) - \text{tr}(\mathbf{X}'\mathbf{Z}\mathbf{Z}'\mathbf{X}) \\ &= \text{tr}(\mathbf{X}'\mathbf{X}) - \text{tr}(\mathbf{Z}'\mathbf{X}\mathbf{X}'\mathbf{Z}) = \text{tr}(\mathbf{X}'\mathbf{X}) - \text{tr}(\mathbf{C}'\mathbf{X}'\mathbf{X}\mathbf{X}'\mathbf{X}\mathbf{C}) \end{aligned}$$

Po podstawieniu tego wyniku, funkcja Lagrange przyjmie postać

$$\begin{aligned}\ell(\mathbf{C}) &= \text{tr}(\mathbf{X}'\mathbf{X}) - \text{tr}[\mathbf{C}'\mathbf{X}'\mathbf{X}\mathbf{X}'\mathbf{X}\mathbf{C}] + \text{tr}(\mathbf{C}'\mathbf{X}'\mathbf{X}\mathbf{C}\mathbf{\Lambda} - \mathbf{\Lambda}) \\ &= \text{tr}(\mathbf{X}'\mathbf{X}) - \text{vec}(\mathbf{C})'(\mathbf{I} \otimes \mathbf{X}'\mathbf{X}\mathbf{X}'\mathbf{X})\text{vec}(\mathbf{C}) \\ &\quad + \text{vec}(\mathbf{C})'(\mathbf{\Lambda}' \otimes \mathbf{X}'\mathbf{X})\text{vec}(\mathbf{C}) - \text{tr} \mathbf{\Lambda}\end{aligned}$$

Warunki pierwszego rzędu mają teraz postać

$$\begin{aligned}\frac{\partial \ell(\mathbf{C})}{\partial \text{vec}(\mathbf{C})} &= -2(\mathbf{I} \otimes \mathbf{X}'\mathbf{X}\mathbf{X}'\mathbf{X})\text{vec}(\mathbf{C}) + 2(\mathbf{\Lambda}' \otimes \mathbf{X}'\mathbf{X})\text{vec}(\mathbf{C}) \\ &= -2\text{vec}(\mathbf{X}'\mathbf{X}\mathbf{X}'\mathbf{X}\mathbf{C}) + 2\text{vec}(\mathbf{X}'\mathbf{X}\mathbf{C}\mathbf{\Lambda}) = 0\end{aligned}$$

a więc

$$\mathbf{X}'\mathbf{X}\mathbf{X}'\mathbf{X}\mathbf{C} = \mathbf{X}'\mathbf{X}\mathbf{C}\mathbf{\Lambda}$$

mnożąc obie strony przez \mathbf{X} i przekształcając otrzymujemy

$$\mathbf{X}'\mathbf{X}\mathbf{C} = \mathbf{C}\mathbf{\Lambda}$$

mnożąc obie strony z prawej strony przez \mathbf{X} i wykorzystując, że z definicji $\mathbf{Z} = \mathbf{X}\mathbf{C}$ otrzymujemy

$$\mathbf{X}\mathbf{X}'\mathbf{Z} = \mathbf{Z}\mathbf{\Lambda}$$

co oznacza, że za macierz \mathbf{Z} możemy przyjąć macierz wektorów własnych macierzy $\mathbf{X}\mathbf{X}'$, przy czym macierz $\mathbf{\Lambda}$ będzie w tym przypadku macierzą diagonalną. Ponieważ $\text{Rank}(\mathbf{X}\mathbf{X}') = K$, więc maksymalny wymiar \mathbf{Z} , dla którego wszystkie diagonalne elementy $\mathbf{\Lambda}$ (wartości własne) są większe od zera wynosi K . Wartość funkcji kryterium jest równa

$$S(\mathbf{C}) = \text{tr}(\mathbf{X}'\mathbf{X}) - \text{tr}[\mathbf{Z}'\mathbf{X}\mathbf{X}'\mathbf{Z}] = \text{tr}(\mathbf{X}'\mathbf{X}) - \text{tr} \mathbf{\Lambda}.$$

Celem minimalizacji funkcji kryterium wybieramy więc wektory własne związane z największymi wartościami własnymi. Jeśli za miarę zmienności \mathbf{X} weźmiemy, $\text{tr}(\mathbf{X}'\mathbf{X})$ to miarą proporcji

całkowitej zmienności \mathbf{X} do zmienności wyjaśnionej przez \mathbf{c}_i będzie

$$w_i = \frac{\lambda_i}{\text{tr}(\mathbf{X}'\mathbf{X})} = \frac{\lambda_i}{\sum_{k=1}^K \lambda_k}$$

