PROGNOZOWANIE W KMNRL

Łukasz Kwiatkowski

Katedra Ekonometrii i Badań Operacyjnych

Plan wykładu

1) Wprowadzenie
2) Predykcja punktowa
    a) Prognoza punktowa
    b) Pomiar niepewności prognozy punktowej
        i. Średni błąd predykcji *ex ante*
        ii. Względny błąd predykcji *ex ante*
3) Predykcja przedziałowa
4) Przykład
5) Predykcja w modelu z logarytmiczną zmienną objaśnianą
6) Prognozowanie probabilistyczne

Wprowadzenie

- Przypomnijmy punkt wyjścia naszych rozważań, czyli zapis równania regresji (w domyśle: spełniającej wszystkie założenia KMNRL):

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

lub zwięźlej, korzystając z rachunku wektorów:

$$y_t = x_t \beta + \varepsilon_t$$

gdzie $x_t = [1 \;\; x_{t1} \;\; x_{t2} \;\; \ldots \;\; x_{tK}]$, zaś $\beta = [\beta_0 \;\; \beta_1 \;\; \beta_2 \;\; \ldots \;\; \beta_K]'$ (← transpozycja konieczna)

- Zaraz we wprowadzeniu do materiału poświęconego testowi *t*-Studenta zadaliśmy sobie kluczowe pytanie: *do czego w praktyce można wykorzystać taki model (oczywiście, po jego oszacowaniu)?*

  Wskazaliśmy na jego dwa główne zastosowania:

  1) Do ***prognozowania*** wartości zmiennej objaśnianej, y_t (przy zadanych wartościach regresorów, czyli "iksów")
  2) Do ***wnioskowania statystycznego*** o prawdziwych (zawsze nieznanych i niepoznawalnych) wartościach parametrów, które odzwierciedlają wpływ poszczególnych zmiennych objaśniających na zmienną objaśnianą

  O ile zagadnienie wnioskowania statystycznego w KMNRL już w dużej mierze omówiliśmy (test *t*, przedziały ufności, test *F*), o tyle tematykę prognozowania (inaczej: predykcji) odraczaliśmy aż dotąd

- W niniejszej prezentacji dowiemy się, jak można wykorzystać KMNRL do prognozowania „y-ka", czyli zmiennej objaśnianej, choć o samych podstawach teoretycznych predykcji w statystyce/ekonometrii powiemy bardzo niewiele – jest to bowiem osobny dział, któremu poświęcony jest osobny przedmiot

Wprowadzenie

- Zanim przejdziemy do konkretów wyjaśnijmy tylko, że sam termin *prognoza* pochodzi od greckiego połączenia słów *pro-* (=przed) i *gnosis* (=wiedza), co w wolnym tłumaczeniu na język polski oznacza prze-widywanie, “przed-wiedzenie”, czyli wiedzę o lub znajomość czegoś zanim to się wydarzy, zanim to nastąpi. W języku polskim, terminem równoważnym do *prognozy* jest *predykcja*, etymologicznie powstały z połączenia łacińskich słów: *prae-* (=przed) i *dicere* (=mówić). Termin ten jest stosowany przede wszystkim właśnie w kontekście *statystycznych/ekonometrycznych* metod przewidywania.

Predykcja punktowa

- Przypomnijmy, że w równaniu KMNRL wyróżniamy dwie części: systematyczną (inaczej zwaną deterministyczną) oraz stochastyczną (składnik losowy, będący w KMNRL jedyną przyczyną losowości obserwowanej w wartościach zmiennej objaśnianej):

$$y_t = \underbrace{\beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_K x_{tK}}_{\substack{część\ systematyczna \\ (in.: deterministyczna)\ modelu}} + \underbrace{\varepsilon_t}_{\substack{część\ stochastyczna \\ modelu\ (składnik\ losowy)}}$$

- Z założeń KMNRL wynika, że gdyby nie występowały zaburzenia losowe (tj. $\varepsilon_t = 0$), wówczas – w świetle powyższego zapisu – przy danych/ustalonych wartościach zmiennych objaśniających $(x_{t1}, x_{t2}, \ldots, x_{tK})$ należy *przewidywać/oczekiwać*, że wartość zmiennej objaśnianej wyniesie dokładnie tyle, ile wynika z części systematycznej modelu, czyli $\beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_K x_{tK}$. Zwróćmy uwagę, że wyrażenie to stanowi nic innego jak wartość teoretyczną zmiennej objaśnianej (przy ustalonych „iksach"), tyle tylko, że jeszcze tą „nieoszacowaną" (bo z parametrami, a nie z ich ocenami). Nieco formalniej, mając na uwadze wartość oczekiwaną zmiennej objaśnianej możemy zapisać:

$$E(y_t) = E(\beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_K x_{tK} + \varepsilon_t) = E(\beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_K x_{tK}) + E(\varepsilon_t)$$
$$= \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_K x_{tK} = x_t \beta$$

przy czym w powyższym rozumowaniu wykorzystano fakt, że wartość oczekiwana sumy pewnych komponentów to suma ich wartości oczekiwanych ($E(X + Y) = E(X) + E(Y)$), oraz założenia KMNRL (wartość oczekiwana składnika losowego wynosi 0, zaś wszystkie „iksy" i parametry to pewne stałe, więc wartość oczekiwana z nich to te same stałe: $E(c) = c$, gdzie c to pewna wielkość stała, tj. nielosowa).

Predykcja punktowa

- Otrzymane wyrażenie: $\beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_K x_{tK}$ (czyli $x_t\beta$) moglibyśmy oznaczyć symbolem $\hat{y}_t$, gdyby nie to, że już wcześniej przywykliśmy nim oznaczać tak naprawdę oszacowanie tego wyrażenia – oszacowanie, w którym w miejsce parametrów wstawiamy ich oceny MNK:

$$\hat{y}_t = x_t\hat{\beta} = \hat{\beta}_0 + \hat{\beta}_1 x_{t1} + \hat{\beta}_2 x_{t2} + \cdots + \hat{\beta}_K x_{tK}$$

  I przy tej konwencji pozostaniemy, żeby nie wprowadzać niepotrzebnego zamętu :)

- Tak naprawdę, co do istoty, to powyższy zapis definiuje nam właśnie wartość prognozy punktowej zmiennej objaśnianej – przy ustalonych na danym poziomie wartościach zmiennych objaśniających. Jednak dla odróżnienia tych wartości teoretycznych w próbie (odpowiadających zaobserwowanym wartościom zmiennej objaśnianej), z którymi mieliśmy do czynienia dotychczas, od tych wartości teoretycznych, które wyznaczamy dla dowolnego, interesującego nas zestawu wartości poszczególnych zmiennych objaśniających – chcąc właśnie przy nich przewidzieć/prognozować wartość "y-ka" – te ostatnie będziemy oznaczać symbolem $\hat{y}_t^P$ (z indeksem górnym "P" – od słowa "predykcja"). Zatem różnica pomiędzy tymi dwiema *de facto* wartościami teoretycznymi tkwi tylko w kontekście, nie zaś w sposobie obliczania.

