MODELE REGRESJI SPROWADZALNE DO LINIOWYCH

Łukasz Kwiatkowski

Katedra Ekonometrii i Badań Operacyjnych

Plan wykładu

1) Wstęp
2) Model potęgowy
3) Model wykładniczy
4) Model potęgowo-wykładniczy („mieszany”)
5) Model potęgowo-wykładniczy – estymacja
6) Ocena dobroci dopasowania modelu z logarytmiczną zmienną objaśnianą

Wstęp

- W praktyce bardzo często modele regresji liniowej buduje się nie tyle bezpośrednio dla zmiennej ekonomicznej (w charakterze zmiennej objaśnianej), co dla jej **logarytmu**, przy czym poszczególne zmienne objaśniające mogą występować bądź w swojej oryginalnej (niezlogarytmowanej) postaci, bądź – podobnie jak zmienna objaśniana – zlogarytmowane. **Przykładem** jest tzw. **model funkcji produkcji Cobba-Douglasa** (znany z mikroekonomii :)

$$\ln Q_t = \beta_0 + \beta_1 \ln C_t + \beta_2 \ln L_t + \varepsilon_t, \qquad (1)$$

gdzie Q_t oznacza wielkość produkcji (np. w różnych przedsiębiorstwach z danej branży, $t = 1, \dots, T$), zaś C_t i L_t są dwoma głównymi czynnikami produkcji: C_t – wielkość nakładu kapitału, L_t – wielkość nakładu pracy. W istocie, powyższy model wywodzi się z teorii ekonomii, gdzie oryginalnie przyjmuje **postać potęgową**:

$$Q_t = \alpha_0 C_t^{\beta_1} L_t^{\beta_2} e^{\varepsilon_t}, \qquad (2)$$

przy czym $\alpha_0 = e^{\beta_0}$.

Zauważmy jednak, że ten ostatni NIE spełnia 1. założenia KM(N)RL – z dwóch powodów:

a) po pierwsze, składnik losowy NIE występuje w postaci addytywnej (czyli „$+\varepsilon_t$”), a w multiplikatywnej („razy e^{ε_t}”);

b) po drugie, część systematyczna (inaczej: deterministyczna) modelu, tj. fragment $\alpha_0 C_t^{\beta_1} L_t^{\beta_2}$, NIE jest LINIOWĄ funkcją parametrów

→ W związku z powyższym, modelu (2) NIE da się zapisać w postaci $y_t = x_t\beta + \varepsilon_t$, więc – powtórzmy – założenie 1 KM(N)RL NIE jest spełnione, wobec czego modelu (2) NIE da się bezpośrednio oszacować

Wstęp

➢ Zauważmy jednak, że model (1):

- otrzymujemy poprzez zlogarytmowanie „wyjściowej” specyfikacji modelu, tj. równania (2) (przyjmując dodatkowo $\beta_0 = \ln \alpha_0$)
- spełnia 1. założenie KM(N)RL:

  a) ma addytywny składnik losowy;

  b) część systematyczna modelu, tj. fragment $\beta_0 + \beta_1 \ln C_t + \beta_2 \ln L_t$, owszem JEST LINIOWĄ funkcją parametrów

→ Wobec powyższego równanie (1) da się zapisać w postaci odpowiadającej 1. założeniu KM(N)RL, tj. $y_t = x_t\beta + \varepsilon_t$:

→ W związku z tym – o ile modelu (2) NIE da się oszacować za pomocą MNK – o tyle model (1) już tak.

Zatem o modelu postaci (2) powiemy, że jest **sprowadzalny do postaci liniowej**. To nas prowadzi już do ogólnych rozważań, zamieszczonych poniżej, a dotyczących dwóch głównych postaci modeli: potęgowej i wykładniczej, które choć pierwotnie NIE spełniają 1. założenia KM(N)RL (więc MNK nie daje się do nich bezpośrednio zastosować), tak jednak w wyniku transformacji logarytmicznej – już tak (i estymacja za pomocą MNK jest już możliwa).