Dla kilku L czynników głównych miarą zmienności będzie $\sum_{k=1}^L w_k$. Ponieważ dla $K = L$ ślad $\text{tr}(\mathbf{\Lambda}) = \text{tr}(\mathbf{X}\mathbf{X}') = \text{tr}(\mathbf{X}'\mathbf{X})$ więc miara ta przyjmuje wartości między 0 i 1. Czynnikiami głównymi nazywamy wektory $\mathbf{z}_i = \mathbf{X}\mathbf{c}_i$. Do dobieramy $L < K$ czynników głównych, takich że $\sum_{k=1}^L w_k$ jest wystarczająco bliska 1. Oznaczmy $\mathbf{C}_L = (\mathbf{c}_1, \dots, \mathbf{c}_L)$ i $\mathbf{Z}_L = (\mathbf{z}_1, \dots, \mathbf{z}_L) = \mathbf{X}\mathbf{C}_L$. Estymator MNK wyliczony dla modelu, w którym oryginalne zmienne zastąpiono czynnikami głównymi będzie miał postać

$$\mathbf{d} = (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{y}.$$

Zauważmy, że

$$\mathbf{Z}'\mathbf{Z} = \mathbf{C}_L\mathbf{X}'\mathbf{X}\mathbf{C}_L = \mathbf{\Lambda}_L,$$

gdzie $\mathbf{\Lambda}_L$ jest macierzą diagonalną zawierającą L największych wartości własnych macierzy $\mathbf{X}'\mathbf{X}$. Z drugiej strony

$$\mathbf{Z}'\mathbf{y} = \mathbf{C}_L'\mathbf{X}'\mathbf{y} = \mathbf{C}_L'(\mathbf{X}'\mathbf{X})\mathbf{b}$$

a $\mathbf{C}_L'(\mathbf{X}'\mathbf{X}) = \mathbf{\Lambda}_L\mathbf{C}_L$, więc

$$\mathbf{d} = \mathbf{C}_L'\mathbf{b}$$

co oznacza, że uzyskany estymator jest prostą funkcją estymatora MNK wyliczonego dla oryginalnych zmiennych. Wektor \mathbf{d} ma L elementów. Aby bliżej zbadać relacje między estymatorem policzonym przy zastosowaniu metody czynników głównych a estymatorem MNK zauważmy, że $\hat{\mathbf{y}}$ dla metody czynników głównych jest równe

$$\hat{\mathbf{y}} = \mathbf{Z}\mathbf{d} = \mathbf{X}\mathbf{C}_L\mathbf{C}_L'\mathbf{b}$$

w związku z tym za odpowiednik estymatora MNK dla metody czynników głównych jest estymator

$$\mathbf{b}_p = \mathbf{C}_L \mathbf{d} = \mathbf{C}_L \mathbf{C}'_L \mathbf{b}$$

Estymator ten będzie równoważny estymatorowi MNK dla $L = K$. W ogólnym przypadku $L < K$ i \mathbf{b}_p będzie estymatorem obciążonym ale o wariancji mniejszej od wariancji estymatora \mathbf{b} .

Najważniejszym problemem przy zastosowaniach metody czynników głównych jest znalezienie interpretacji dla policzonych kombinacji liniowych oryginalnych zmiennych. Ponieważ, kombinacje te są znajdowane mechanicznie a nie na podstawie teoretycznych rozważań, więc z reguły trudno jest im nadać jakąś określoną interpretację. Zauważmy jednak, że wartość funkcji celu w przypadku, kiedy $1 < L < K$ jest równa

$$S(\alpha_1, \dots, \alpha_K) = S(\mathbf{C}_L) = \text{tr}(\mathbf{X}'\mathbf{X}) + \text{tr}\left[\mathbf{C}'_L (\mathbf{X}'\mathbf{X})^2 \mathbf{C}_L\right] = \text{tr}(\mathbf{X}'\mathbf{X}) + \sum_{k=1}^L \lambda_k^2$$

Wartość funkcji kryterium pozostanie niezmienniona jeśli zamiast \mathbf{C}_L użyjemy $\mathbf{C}_L^* = \mathbf{C}_L \mathbf{D}$, gdzie $\mathbf{D}_{L \times L}$ jest dowolną macierzą ortonormalną. Oznacza to, że można zastosować inną normalizację niż ta implikowana przez ograniczenia $\mathbf{z}'_s \mathbf{z}_s = 1$ i $\mathbf{z}'_s \mathbf{z}_r = 0$ dla $s \neq r$. Rzeczywiście

$$\begin{aligned} \text{tr}\left[\mathbf{C}_L^* (\mathbf{X}'\mathbf{X})^2 \mathbf{C}_L^*\right] &= \text{tr}\left[\mathbf{D}' \mathbf{C}'_L (\mathbf{X}'\mathbf{X})^2 \mathbf{C}_L \mathbf{D}\right] = \text{tr}\left(\mathbf{D} \mathbf{D}' \mathbf{C}'_L (\mathbf{X}'\mathbf{X})^2 \mathbf{C}_L\right) \\ &= \text{tr}\left(\mathbf{C}'_L (\mathbf{X}'\mathbf{X})^2 \mathbf{C}_L\right) \end{aligned}$$

Może się zdarzyć, że jesteśmy w stanie znaleźć taką macierz \mathbf{D} , że czynniki głównym danym wzorem $\mathbf{Z}_L = \mathbf{X} \mathbf{C}_L^*$, można nadać jakąś interpretację.

Literatura: Steward (1991) str. 88-95, Green (1997) str. 255-259, Goldberger (1972) str. 252-254, Theil (1979) str. 164-170.