- Predykcji "y-ka" dokonujemy przy zadanych wartościach zmiennych objaśniających. W związku z tym, w kontekście predykcji będziemy także i te ostatnie wyróżniać poprzez umieszczenie symbolu "P" w indeksie górnym: x_{t1}^P, x_{t2}^P, …, x_{tK}^P

Predykcja punktowa

- Podsumowując, chcąc wykorzystać oszacowany model regresji do prognozy, w pierwszej kolejności zadajemy (w jakiś sposób, póki co "nieważne jaki") wartości zmiennych objaśniających i zestawiamy je w ich wektorze:

$$x_t^P = [1 \quad x_{t1}^P \quad x_{t2}^P \quad \dots \quad x_{tK}^P],$$

a następnie – wraz z wektorem ocen parametrów, $\hat{\beta}$ – podstawiamy do dobrze znanego wzoru na wartość teoretyczną, tym samym uzyskując wartość **prognozy punktowej** (w skrócie, po prostu: **prognozę**) zmiennej objaśnianej:

$$\hat{y}_t^P = x_t^P \hat{\beta}$$

- Można pokazać, że wzór ten definiuje tzw. **optymalny predyktor** – taki, który minimalizuje wartość oczekiwaną kwadratu błędu prognozy, gdzie przez ten ostatni rozumie się różnicę pomiędzy prognozowaną wartością (inaczej: wartością, która podlega prognozie, czyli y_t^P) a wartością prognozy ($\hat{y}_t^P$).

$\rightarrow$ Pewien "niepokój" może budzić sformułowanie "prognozowana wartość" – użyte powyżej w odniesieniu do y_t^P, a nie – jak można by się spodziewać – do $\hat{y}_t^P$. Tę subtelną kwestię językową wyjaśniamy na kolejnym slajdzie :)

Predykcja punktowa

- Wtręt terminologiczno-językowy

  W zagadnieniu prognozowania mamy do czynienia jednocześnie z dwiema wielkościami: prawdziwą/faktyczną/empiryczną wartością prognozowanej zmiennej (oznaczaną symbolem y_t^P) oraz jej prognozą ($\hat{y}_t^P$). Dobrze jest, w związku z tym, uporządkować sobie znaczenie i możliwości stosowania absolutnie podstawowego terminu, którym jest "**prognozowana wartość**". Jego użycie – choć powszechne – może w istocie być obarczone pewną niejednoznacznością, wynikającą z tego, że – jak to tu wyjaśnimy – sformułowanie to może być użyte zarówno w odniesieniu do y_t^P, jak i $\hat{y}_t^P$. Z jednej strony, pod pojęciem "prognozowana wartość" możemy rozumieć faktyczną (tą prawdziwą) wartość prognozowanej zmiennej, czyli y_t^P – jednak tylko wtedy, gdy owo sformułowanie pełni funkcję skrótu myślowego od "wartość, która podlega prognozie, która jest przedmiotem przeprowadzanego prognozowania". W tym właśnie znaczeniu zostało ono użyte na poprzednim slajdzie.

  Natomiast z drugiej strony, ten sam zwrot "prognozowana wartość" może oznaczać *de facto* uzyskaną już wartość samej prognozy, czyli $\hat{y}_t^P$ – wtedy, gdy przez to sformułowanie rozumiemy "wartość będącą efektem, skutkiem prognozowania; wy-prognozowaną wartość". Wydaje się, że to drugie użycie jest i powszechniejsze, i tym samym bezpieczniejsze…

Predykcja punktowa

- Przechodząc łagodnie do kolejnej kwestii podkreślmy raz jeszcze, że prognoza w modelu regresji ma charakter **warunkowy**, gdyż odbywa się przy ustalonych wartościach regresorów (przyjęcie innych wartości "iksów" będzie z reguły skutkowało inną wartością prognozy). Jeśli zaś chodzi o **sposób ustalania wartości zmiennych objaśniających**, przy których przeprowadzamy predykcję, to przede wszystkim najlepiej jest skorzystać z jakichś metod pomocniczych służących prognozie (o którym parę słów poniżej), lub ewentualnie zadać je samodzielnie (być może po prostu nas interesuje predykcja "y-ka" przy jakiś konkretnych "naszych iksach"; można także przeprowadzać prognozy scenariuszowe, rozważając jednocześnie różne zestawy "iksów"). W celu wyjaśnienia tych dwóch podejść rozważmy dwa konteksty – wyłaniające się z uwagi na typ danych, dla których zbudowano dane model regresji: *danych przekrojowych* albo *szeregów czasowych*.

Predykcja punktowa

➢ Predykcja w kontekście danych **przekrojowych** (in.: przestrzennych)

Tu wartości x_{t1}^P, x_{t2}^P, ..., x_{tK}^P możemy zadać zgodnie z tym, "co nas interesuje", bez konieczności posługiwania się jakimiś dodatkowymi metodami prognozy każdego z "iksów" z osobna. Jako przykład rozważmy sytuację, w której zbudowaliśmy model dla zaobserwowanych cen sprzedaży lokali mieszkaniowych w Krakowie w 2019 r., a do wyjaśnienia ich zróżnicowania (zmienności) posłużyliśmy się wybranymi ich trzema charakterystykami: liczba pokoi (x_{t1}), numer piętra (x_{t2}) i metraż (x_{t3}). Dajmy na to, że z jakichś przyczyn (bo np. znajomy szuka mieszkania do kupienia) chcemy się dowiedzieć, ile orientacyjnie (czyli przeciętnie, średnio rzecz biorąc) może kosztować 4-pokojowe mieszkanie na 2. piętrze, o łącznej powierzchni 70 m^2. Przyjmujemy zatem: $x_{t1}^P = 4$, $x_{t2}^P = 2$ oraz $x_{t3}^P = 70$, i zestawiamy te wartości w wektorze x_t^P:

$$x_t^P = [1 \quad x_{t1}^P \quad x_{t2}^P \quad x_{t3}^P] = [1 \quad 4 \quad 2 \quad 70]$$

Zauważmy, że w związku z (domyślnie) występującym w modelu wyrazem wolnym (jako pierwszy parametr w równaniu, β_0), na pierwszej współrzędnej wektora x_t^P musi się znajdować wartość „1".

Oszacowawszy wcześniej model na podstawie dostępnych danych (czyli pewnej próby mieszkań), prognozę punktową ceny takiego mieszkania uzyskujemy za pomocą opisanego wyżej wzoru na $\hat{y}_t^P$:

$$\hat{y}_t^P = x_t^P \hat{\beta}$$

...i *voilà*!

Predykcja punktowa

➢ Predykcja w kontekście **szeregów czasowych**

Przejdźmy teraz do drugiego kontekstu, a mianowicie do prognozy w modelu zbudowanego dla *danych czasowych*.

Dajmy na to, że zbudowaliśmy bardzo prosty model(-ik) dla szeregu czasowego przeciętnych miesięcznych cen mieszkań w Krakowie (y_t, w zł), a w charakterze zmiennych objaśniających – w sumie tylko dwóch – wykorzystaliśmy:

- $x_{t1} \equiv t$ czyli zwykły trend liniowy, ujmujący swoistą tendencją rozwojową cen na krakowskim rynku nieruchomości; tak więc $t = 1, 2, \dots, T$ oznacza numer kolejnych obserwacji – o częstotliwości miesięcznej, dajmy na to pochodzących z lat 2018 i 2019 (zatem dysponujemy $T = 24$ obserwacjami)
- x_{t2} – przeciętny miesięczny dochód gospodarstwa domowego w Krakowie (mierzący poziom zamożności Krakowian); zmienna wyrażona w zł/osobę w g.d.