Model potęgowy

- Niech $A_t \in \mathbb{R}_+$ (tj. $A_t > 0$) oznacza pewną zmienną ekonomiczną, której zmienność chcemy wyjaśnić za pomocą zmiennych objaśniających $Z_{ti} \in \mathbb{R}_+$ (tj. $Z_{ti} > 0$).
- **Model potęgowy:**

$$A_t = \alpha_0 Z_{t1}^{\beta_1} Z_{t2}^{\beta_2} \dots Z_{tK}^{\beta_K} e^{\varepsilon_t}$$

→ Powyższe równanie nie spełnia założenia 1 KMNRL, gdyż: i) składnik losowy nie jest addytywny, ii) część deterministyczna modelu nie stanowi liniowej kombinacji parametrów i zmiennych objaśniających. Wobec tego logarytmujemy wyjściowe równanie i reparametryzujemy $\ln \alpha_0$, przyjmując $\beta_0 = \ln \alpha_0$ (wszak dowolna funkcja parametru też jest parametrem, czyli nieznaną stałą):

$$\underbrace{\ln A_t}_{y_t} = \beta_0 + \beta_1 \underbrace{\ln Z_{t1}}_{x_{t1}} + \beta_2 \underbrace{\ln Z_{t2}}_{x_{t2}} + \dots + \beta_K \underbrace{\ln Z_{tK}}_{x_{tK}} + \varepsilon_t$$

Tym samym model spełnia już założenie 1 KMNRL (daje się zapisać w postaci $y_t = x_t\beta + \varepsilon_t$).

→ Terminy „zmienna objaśniana" i „zmienna objaśniająca" odnoszą się do modelu *ekonometrycznego*, a nie wyjściowej (potęgowej) specyfikacji, zatem są tu nimi, odpowiednio, $y_t \equiv \ln A_t$ oraz $x_{ti} \equiv \ln Z_{ti}$.

→ Zauważmy, że $\beta_i = \frac{\partial \ln A_t}{\partial \ln Z_{ti}} \equiv El_{A_t|Z_{ti}}$, a zatem β_i jest **elastycznością** zmiennej A_t (już bez logarytmu) względem zmiennej Z_{ti} (także bez logarytmu). Wobec tego, ocenę $\hat{\beta}_i$ interpretujemy tak:

Jeżeli wartość zmiennej Z_{ti} (← bez logarytmu) *byłaby wyższa o 1%* (*przy ustalonych wartościach pozostałych zmiennych* $Z_{t1}, \dots, Z_{t,i-1}, Z_{t,i+1}, \dots, Z_{tK}$), *wówczas wartość zmiennej* A_t *byłaby wyższa* ($\hat{\beta}_i > 0$)/niższa ($\hat{\beta}_i < 0$) <u>*przeciętnie*</u> *o* $\lvert\hat{\beta}_i\rvert$%. (← zapisano w module, bo w tym miejscu znak już pomijamy; wartości (modułu) oceny parametru NIE przemnażamy już przez 100%).

Model wykładniczy

- Niech $A_t \in \mathbb{R}_+$ oznacza pewną zmienną ekonomiczną, której zmienność chcemy wyjaśnić za pomocą zmiennych objaśniających $Z_{ti} \in \mathbb{R}$ (Z_{ti} nie musi przyjmować wartości tylko w $\mathbb{R}_+$).
- **Model wykładniczy:**

$$A_t = \alpha_0 \alpha_1^{Z_{t1}} \alpha_2^{Z_{t2}} \dots \alpha_K^{Z_{tK}} e^{\varepsilon_t}$$

→ Powyższe równanie nie spełnia założenia 1 KMNRL – z tych samych dwóch powodów, co model potęgowy. Wobec tego logarytmujemy wyjściowe równanie i reparametryzujemy $\ln \alpha_i$ $(i = 0, 1, \dots, K)$, przyjmując $\beta_i = \ln \alpha_i$:

$$\underbrace{\ln A_t}_{y_t} = \beta_0 + \beta_1 \underbrace{Z_{t1}}_{x_{t1}} + \beta_2 \underbrace{Z_{t2}}_{x_{t2}} + \dots + \beta_K \underbrace{Z_{tK}}_{x_{tK}} + \varepsilon_t$$

Tym samym model spełnia już założenie 1 KMNRL (daje się zapisać w postaci $y_t = x_t \beta + \varepsilon_t$).