Tak więc rozważamy model postaci: $y_t = \beta_0 + \beta_1 t + \beta_2 x_{t2} + \varepsilon_t$.

Przypuśćmy, że (całkiem naturalnie) interesowałaby nas prognoza przeciętnej miesięcznej ceny mieszkań w Krakowie na kolejny po ostatniej obserwacji w próbie okres wprzód, czyli na styczeń 2020 r.

Predykcja punktowa

➢ Predykcja w kontekście szeregów czasowych – c.d. (1)

Powstaje teraz **pytanie**: jakie należałoby przyjąć wartości zmiennych objaśniających, tj. t^P i x_{t2}^P?

Dla **trendu** odpowiedź jest oczywista: $t^P = T + 1 = 25$ ponieważ prognozujemy na pierwszy okres (miesiąc) poza próbą (gdybyśmy prognozowali na luty 2020, wówczas oczywiście $t^P = T + 2 = 26$, na marzec: $t^P = T + 3 = 27$ itd.).

Jeśli zaś chodzi o **wartość** $\boldsymbol{x_{t2}^P}$ (przeciętny miesięczny dochód g.d. w Krakowie), to tutaj musimy się zastanowić. Jasne, że możemy sobie zadać jakąś zupełnie arbitralną wartość („z kosmosu"). Nie wiadomo jednak, czy miałaby ona w ogóle sens, bo jeśli – przykładowo – przez ostatni kwartał 2019 r. (czyli 3 ostatnie obserwacje w próbie) odnotowaliśmy kolejno takie wartości zmiennej x_{t2}: 1500 zł, 1550 zł i 1580 zł, to czy jest sens przyjmować, że w kolejnym miesiącu (czyli już w styczniu 2020 r., tj. właśnie w naszym okresie prognozy) wartość ta wyniesie np. $x_{t2}^P = 2500$ zł? Dobrze mówicie(!): No raczej nie :) Choć, istotnie, od strony technicznej nic nie stoi po temu na przeszkodzie (wzory matematyczne są cierpliwe prawie jak papier – wiele przyjmą ;), bo jak do wzoru na prognozę wstawimy taką wartość, to z pewnością „jakiś tam" wynik otrzymamy – ale czy taka predykcja będzie wiarygodna? Taaak, dobrze mówicie(!): Nie ;)

→ Krótko mówiąc: żeby prognoza wartości zmiennej objaśnianej miała sens, to i przyjęte wartości zmiennych objaśniających muszą go mieć (a nieco bardziej formalnie: ekstrapolacja – czyli „przedłużenie" – wartości zmiennych objaśniających poza okres próby musi być uzasadniona :)

Predykcja punktowa

➢ Predykcja w kontekście szeregów czasowych – c.d. (2)

Powyższe rozumowanie prowadzi nas do kolejnego (znowu) pytania: To skąd wziąć – w naszym przykładzie – sensowną wartość x_{t2}^{P}, przy której należałoby sporządzić prognozę ceny mieszkań w Krakowie? Podejść jest kilka – jedne mniej, inne bardziej arbitralne (czyli uznaniowe). Odnośnie tych drugich, to najzwyczajniej w świecie my – jako autorzy tej prognozy – możemy po prostu uznać, że w styczniu 2020 r. wartość x_{t2} będzie wynosiła 1600 zł – spoglądając na poprzedzające ten okres wcześniejszy poziom i tendencję rozwojową tej zmiennej. Taka wartość wygląda na całkiem „sensowną", jednak nie bardziej niż np. 1590 zł czy 1610 zł. Wszystkie te „propozycje" są jednak w dalszym ciągu czysto uznaniowe (a zatem arbitralne), choć owszem, zadając je na „takim, a nie innym" poziomie sposób staramy się, aby nawiązywały one do tego, co ze zmienną x_{t2} działo się do momentu wyznaczania prognozy. Chcąc jednak być nieco mniej arbitralnym, najzwyczajniej w świecie możemy skonstruować dla zmiennej x_{t2} jakiś osobny model („na boku") i z niego wyznaczyć *prognozę* jej wartości na ów styczeń 2020 r., którą to prognozę, jako x_{t2}^{P}, możemy następnie wykorzystać do obliczenia prognozy y_t^P.

Predykcja punktowa

➢ Predykcja w kontekście szeregów czasowych – c.d. (3)

Także i tu powstaje jednak pytanie: jaki „pomocniczy” model zbudować dla zmiennej? Wybór jest dość szeroki – obok możliwości zbudowania po prostu jakiegoś osobnego modelu regresji, w którym zmienna x_{t2} była akurat zmienną objaśnianą (i, tym samym, prognozowaną), istnieje cała grupa (klasa) tzw. *modeli adaptacyjnych*, różniących się w swojej konstrukcji zasadniczo od typowych modeli regresji tym, że nie ma w nich zmiennych objaśniających, a całą dynamikę (trend, wahania sezonowe) danej zmiennej modeluje się za pomocą różnych technik „wygładzania” (ang. *smoothing*) jej zaobserwowanych wartości. Wśród nich najpopularniejsza jest rodzina tzw. modeli wygładzania wykładniczego (ang. *exponential smoothing models/methods*). Najprawdopodobniej zapoznają się z nimi Państwo na innym przedmiocie :)

Predykcja punktowa

➢ Predykcja w kontekście szeregów czasowych – c.d. (4)

Nierzadko, jako takie „pomocnicze” modele prognostyczne wykorzystywane są też tzw. *modele autoregresyjne* (ang. *autoregression models*), w których – jak sama nazwa wskazuje – do objaśnienia *bieżącej* wartości danej zmiennej (w naszym przykładzie byłaby to zmienna x_{t2}) wykorzystuje się zmienne objaśniające (jak w regresji) będące jednak tylko *opóźnionymi wartości samej zmiennej objaśnianej* (czyli tu: $x_{t-1,2}$, $x_{t-2,2}$ itd.), a nie jakimiś dodatkowymi, „zewnętrznymi” regresorami (jak w typowej regresji). W ogólnej postaci – dla pewnej zmiennej z_t – model autoregresyjny rzędu $p \in \{1, 2, \dots\}$ (oznaczany w dalszym ciągu jako AR(*p*)) możemy zapisać w następujący sposób:

$$z_t = \alpha_0 + \alpha_1 z_{t-1} + \alpha_2 z_{t-2} + \cdots \alpha_p z_{t-p} + u_t$$

gdzie z_{t-i} oznacza wartość i-tego opóźnienia zmiennej objaśnianej ($i = 1, 2, \dots, p$), $\alpha_0, \alpha_1, \dots, \alpha_p$ są parametrami, zaś u_t – składnikiem losowym. Wartość p nazywamy *rzędem autoregresji* i oznacza ona numer najdalszego opóźnienia uwzględnionego w modelu. Rząd autoregresji ustalamy arbitralnie lub za pomocą stosownych mierników dopasowania. Model AR może być traktowany jako regresja liniowa, lecz już nie jako KMNRL, gdyż nie jest spełnione 2. założenie tego ostatniego, a mianowicie o wartościach zmiennych objaśniających **nie** możemy już założyć (mniej lub bardziej „frywolnie”), że są wielkościami nielosowymi (jak to się właśnie zakłada w KMNRL), gdyż na mocy konstrukcji powyższego równania wszystkie wartości zmiennej z_t (dla dowolnego t, więc także z_{t-1}, z_{t-2} itd.) – będąc funkcjami składników losowych (a zatem zmiennych losowych) – są zmiennymi losowymi. Niemniej, modele AR możemy szacować za pomocą znanej nam już MNK, o czym powiemy sobie jednak innym razem...

Predykcja punktowa

- **Podsumowując:** w KMNRL, posiadając oceny parametrów modelu, zestawione w wektorze $\hat{\beta}$, oraz "jakoś tam" dobrane/zadane wartości zmiennych objaśniających, zestawione w wektorze

$$x_t^P = [1 \quad x_{t1}^P \quad x_{t2}^P \quad \dots \quad x_{tK}^P]$$

prognozę punktową zmiennej objaśnianej, y_t, obliczamy za pomocą formuły:

$$\hat{y}_t^P = x_t^P \hat{\beta}$$

- **Jednostka**, w której wyrażona jest wartość prognozy jest oczywiście ta sama, co i prognozowanej zmiennej (czyli zmiennej objaśnianej)

- **Interpretację** wartości prognozy punktowej – zresztą bardzo naturalną i oczywistą – omówimy na przykładach dopiero w dalszej części tego opracowania

- Tymczasem przejdźmy do omówienia sposobu pomiaru **niepewności predykcji punktowej**, gdyż – podobnie jak się to miało z ocenami parametrów – samo stwierdzenie, że prognoza "czegoś tam" wynosi np. 300 000 zł jeszcze "wiosny nie czyni", bowiem *przewidywanie punktowe* (podobnie jak punktowe oceny parametrów modelu) musimy jeszcze uzupełnić o ocenę naszej *niepewności* z nim związanej, czyli innymi słowy, o informację: o ile, mniej więcej, możemy się mylić przewidując "coś tam" na poziomie tych 300 000 zł. Przechodzimy zatem do...

Pomiar niepewności prognozy punktowej: średni błąd predykcji *ex ante*

- Ocena niepewności związanej z prognozą punktową polega – całkiem intuicyjnie – na określeniu tego, o ile (średnio rzecz biorąc) możemy się mylić prognozując wartość danej zmiennej objaśnianej na poziomie właśnie jej prognozy punktowej – oczywiście:
  - w stosunku do prawdziwej/faktycznej wartości tej zmiennej;
  - w jednostce tej samej, co i prognozowana zmienna;
  - przy danym zestawie wartości zmiennych objaśniających.
- Ponieważ przedmiotem tego pomiaru jest przeciętna różnica pomiędzy faktyczną wartością danej zmiennej a wartością jej prognozy (czyli pewną wartością "teoretyczną"), do jego przeprowadzenia potrzebujemy jakiegoś "odpowiednika" – uwaga: nie, nie błędu średniego szacunku (mierzącego przeciętną różnicę pomiędzy prawdziwą a oszacowaną wartością *parametru*), ale – odchylenia resztowego (*s*, wyznaczanego w celu określenia przeciętnej różnicy pomiędzy zaobserwowanymi a teoretycznymi wartościami zmiennej objaśnianej). Tym "odpowiednikiem" odchylenia resztowego wyznaczanym tu dla pojedynczej prognozy jest tzw. **średni błąd predykcji *ex ante*.**[1]

[1] Ten łaciński dopisek "*ex ante*" występujący w nazwie rozważanego miernika oznacza, że wartość tego ostatniego wyznaczamy *przed* poznaniem/zaobserwowaniem faktycznej wartości zmiennej – wtedy bowiem nie jesteśmy w stanie określić dokładnego błędu prognozy (czyli zwykłej różnicy pomiędzy wartością prognozy a *zaobserwowaną* wartością prognozowanej zmiennej), więc możemy "tylko" określić, jakiej mniej więcej (czyli przeciętnie, średnio rzecz biorąc) pomyłki możemy się *spodziewać*. Nie będziemy się tu jednak zajmować tematem tzw. błędów predykcji *ex post* (czyli tych, które wyznaczamy *po* zaobserwowaniu i w stosunku właśnie do faktycznej, zrealizowanej wartości zmiennej prognozowanej) – prawdopodobnie zostaną one omówione w ramach osobnego przedmiotu.

Pomiar niepewności prognozy punktowej: średni błąd predykcji *ex ante*

- Zanim podamy jego wzór, chciejmy wpierw zrozumieć, czym w ogóle jest. W tym celu wprowadźmy tu dość oczywistą w tym kontekście wielkość – po prostu, „zwyczajny" **błąd prognozy:**

$$\hat{\varepsilon}_t^P = y_t^P - \hat{y}_t^P$$

który stanowi „zwykłą" różnicę pomiędzy faktyczną wartością prognozowanej zmiennej (y_t^P) a prognozowaną wartością tej zmiennej (inaczej, po prostu: wartością prognozy, $\hat{y}_t^P$)

- Pomijając tu szczegółowe wyprowadzenie, powiedzmy sobie tylko tyle, że zapowiedziany wcześniej **średni błąd predykcji *ex ante*** stanowi nic innego, jak oszacowanie odchylenia standardowego błędu prognozy (podobnie jak odchylenie resztowe stanowi oszacowanie odchylenia standardowego reszt) i dlatego informuje nas o przeciętnej różnicy pomiędzy y_t^P a $\hat{y}_t^P$. Można pokazać, że wariancja błędu prognozy wynosi $\sigma^2[1 + x_t^P(X'X)^{-1}x_t^{P\prime}]$, więc przechodząc na odchylenie standardowe pierwiastkujemy to wyrażenie; dodatkowo, w miejsce *nieznanej* wariancji składnika losowego (σ^2) wstawiamy jej *oszacowanie*, czyli wariancję resztową, s^2.

- W związku z powyższym wyjaśnieniem, **średni błąd predykcji *ex ante*** dany jest formułą:

$$s_t^P = \sqrt{s^2[1 + x_t^P(X'X)^{-1}x_t^{P\prime}]}$$

→ Uwaga: ponieważ w praktyce możemy chcieć wyznaczać kilka prognoz jednocześnie, dlatego indeks „t" w oznaczeniu s_t^P nie powienien być pomijany (dla różnych prognoz możemy wtedy zapisać: s_1^P, s_2^P itd.). Jeśli jednak przedmiotem prognozowania jest tylko jedna wartość, wówczas indeks ten jest nadmiarowy – można go pominąć (zapisując s^P), a można też „po prostu" zostawić (zapisując s_t^P).

Pomiar niepewności prognozy punktowej: średni błąd predykcji *ex ante*

- Przyglądając się formule na s_t^P zwróćmy uwagę, **czego potrzebujemy** do jej obliczenia:
  - wariancji resztowej (s^2) – oczywiście tej w modelu, za pomocą którego dokonujemy prognozy
  - wektora x_t^P oraz jego transpozycji ($x_t^{P\prime}$)
  - macierzy $(X'X)^{-1}$ – tę albo już mamy (jeśli formułę na $\hat{\beta} = (X'X)^{-1}X'y$ wprowadzaliśmy w arkuszu ręcznie „krok po kroku"), albo potrzebujemy ją właśnie teraz obliczyć gdzieś „na boku"

  → **Pytanie na pomyślunek**: jakich wymiarów będzie wyrażenie $x_t^P(X'X)^{-1}x_t^{P\prime}$, a w konsekwencji także samo s_t^P?