→ Terminy „zmienna objaśniana” i „zmienna objaśniająca” odnoszą się do modelu *ekonometrycznego*, a nie wyjściowej (potęgowej) specyfikacji, zatem są tu nimi, odpowiednio, $y_t \equiv \ln A_t$ oraz $x_{ti} \equiv Z_{ti}$.

Model wykładniczy

→ Jak interpretować ocenę parametru β_i, stojącego przy danej zmiennej Z_{ti} $(i = 1, 2, \ldots, K)$?

Zauważmy, że jeżeli wartość zmiennej Z_{ti} wzrosłaby o 1 swoją jednostkę (przy ustalonych wartościach pozostałych zmiennych objaśniających), to wartość zmiennej objaśnianej, $\ln A_t$, zmieni się przeciętnie o (właśnie) $\hat{\beta}_i$ (wzrośnie, gdy $\hat{\beta}_i > 0$, lub spadnie, gdy $\hat{\beta}_i < 0$). Jak zmiana $\ln A_t$ o $\hat{\beta}_i$ przekłada się na zmianę bezpośrednio zmiennej ekonomicznej A_t? Rozpiszmy to sobie:

$$\ln A_t + \hat{\beta}_i = \ln A_t + \ln e^{\hat{\beta}_i} = \ln\left(A_t e^{\hat{\beta}_i}\right)$$

Zatem zmiana $\ln A_t$ o $\hat{\beta}_i$ jest równoznaczna z $e^{\hat{\beta}_i}$-krotną zmianą „samego" A_t. Ponieważ teraz mówimy o relatywnej (multiplikatywnej) zmianie wartości zmiennej A_t, z reguły lepiej będzie ją wyrazić w procentach. Przykładowo, gdyby czynnik $e^{\hat{\beta}_i}$ był równy 2, wówczas $A_t \cdot 2$ oznacza dwukrotny wzrost A_t, czyli wzrost o 100%. Podobnie, gdyby $e^{\hat{\beta}_i} = 0{,}7$, wówczas $A_t \cdot 0{,}7$ oznacza spadek wartości A_t o 30%. W ten sposób możemy zapisać ogólnie, że zmiana $\ln A_t$ o $\hat{\beta}_i$, powodując $e^{\hat{\beta}_i}$-krotną zmianą „samego" A_t, powoduje wzrost/spadek wartości A_t przeciętnie o $\left(e^{\hat{\beta}_i} - 1\right)100\%$.

Zauważmy, że zawsze $e^{\hat{\beta}_i} > 0$. Z drugiej strony, gdy $\hat{\beta}_i > 0$, wówczas $\left(e^{\hat{\beta}_i} - 1\right)100\% > 0$ (ponieważ $e^{\hat{\beta}_i} > 1$), natomiast przy $\hat{\beta}_i < 0$, wówczas $\left(e^{\hat{\beta}_i} - 1\right)100\% < 0$ (ponieważ $e^{\hat{\beta}_i} \in (0, 1)$). Oznacza to, że podobnie jak dotychczas – i zgodnie z intuicją – znak oceny parametru informuje nas bezpośrednio o kierunku zmiany (wzroście/spadku) wartości zmiennej A_t.

Model wykładniczy

→ Podsumowując, ocenę parametru β_i w modelu wykładniczym interpretujemy według szablonu:

Jeżeli wartość zmiennej Z_{ti} (← bez logarytmu) *byłaby wyższa o 1* <u>*jednostkę*</u> (*przy ustalonych wartościach pozostałych zmiennych* $Z_{t1},\dots,Z_{t,i-1},Z_{t,i+1},\dots,Z_{tK}$), *wówczas wartość zmiennej* A_t *byłaby wyższa* ($\hat{\beta}_i > 0$)/niższa ($\hat{\beta}_i < 0$) <u>*przeciętnie*</u> *o* $\left|\left(e^{\widehat{\beta}_i}-1\right)100\right|\%$. (← zapisano w module, bo w tym miejscu znak już pomijamy).

Model wykładniczy

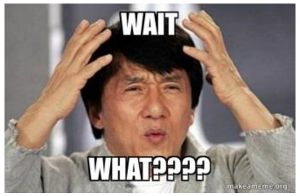
- Kiedy do modelowania pewnej zmiennej ekonomicznej możemy chcieć zastosować ***model wykładniczy*** (zmienna objaśniana – w postaci logarytmicznej; zmienne objaśniające – bez takiej transformacji, tj. w swojej wyjściowej postaci) – trzy główne powody:

1) W modelowaniu szeregów czasowych – gdy chcemy uwzględnić **trend deterministyczny** w modelu (przykładowo, w równaniu $\ln A_t = \beta_0 + \beta_1 t + \cdots + \varepsilon_t$ zmienna czasowa t jest BEZ logarytmu):
   - Gdy A_t przejawia wzrost wykładniczy, wówczas $\ln A_t$ charakteryzuje się trendem liniowym

   rysunek?

   - Ze względu na interpretację oceny parametru przy trendzie – w powyższym równaniu, gdzie zmienna t jest niezlogarytmowana, moglibyśmy sformułować interpretację w stylu *„Z okresu na okres (przy ustalonych wartościach pozostałych zmiennych objaśniających), wartość zmiennej A_t wzrasta/spada (w zależności od znaku $\hat{\beta}_1$) przeciętnie o* $\left|\left(e^{\hat{\beta}_i} - 1\right)100\right|$ %.

   → Zauważmy, że gdyby zmienną t uwzględnić w modelu w postaci zlogarytmowanej, tj. $\ln A_t = \beta_0 + \beta_1 \ln t + \cdots + \varepsilon_t$, wówczas parametr β_1 byłby elastycznością A_t względem czasu, a zatem początek interpretacji brzmiałby tak: *„Jeżeli wartość zmiennej czasowej wzrosłaby o 1% (przy ustalonych itd…) itd…”*, co jest bardzo nieintuicyjne, bo co niby miałby oznaczać *wzrost numeru okresu o 1%...?*



Model wykładniczy

- Kiedy do modelowania pewnej zmiennej ekonomicznej możemy chcieć zastosować ***model wykładniczy*** (zmienna objaśniana – w postaci logarytmicznej; zmienne objaśniające – bez takiej transformacji, tj. w swojej wyjściowej postaci) – trzy główne powody (c.d.):

2) Jeżeli w modelu chcemy uwzględnić jakiś regresor będący zmienną ekonomiczną wyrażoną w **procentach**, np. stopę bezrobocia, stopę inflacji, stopę procentową – także z powodu interpretacji oceny stojącego przy niej parametru. Jeżeli NIE zlogarytmujemy takiej zmiennej, wówczas w intepretacji będziemy mogli mówić o wzroście jej wartości o 1 *punkt procentowy* (np. wzrost stopy bezrobocia z 10% do 11%). W przeciwnym razie, gdybyśmy zlogarytmowali taką zmienną, wtedy musielibyśmy przejść na interpretację elastycznościową, która jest raczej nieintuicyjna w przypadku zmiennych wyjściowo wyrażonych w procentach, gdyż wymagałaby rozważania wzrostu takiej zmiennej dosłownie o 1% (zamiast punktu procentowego), co oznacza konieczność wyobrażenia sobie „*procentu z procentu*”.

Model wykładniczy

- Kiedy do modelowania pewnej zmiennej ekonomicznej możemy chcieć zastosować ***model wykładniczy*** (zmienna objaśniana – w postaci logarytmicznej; zmienne objaśniające – bez takiej transformacji, tj. w swojej wyjściowej postaci) – trzy główne powody (c.d.):

  3) Jeżeli w modelu chcemy uwzględnić jakiś regresor będący **zmienną licznikową**, tj. przyjmującą (z reguły niewielkie) wartości ze zbioru liczb naturalnych, np. liczba dzieci w rodzinie. Powodem, dla którego takiej zmiennej również nie poddajemy transformacji logarytmicznej jest znowu kwestia interpretacji – gdyby w sytuacji zmiennej objaśnianej w postaci zlogarytmowanej, zmienną objaśniającą licznikową również poddano logarytmizacji, wówczas parametr stojący przy tej ostatniej byłby elastycznością. Wtedy może okazać się „nienaturalne” mówienie o 1%-wym wzroście owej zmiennej licznikowej (bo np. cóż miałby oznaczać wzrost liczby dzieci w rodzinie o 1%??? :)