- Wartość s_t^P (która jest wyrażona w tej samej jednostce, co i prognozowana zmienna) informuje nas, o ile średnio rzecz biorąc możemy się mylić prognozując wartość zmiennej objaśnianej na poziomie $\hat{y}_t^P$ (przy zadanych wartościach zmiennych objaśniających) – oczywiście w stosunku do prawdziwej wartości prognozowanej zmiennej.

Pomiar niepewności prognozy punktowej: względny błąd predykcji *ex ante*

- Zauważmy, że podanie samej wartości s_t^P, przykładowo $s_t^P = 1000$ zł jeszcze nie pozwala nam „odczuć" tego, czy to dużo, czy też mało. Odruchowo chcielibyśmy odnieść tę wartość do wartości samej prognozy, celem określenia *względnego* błędu predykcji, wyrażonego procentowo jako ułamek wyznaczonej wcześniej wartości prognozy punktowej.
- Powyższe rozumowanie prowadzi nas do zdefiniowania tzw. **względnego błędu predykcji *ex ante*:**

$$V_t^P = \frac{s_t^P}{|\hat{y}_t^P|} 100$$

gdzie przemnożenie przez czynnik „100" zapewnia wyrażenie wartości miernika w procentach.

- Do **interpretacji** wartości V_t^P "zabieramy" się analogicznie, jak w przypadku współczynnika zmienności resztowej, V_ε (przypomnijmy: wykorzystywanego w ocenie dobroci dopasowania modelu do danych). O ile wartość V_ε interpretowaliśmy w połączeniu z odchyleniem resztowym, s (w jednym, rozbudowanym zdaniu), o tyle w ten sam sposób postąpimy tutaj, łącząc w jednym zdaniu interpretację obydwu rozważanych błędów predykcji *ex ante*: średniego (s_t^P) i względnego (V_t^P). Konkretny **przykład** zostanie zaprezentowany w dalszej części tego materiału (pod koniec).

- Względny błąd predykcji *ex ante* można wykorzystać do określenia tzw. **dopuszczalności prognozy**: mówimy, że prognoza na poziomie $\hat{y}_t^P$ jest **dopuszczalna**, jeśli V_t^P **nie przekracza** 10% ($V_t^P \leq 10\%$). Zaznaczmy jednak, że ta progowa wartość 10% jest tylko umowna, zwyczajowa – nie wynika bowiem z żadnych *formalnych* przesłanek; jest jednak szeroko przyjęta wśród badaczy, w szczególności w obszarze nauk społeczno-gospodarczych.

Predykcja przedziałowa

- Podobnie, jak w przypadku *estymacji parametrów* modelu regresji, także i w przewidywaniu wartości zmiennej objaśnianej możemy chcieć wyznaczyć – oprócz prognozy punktowej i mierników jej niepewności – pewien zakres/przedział liczbowy najbardziej wiarygodnych/prawdopodobnych wartości prognozowanej zmiennej, czyli taki odpowiednik przedziału ufności konstruowanego dla pojedynczego współczynnika regresji (parametru).
- W kontekście prognozowania taki przedział nazywamy **przedziałem predykcji** (ang. *prediction interval*) lub po prostu **prognozą przedziałową** (ang. *interval forecast/prediction*)[2]
- Na kolejnym slajdzie zaprezentowano szkic wyprowadzenia formuły na prognozę przedziałową w KMNRL – nieobowiązkowe, a tylko dla "ciekawskich" :)

[2] Choć być może nasuwają się też sformułowania w stylu *przedział predyktywny* lub *przedział prognostyczny*, tak jednak terminy te nie są oficjalnie stosowane w literaturze polskojęzycznej (z wyjątkiem pojęcia *przedziału predyktywnego*, lecz na gruncie nieco innego rodzaju statystyki, a mianowicie tzw. statystyki bayesowskiej, którą się tutaj nie zajmujemy). Przy okazji warto też zwrócić uwagę, że sformułowanie *przedział ufności dla prognozy* jest fundamentalnie niepoprawne (z przyczyn, których wyjaśnienie w tym miejscu pomijamy), jakkolwiek skojarzenie prognozy przedziałowej z przedziałem ufności dla parametru jest w istocie dość "odruchowe".

Predykcja przedziałowa

- Ustalmy oznaczenia:
  - $y_t^P = x_t^P \beta + \varepsilon_t^P$ – prawdziwa, nieznana (niezaobserwowana – być może jeszcze) wartość zmiennej objaśnianej (przy zadanym zestawie wartości zmiennych objaśniających, x_t^P)

    $\rightarrow$ to właśnie y_t^P chcemy prognozować i czynimy to za pomocą...
  - $\hat{y}_t^P = x_t^P \hat{\beta}$ – prognoza punktowa wartości y_t^P
- Przy założeniach KMNRL można pokazać, że wyrażenie $t(\hat{y}_t^P) = \frac{y_t^P - \hat{y}_t^P}{s_t^P}$ ma rozkład t-Studenta o $T-k$ stopniach swobody, co zapisujemy: $t(\hat{y}_t^P) = \frac{y_t^P - \hat{y}_t^P}{s_t^P} \sim St(T-k)$
- Przyjmując, że naszym celem jest skonstruowanie takiego przedziału, w którym wartość prognozowanej zmiennej (y_t^P) znajdowałaby się z pewnym odgórnie ustalonym, wysokim (bliskim 1) prawdopodobieństwem $1-\alpha$ (typowe wartości to oczywiście: 0,9; 0,95, 0,99), możemy wykorzystać powyższy fakt, podobnie, jak czyniliśmy to przy wyprowadzaniu przedziału ufności dla pojedynczego parametru modelu regresji, a mianowicie:

$$\Pr\left\{\left|\frac{y_t^P - \hat{y}_t^P}{s_t^P}\right| < t_{\alpha/2}\right\} = 1-\alpha$$

$$\Leftrightarrow \Pr\left\{-t_{\alpha/2} < \frac{y_t^P - \hat{y}_t^P}{s_t^P} < t_{\alpha/2}\right\} = 1-\alpha$$

$$\Leftrightarrow \Pr\left\{\hat{y}_t^P - t_{\alpha/2}\, s_t^P < y_t^P < \hat{y}_t^P + t_{\alpha/2}\, s_t^P\right\} = 1-\alpha$$

Predykcja przedziałowa

- Z powyższych rozważań otrzymujemy wzór na $(\mathbf{1}-\boldsymbol{\alpha})\cdot\mathbf{100\%}$ **prognozę przedziałową** (czy inaczej **przedział predykcji**):

$$\left(\hat{y}_t^P - t_{\alpha/2}\, s_t^P;\ \ \hat{y}_t^P + t_{\alpha/2}\, s_t^P\right)$$

  gdzie $t_{\alpha/2}$ oznacza „tradycyjną", dwustronną wartość krytyczną w rozkładzie *t*-Studenta o $T-k$ stopniach swobody – dokładnie tę samą, którą wykorzystujemy w budowie przedziałów ufności.