Model potęgowo-wykładniczy („mieszany”)

- **Model potęgowo-wykładniczy:**
  - Zmienna objaśniana – w logarytmie
  - Niektóre zmienne objaśniające – w logarytmie; pozostałe – w swojej wyjściowej postaci

$$A_t = \alpha_0 Z_{t1}^{\beta_1} Z_{t2}^{\beta_2} \dots Z_{tK_1}^{\beta_{K_1}} \alpha_1^{W_{t1}} \alpha_2^{W_{t2}} \dots \alpha_{K_2}^{W_{tK_2}} e^{\varepsilon_t} \quad /\ln$$

Przyjmując $\ln \alpha_0 = \beta_0$ oraz $\ln \alpha_i = \beta_{K_1+i}$ (dla $i = 1, 2, \dots, K_2$), otrzymujemy:

$$\underbrace{\ln A_t}_{y_t} = \beta_0 + \beta_1 \underbrace{\ln Z_{t1}}_{x_{t1}} + \cdots \beta_{K_1} \underbrace{\ln Z_{tK_1}}_{x_{tK_1}} + \beta_{K_1+1} \underbrace{W_{t1}}_{x_{t,K_1+1}} + \cdots \beta_{K_1+K_2} \underbrace{W_{tK_2}}_{x_{t,K_1+K_2}} + \varepsilon_t$$

Model potęgowo-wykładniczy – estymacja

Przypomnijmy

$$\underbrace{\ln A_t}_{y_t} = \beta_0 + \beta_1 \underbrace{\ln Z_{t1}}_{x_{t1}} + \cdots \beta_{K_1} \underbrace{\ln Z_{tK_1}}_{x_{tK_1}} + \beta_{K_1+1} \underbrace{W_{t1}}_{x_{t,K_1+1}} + \cdots \beta_{K_1+K_2} \underbrace{W_{tK_2}}_{x_{t,K_1+K_2}} + \varepsilon_t$$

Powyższy model spełnia już 1. założenie KMNLR (posiada addytywny składnik losowy, zaś część systematyczna jest liniowa względem parametrów). Zatem:

1) Da się zapisać w ogólnej postaci:

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_K x_{tK} + \varepsilon_t$$

gdzie:

- $K = K_1 + K_2$ (liczba parametrów BEZ wyrazu wolnego, β_0); $k = K + 1$ (… i Z wyrazem wolnym)
- $y_t \equiv \ln A_t$
- $x_{ti} \equiv \ln Z_{ti}$ dla $i = 1, \ldots, K_1$
- $x_{ti} \equiv W_{ti}$ dla $i = K_1 + 1, \ldots, K_1 + K_2$

2) jego parametry możemy oszacować za pomocą MNK: $\hat{\beta} = (X'X)^{-1}X'y$, przy czym:

$$y = \begin{bmatrix} \ln A_1 \\ \ln A_2 \\ \vdots \\ \ln A_T \end{bmatrix}, \qquad X = \begin{bmatrix} 1 & \ln Z_{11} & \ldots & W_{1,K_2} \\ 1 & \ln Z_{21} & \ldots & W_{2,K_2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \ln Z_{T1} & \ldots & W_{T,K_2} \end{bmatrix}_{(T\times k)}$$

Ocena dobroci dopasowania modelu z logarytmiczną zmienną objaśnianą

Ogólnie, zmienną objaśnianą (a ściślej, jej pojedynczą wartość o nr t) oznaczamy symbolem y_t (to nie nowość). Z powyższych rozważań wiemy już i musimy pamiętać, że pod terminem / oznaczeniem *zmiennej objaśnianej* możemy rozumieć – w zależności od specyfikacji modelu – pewną zmienną ekonomiczną (np. przychody ze sprzedaży, np. w tysiącach zł) ALBO jej *logarytm*. Jeśli wspomniane przychody ze sprzedaży oznaczymy jako S_t (ang. sales), wówczas:

- W pierwszym przypadku (tj. gdy y_t jest bezpośrednio zmienną ekonomiczną) zapiszemy $y_t \equiv S_t$
  → tym samym y_t jest wyrażony w tej samej co S_t jednostce, czyli w tysiącach zł.

- W drugim przypadku (tj. gdy y_t jest *logarytmem* zmiennej ekonomicznej) zapiszemy $y_t \equiv \ln S_t$
  → jednostką y_t jest wtedy *logarytm* jednostki zmiennej S_t, czyli *logarytm* tysiąca zł (jakkolwiek dziwnie to brzmi :)

Podsumowując, "samo" y_t oznacza zawsze *zmienną objaśnianą* (i zarazem wartości tejże), a tą ostatnią może być pewna zmienna ekonomiczna lub jej logarytm. W kontekście oceny dopasowania modelu regresji do danych ma to o tyle znaczenie, że w tym drugim przypadku oceniamy, w istocie, dopasowanie modelu regresji do wartości będących *logarytmem* oryginalnej zmiennej ekonomicznej (a nie do wartości samej zmiennej ekonomicznej). Będzie to musiało być uwzględnione na etapie interpretacji wartości obliczanych mierników

Ocena dobroci dopasowania modelu z logarytmiczną zmienną objaśnianą

- Przypomnijmy mierniki:
  - [tu wpisz nazwę] $s = \sqrt{s^2}$
  - [tu wpisz nazwę] $V_\varepsilon = \frac{s}{|\bar{y}|} \cdot 100$
  - [tu wpisz nazwę] $\varphi^2 = \frac{\sum_{t=1}^{T}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{T}(y_t - \bar{y})^2} = \frac{SSE}{\sum_{t=1}^{T}(y_t - \bar{y})^2}$
  - [tu wpisz nazwę] $R^2 = 1 - \varphi^2$

- Przejdźmy do dwóch przykładów – różniących się tylko i wyłącznie postacią zmiennej objaśnianej:
  - W przykładzie 1 będzie to zmienna bez logarytmu
  - W przykładzie 2 – zmienna zlogarytmowana

Ocena dobroci dopasowania modelu z logarytmiczną zmienną objaśnianą

Przykład 1. Rozważmy pewien model:

$$E_t = -34{,}2 + 7{,}1 I_t + 27{,}3 H_t + \hat{\varepsilon}_t,$$

oszacowany na podstawie 10 gospodarstw domowych, celem modelowania wielkości miesięcznych ich wydatków na transport (zmienna E_t, w zł/osobę w g.d.) w zależności od wielkości miesięcznych dochodów gospodarstwa domowego (zmienna I_t, w setkach zł / osobę w g.d.) i typu gospodarstwa: zmienna H_t typu 0-1, przy czym $H_t = 1$ dla gospodarstw pracowniczych, zaś $H_t = 0$ dla gospodarstw emerytów i rencistów. Celem oceny dobroci dopasowania rozważanego modelu do danych obliczono wartości stosownych mierników i otrzymano wyniki: $s = 15{,}29$ (← Pytanie: jaka jest jednostka?); $V_\varepsilon = 5{,}67\%$; $R^2 = 0{,}78$; $\varphi^2 = 0{,}22$ (Przypomnijmy: zachodzi $R^2 + \varphi^2 = 1$, gdyż w modelu występuje wyraz wolny – jego ocena to $\hat{\beta}_0 = -34{,}2$).

Interpretacje: (bierzemy tu pod uwagę konkretne znaczenie zmiennej objaśnianej, którą jest tu $y_t \equiv E_t$)

- $s = 15{,}29$; $V_\varepsilon = 5{,}67\%$ (obydwie wartości interpretujemy razem, w jednym zdaniu):
  Zaobserwowane wartości miesięcznych wydatków gospodarstw domowych na transport różnią się od ich wartości teoretycznych przeciętnie o ±15,29 zł/os. w g.d. [do tego miejsca mamy interpretację s; dalsza część zdania dotyczy V_ε], co stanowi 5,67% średniej arytmetycznej zaobserwowanych wartości tych wydatków.