- Powyższa formuła – zanim jeszcze wstawimy do niej *konkretne wartości prognozy punktowej i średniego błędu ex ante* uzyskane dla *konkretnego* zbioru danych (czyli próby) – stanowi pewien **przedział losowy**. Natomiast po dokonaniu owego podstawienia uzyskamy oczywiście już *konkretny* **przedział liczbowy**, stanowiący pojedynczą (bo uzyskaną dla „danych danych") realizację tego wcześniejszego.

- Podobnie jak można to wykazać w przypadku przedziału ufności skonstruowanego dla parametru modelu regresji, że jest on najkrótszy (dot. tego *losowego* przedziału ufności, a nie liczbowego, będącego tylko pojedynczą realizacją tego pierwszego), tak i przedział predykcji – ten losowy, dany powyższym wzorem – przy założeniach KMNRL – też jest **najkrótszym** spośród wszystkich (nieprzeliczalnie wielu) takich przedziałów, które **z prawdopodobieństwem** $1-\alpha$ zawierają w sobie prawdziwą wartość prognozowanej zmiennej (oczywiście, przy zadanych wartościach regresorów).

Predykcja przedziałowa

- Jeśli zaś chodzi o **interpretację** już konkretnego, liczbowego przedziału predykcji, to i tu – znowu, podobnie jak w przypadku przedziałów ufności dla parametrów – natrafiamy na pewne problemy. Darując sobie jednak ich dyskusję (a przyznam, że w moim odczuciu byłaby ona jeszcze subtelniejsza niż w przypadku przedziałów ufności :), przyjmijmy, że „dostatecznie poczciwą" interpretacją konkretnego już liczbowego przedziału predykcji, dajmy na to (a, b), będzie jedno z dwóch zaproponowanych niżej szablonowych sformułowań (do wyboru):
  - „Istnieje $(1-\alpha)\cdot 100\%$ szans, że (faktyczna) wartość prognozowanej zmiennej – przy ustalonych wartościach zmiennych objaśniających – mieści się w przedziale od a do b."
  - „Z wiarygodnością $(1-\alpha)\cdot 100\%$ (faktyczna) wartość prognozowanej zmiennej – przy ustalonych wartościach zmiennych objaśniających – mieści się w przedziale od a do b."

  $\rightarrow$ Oczywiście, końce przedziału predykcji (a, b) są wyrażone w tej samej jednostce, co i prognozowana zmienna

- O co chodzi w sformułowaniach: „istnieje $(1-\alpha)\cdot 100\%$ szans" lub równoważnie „z wiarygodnością $(1-\alpha)\cdot 100\%$"? Znaczą one tyle, co stwierdzenie, że – czysto hipotetycznie (bo w praktyce jest to absolutnie niemożliwe) – gdybyśmy nieskończenie wiele razy generowali losowo zbiór modelowanych obserwacji (czyli próbę, wektor y), uzyskując następnie na ich podstawie też nieskończenie wiele wektorów ocen parametrów ($\hat{\beta}$) w danym modelu, a przez to także i nieskończenie wiele wartości prognoz $\hat{y}_t^P$ (zawsze przy tym samym wektorze x_t^P), wówczas prawdziwa wartość prognozowanej zmiennej, czyli y_t^P (bez daszka!), mieściła by się w $(1-\alpha)\cdot 100\%$ przypadków spośród tych wszystkich przedziałów.

Przykład

- Czas na jakiś przykład. Wykorzystajmy po temu jedną z dwóch wcześniej już zarysowanych ilustracji prezentujących podstawowe kwestie związane z prognozowaniem osobno w przypadku modelowania *danych przekrojowych* i tych w postaci *szeregów czasowych*. Na potrzeby tego przykładu zajmiemy się tym pierwszym przypadkiem – modelem dla danych przekrojowych.

Przykład [Model dla cen mieszkań w Krakowie – dane przekrojowe] (zbiór danych: „Czyżyny.xlsx")

Rozważmy model postaci: $P_t = \beta_0 + \beta_1 R_t + \beta_2 F_t + \beta_3 A_t + \varepsilon_t$, gdzie:

- P_t – cena transakcyjna sprzedaży mieszkania w Krakowie w 2019 r. (w zł – za cały lokal, nie za 1 m^2)
- R_t – liczba pokoi danego mieszkania
- F_t – numer kondygnacji, na której znajduje się lokal ($F_t = 0$ oznacza parter)
- A_t – wielkość powierzchni lokalu (w m^2)

Powyższy model oszacowano na podstawie odpowiedniego zbioru danych (dotyczącego $T = 630$ lokali), otrzymując równanie: $\hat{P}_t = 55326{,}4 - 6536{,}85R_t + 740{,}93F_t + 5472{,}47A_t$. **Nasze zadanie**: Jakiej ceny należy się spodziewać w przypadku 4-pokojowego mieszkania na 2. piętrze, o łącznej powierzchni 70 m^2?

$\rightarrow$ Zauważmy, że tak sformułowane pytanie *nie wydaje nam konkretnego polecenia* "zrób to lub tamto", w sensie: "wyznacz prognozę punktową ceny mieszkania", albo "wyznacz prognozę przedziałową ceny mieszkania". I tak też to często wygląda w życiu :) "Szef" zadaje tylko pytanie, natomiast to po naszej stronie leży świadomość możliwych do wykorzystania narzędzi, umiejętność ich wykorzystania i interpretacji wyników – komunikatywnie, tak, żeby "szef" (niekoniecznie ekonometryk) zrozumiał :)

Przykład – c.d. (1)

- Od strony operacyjnej – podsumowując tę prezentację – znamy nasz repertuar narzędzi:
  - Prognoza punktowa – koniecznie wraz z oceną jej niepewności za pomocą:
    - Średniego błędu predykcji *ex ante*
    - Względnego błędu predykcji *ex ante*
  - Prognoza przedziałowa
- Czego potrzebujemy do wyznaczenia tych powyższych – to powiedzą nam już same wzory.
- Zaczynamy od <u>prognozy punktowej</u>:

Zapiszmy wzór ogólny: $\hat{y}_t^P = x_t^P \hat{\beta}$

Ta transpozycja jest tu konieczna

„U nas”:

$y_t \equiv P_t, \quad x_t^P = [1 \;\; R_t^P \;\; F_t^P \;\; A_t^P] = [1 \;\; 4 \;\; 2 \;\; 70]$, zaś $\hat{\beta} = [55326{,}4 \quad -6536{,}85 \quad 740{,}93 \quad 5472{,}47]'$

Wobec czego:

Od razu zapiszmy jednostkę

$$\hat{P}_t^P = [1 \;\; 4 \;\; 2 \;\; 70] \begin{bmatrix} 55326{,}4 \\ -6536{,}85 \\ 740{,}93 \\ 5472{,}47 \end{bmatrix} = 413733{,}76 \text{ zł}$$

→ Interpretację zostawimy sobie na koniec, gdy już otrzymamy wszystkie wyniki liczbowe

→ Zauważmy, że nie ma to znaczenia, czy lokal o rozważanych charakterystykach x_t^P znalazł się wcześniej w próbie, na podstawie której oszacowano model, czy też nie

Przykład – c.d. (2)

➢ Średni błąd predykcji *ex ante*:

Zapiszmy wzór: $$s_t^P = \sqrt{s^2[1 + x_t^P (X'X)^{-1} x_t^{P\prime}]}$$

→ Zatem potrzebujemy:

- wariancji resztowej – powiedzmy, że ją już policzyliśmy i otrzymaliśmy:

  $s^2 = 2\,062\,365\,262{,}669$ (zł²) (niestety, faktycznie wychodzi taki „brzydki" wynik)

- macierzy $(X'X)^{-1}$ – łatwą ją obliczyć w arkuszu Excel – tu podaję wynik dla macierzy $X'X$, gdyż to obejdzie się akurat bez konieczności dokonywania nadmiernych zaokrągleń (nieuniknionych dla odwrotności tej macierzy); macierz odwrotną mogą Państwo łatwo uzyskać w Excelu:

$$X'X = \begin{bmatrix} 630 & 1710 & 2000 & 38204{,}24 \\ & 5138 & 5564 & 113019{,}63 \\ & & 10408 & 122354{,}83 \\ & & & 2550345{,}195 \end{bmatrix}$$

→ Przy okazji, co z brakującymi elementami tej macierzy?