  Ponadto, z uwagi na to, że $V_\varepsilon \leq 10\%$ możemy stwierdzić, że rozważany model charakteryzuje się bardzo dobrym dopasowaniem do danych.

Ocena dobroci dopasowania modelu z logarytmiczną zmienną objaśnianą

<u>Przykład 1.</u> - c.d.

- $R^2 = 0{,}78$ (w dwóch równoważnych brzmieniach – do wyboru):

  → Wersja 1: Całkowita zmienność wielkości miesięcznych wydatków gospodarstw domowych na transport została wyjaśniona w ramach oszacowanego modelu w 78%.

  → Wersja 2: 78% całkowitej zmienności wielkości miesięcznych wydatków gospodarstw domowych na transport zostało wyjaśnione w ramach oszacowanego modelu.

- $\varphi^2 = 1 - R^2 = 0{,}22$ (analogicznie jak współczynnik determinacji, tyle tylko, że dopisując „nie”…):

  → Wersja 1: Całkowita zmienność wielkości miesięcznych wydatków gospodarstw domowych na transport NIE została wyjaśniona w ramach oszacowanego modelu w 22%.

  → Wersja 2: 22% całkowitej zmienności wielkości miesięcznych wydatków gospodarstw domowych na transport NIE zostało wyjaśnione w ramach oszacowanego modelu.

Ocena dobroci dopasowania modelu z logarytmiczną zmienną objaśnianą

<u>Przykład 2.</u> Przykład ten stanowi subtelną (acz znaczącą :) modyfikację poprzedniego, przy czym psikus polega na tym, że tu zmienną objaśnianą jest *logarytm* owej wielkości miesięcznych wydatków gospodarstw domowych na transport:

$$\ln E_t = -34{,}2 + 7{,}1 I_t + 27{,}3 H_t + \hat{\varepsilon}_t.$$

Załóżmy, że wszystkie wyniki pozostają bez zmian (Pytanie: w jakiej jednostce jest teraz $s = 15{,}29$?). Jak logarytmiczna postać zmiennej objaśnianej ($y_t \equiv \ln E_t$) rzutuje na brzmienie poszczególnych interpretacji?

<u>Interpretacje:</u>

- $s = 15{,}29$; $V_\varepsilon = 5{,}67\%$ (interpretujemy razem, w jednym zdaniu):
  Zaobserwowane wartości logarytmu miesięcznych wydatków gospodarstw domowych na transport różnią się od ich wartości teoretycznych przeciętnie o ±15,29 logarytmu zł/os. w g.d. [do tego miejsca mamy interpretację s; dalsza część zdania dotyczy V_ε], co stanowi 5,67% średniej arytmetycznej zaobserwowanych wartości logarytmu tych wydatków.

  Ponadto, z uwagi na to, że $V_\varepsilon \leq 10\%$ możemy stwierdzić, że rozważany model charakteryzuje się bardzo dobrym dopasowaniem.

Ocena dobroci dopasowania modelu z logarytmiczną zmienną objaśnianą

<u>Przykład 2.</u> - c.d.

- $R^2 = 0{,}78$ (w dwóch równoważnych brzmieniach – do wyboru):

  → Wersja 1: Całkowita zmienność logarytmu wielkości miesięcznych wydatków gospodarstw domowych na transport została wyjaśniona w ramach oszacowanego modelu w 78%.

  → Wersja 2: 78% całkowitej zmienności logarytmu wielkości miesięcznych wydatków gospodarstw domowych na transport zostało wyjaśnione w ramach oszacowanego modelu.

- $\varphi^2 = 1 - R^2 = 0{,}22$ (analogicznie jak współczynnik determinacji, tyle tylko, że dopisując „nie”…):

  → Wersja 1: Całkowita zmienność logarytmu wielkości miesięcznych wydatków gospodarstw domowych na transport NIE została wyjaśniona w ramach oszacowanego modelu w 22%.

  → Wersja 2: 22% całkowitej zmienności logarytmu wielkości miesięcznych wydatków gospodarstw domowych na transport NIE zostało wyjaśnione w ramach oszacowanego modelu.