Przykład – c.d. (3)

→ Aby obliczyć w Excelu formułę na średni błąd predykcji *ex ante*, będzie potrzebne kilka operacji:

- pierwiastek (z liczby *x*): =PIERWIASTEK(*x*) albo =*x*^0,5
- zwykłe mnożenie dwóch skalarów: s^2 i wartości wyrażenia w nawiasach kwadratowych (który też będzie skalarem, czyli wymiarów 1×1)
- mnożenie macierzy, ALE TRÓJargumentowe: x_t^P razy $(X'X)^{-1}$ razy $x_t^{P\prime}$, gdzie przy ostanim argumencie musimy pamiętać o transpozycji (=TRANSPONUJ())

  → **Pytanie na pomyślunek**: skoro funkcja =MACIERZ.ILOCZYN, dostępna w arkuszu Excel, jest tylko DWUargumentowa, tj. dopuszcza („obsługuje”) mnożenie przez siebie tylko dwóch macierzy, to co zrobić, aby przemnożyć przez siebie TRZY macierze?

  (Odpowiedź już na następnej stronie, ale najpierw pomyśl :)

Przykład – c.d. (4)

→ Ponieważ funkcja =MACIERZ.ILOCZYN jest tylko DWUargumentowa, w celu przemnożenia przez siebie TRZECH macierzy potrzebne będzie dodatkowe zagnieżdżenie (wywołanie) tej funkcji w niej samej – w miejscu jej pierwszego albo drugiego argumentu (obojętnie). Tak więc wyrażenie $x_t^P (X'X)^{-1} x_t^{P\prime}$ możemy obliczyć na dwa, zupełnie równoważne sposoby (więc w praktyce będzie to obojętne, z którego skorzystamy):

- z zagnieżdżeniem na pierwszym argumencie:
  =MACIERZ.ILOCZYN(MACIERZ.ILOCZYN(x_t^P; $(X'X)^{-1}$); TRANSPONUJ(x_t^P))

- z zagnieżdżeniem na drugim argumencie:
  =MACIERZ.ILOCZYN(x_t^P; MACIERZ.ILOCZYN($(X'X)^{-1}$; TRANSPONUJ(x_t^P))

→ W obydwu wyrażeniach proszę zwrócić uwagę na nawiasy!

→ W tym momencie możemy już obliczyć wartość s_t^P – spróbuj to zrobić samodzielnie w arkuszu; prawidłowy wynik to:

Od razu zapiszmy jednostkę

$$s_t^P \approx 45629{,}57 \text{ zł}$$

→ Interpretację zostawiamy na koniec

Przykład – c.d. (5)

- Względny błąd predykcji *ex ante*:

  Zapiszmy wzór:

$$V_t^P = \frac{s_t^P}{|\hat{P}_t^P|} 100$$

  Po prostym podstawieniu otrzymujemy wynik:

$$V_t^P \approx 11{,}03\%$$

  → Interpretację zostawiamy na koniec

- Przedział predykcji

  Zapiszmy wzór: $\left(\hat{P}_t^P - t_{\alpha/2}\, s_t^P;\ \ \hat{P}_t^P + t_{\alpha/2}\, s_t^P\right)$

  Poziom ufności – przyjmujemy domyślny (gdyż żaden inny nie został nam zadany): $1 - \alpha = 0{,}95$

  Odpowiada mu poziom istotności $\alpha = 0{,}05$, co przy liczbie stopni swobody $T - k = 630 - 4 = 626$ prowadzi do wyznaczenia (dwustronnej) wartości krytycznej w rozkładzie *t*-Studenta: $t_{\alpha/2} = 1{,}964$

  Po podstawieniu otrzymujemy wynik:

$$(324117{,}28 \text{ zł};\ \ 503350{,}20 \text{ zł})$$

  ← pamiętamy o jednostkach

Przykład – c.d. (6)

➢ Zestawienie wyników:

- Wyniki dla prognozy punktowej (z oceną niepewności):

$$\hat{P}_t^P = 413733{,}76 \text{ zł}$$

$$s_t^P \approx 45629{,}57 \text{ zł}$$

$$V_t^P \approx 11{,}03\%$$

- 95% prognoza przedziałowa: $(324117{,}28 \text{ zł};\ 503350{,}20 \text{ zł})$

➢ Interpretacje – tu sformułowane zbiorczo:

Prognozowana wysokość ceny 4-pokojowego mieszkania na 2. piętrze, o łącznej powierzchni 70 m^2 wynosi ok. 413 733,76 zł, przy czym – w stosunku do faktycznej ceny takiego lokalu – możemy się mylić przeciętnie o *plus/minus* 45 629,57 zł, co stanowi ok. 11,03% wartości prognozy. Z uwagi na to, że ta ostatnia wartość (tj. względny błąd predykcji *ex ante*) przekracza 10% wydaje się, że otrzymana prognoza powinna być uznana za niedopuszczalną. Z drugiej strony, przekroczenie to jest nieznaczne (zaledwie 1,03 punktu procentowego), wobec czego można je uznać za zaniedbywalne, a samą prognozę – za dopuszczalną / na granicy dopuszczalności.

W końcu, z wiarygodnością 95% faktyczna cena sprzedaży rozważanego lokalu mieszkaniowego mieści się w przedziale od 324 117,28 zł do 503 350,20 zł.

Predykcja w modelu z logarytmiczną zmienną objaśnianą…

- … czyli o tym, co zrobić w sytuacji, gdy w danym modelu regresji zmienną objaśnianą jest *logarytm* pewnej zmiennej ekonomicznej, a my chcemy prognozować (oczywiście!) wartość właśnie tej ostatniej, a nie jej logarytmu.

- Zauważmy, że wszystkie przedstawione dotychczas treści dotyczyły po prostu prognozowania wartości zmiennej objaśnianej w modelu regresji liniowej, bez względu na to, "co" jest tą zmienną – a jak pamiętamy, może nią być albo pewna zmienna ekonomiczna, albo jej logarytm (do takich dwóch wariantów się ograniczamy, choć w bardziej zaawansowanych modelach – zwłaszcza prognostycznych – czasami wykorzystuje się bardziej "fikuśne" transformacje, jak np. tzw. transformację Boxa-Coxa).

- Tak więc przedstawiona dotychczas metodyka w całości i bezpośrednio pozwala nam przeprowadzić proces predykcji w sytuacji, gdy zmienną objaśnianą jest pewna zmienna ekonomiczna, w swej "czystej", nietrasformowanej postaci. Natomiast w przypadku, gdy zmienną objaśnianą jest jednak *logarytm* pewnej zmiennej ekonomicznej, wówczas wszystkie omówione dotychczas formuły odnoszą się do prognozy punktowej i przedziałowej właśnie *logarytmu* tejże zmiennej, nie zaś niej samej.

Predykcja w modelu z logarytmiczną zmienną objaśnianą…

- W dalszych rozważaniach przyjmiemy następującą **konwencję notacyjną:**
  - Wszystkie wprowadzone dotychczas oznaczenia ($y_t^P, \hat{y}_t^P, s_t^P, V_t^P$) – tak jak dotychczas – odnoszą się do prognozy zmiennej objaśnianej (tu akurat będącej logarytmem zmiennej ekonomicznej)
  - Natomiast w oznaczeniach odnoszących się już do zmiennej ekonomicznej wprowadzamy dodatkową "gwiazdkę" ($y_t^{*P}, \hat{y}_t^{*P}, s_t^{*P}, V_t^{*P}$)
  - Dla omówionego wyżej wariantu „bez logarytmu" zachodzi oczywiście równość pomiędzy wielkością z „gwiazdką" i tą bez niej

Predykcja w modelu z logarytmiczną zmienną objaśnianą…

➢ Ustalmy zatem uwagę – rozważamy typowy zapis modelu regresji liniowej:

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

- postaci poszczególnych zmiennych objaśniających są tu obojętne – mogą być w logarytmie lub w swej oryginalnej postaci (niektóre lub wszystkie, obojętne)
- zmienna objaśniana y_t jest tu *logarytmem* pewnej zmiennej ekonomicznej – tę ostatnią oznaczmy pomocniczo symbolem y_t^*; zatem:

$$y_t = \ln y_t^*$$

→ Oczywiście, gdyby y_t było bezpośrednio pewną zmienną ekonomiczną (a nie jej logarytmem), wówczas zachodziłoba po prostu równość $y_t = y_t^*$

→ Związek odwrotny, pozwalający „odzyskać” wartość zmiennej ekonomicznej z wartości jej logarytmu:

$$y_t^* = e^{y_t}$$

lub w nieco innym zapisie:

$$y_t^* = \exp(y_t)$$ ← **funkcja w Excelu: =EXP(...)**

→ Zauważmy, że $y_t^* > 0$ – zawsze (w przeciwnym razie nie można by obliczyć $\ln y_t^*$)

→ W praktyce, w konkretnym modelu notacja zwykle jest prostsza, gdyż oznaczenie konkretnej zmiennej ekonomicznej nie wymaga dodatkowego symbolu, którym jest tu „*”. Przykładowo, modelując logarytm ceny (ang. *price*) możemy przyjąć oznaczenie całej takiej zmiennej jako $y_t = \ln P_t$, gdzie P_t (a nie y_t^*) oznacza bezpośrednio zmienną ekonomiczną (cenę).

Predykcja w modelu z logarytmiczną zmienną objaśnianą…

- Dotychczasowe oznaczenia, które w obecnym kontekście odnoszą się do predykcji zmiennej objaśnianej będącej w postaci logarytmicznej (czyli logarytmem pewnej zmiennej ekonomicznej), nie zmieniają się:

  - **prognoza punktowa (y-ka, tj. zmiennej logarytmicznej)**:
  $$\hat{y}_t^P = x_t^P \hat{\beta}$$
  - **średni błąd predykcji *ex ante***:
  $$s_t^P = \sqrt{s^2[1 + x_t^P (X'X)^{-1} x_t^{P\prime}]}$$
  - **względny błąd predykcji *ex ante***:
  $$V_t^P = \frac{s_t^P}{|\hat{y}_t^P|} 100$$
  - **prognoza przedziałowa:**
  $$\left(\hat{y}_t^P - t_{\alpha/2}\, s_t^P;\ \ \hat{y}_t^P + t_{\alpha/2}\, s_t^P\right)$$

Natomiast... (patrz kolejny slajd)

Predykcja w modelu z logarytmiczną zmienną objaśnianą…

- Ich odpowiedniki odnoszące się już bezpośrednio do zmiennej ekonomicznej dane są wzorami:
  - **prognoza punktowa**:

$$\hat{y}_t^{*P} = e^{\hat{y}_t^P + \frac{1}{2}\left(s_t^P\right)^2} \quad \text{lub równoważnie} \quad \hat{y}_t^{*P} = \exp\left(\hat{y}_t^P + \frac{1}{2}(s_t^P)^2\right)$$

→ Zatem NIE wystarcza proste odlogarytmowanie wartości prognozy zmiennej w postaci logarytmicznej: $\hat{y}_t^{*P} \neq e^{\hat{y}_t^P}$. Jak zostało to wyjaśnione na wykładzie, sama wartość $e^{\hat{y}_t^P}$ (bez uwzględnienia składnika $\frac{1}{2}(s_t^P)^2$ w wykładniku) daje, w istocie, prognozę obciążoną *in minus* (tj. niedoszacowaną).

  - **średni błąd predykcji *ex ante*** (wzór tylko aproksymacyjny):

$$s_t^{*P} \approx s_t^P \hat{y}_t^{*P}$$

→ Zatem wystarczy tylko przemnożyć: średni błąd predykcji *ex ante* dla zmiennej w postaci logarytmicznej przez "odlogarytmowaną" prognozę punktową.

  - **względny błąd predykcji *ex ante*** – korzystając z powyższej aproksymacji:

$$V_t^{*P} = \frac{s_t^{*P}}{\hat{y}_t^{*P}} 100 \approx 100 s_t^P$$

→ Zatem V_t^{*P} jest (choć tylko w przybliżeniu) równy średniemu błędowi predykcji *ex ante* dla zmiennej w postaci logarytmicznej (po przeskalowaniu przez 100%).

- **Interpretacje** tych wielkości są takie same, jak te prezentowane wcześniej w Przykładzie (gdzie przedmiotem prognozy była bezpośrednio zmienna ekonomiczna – cena sprzedaży mieszkania)

Predykcja w modelu z logarytmiczną zmienną objaśnianą…

- Natomiast **prognoza przedziałowa** zmiennej ekonomicznej jest tu wyznaczana w prosty sposób: najpierw konstruujemy przedział predykcji (na zadanym poziomie ufności $1-\alpha$) dla zmiennej objaśnianej w postaci logarytmicznej – czyli "standardowo":

$$\left(\hat{y}_t^P - t_{\alpha/2}\, s_t^P;\ \ \hat{y}_t^P + t_{\alpha/2}\, s_t^P\right)$$

a następnie obydwa jego końce poddajemy transformacji poprzez funkcję eksponencjalną (*eksponens*, czyli "*e* do *x*"). Tak więc $(\mathbf{1}-\boldsymbol{\alpha})\cdot\mathbf{100\%}$ **przedział predykcji** (dla "odlogarytmowanej" zmiennej ekonomicznej) możemy zapisać w następujący sposób:

$$\left(\exp(\hat{y}_t^P - t_{\alpha/2}\, s_t^P);\ \ \exp(\hat{y}_t^P + t_{\alpha/2}\, s_t^P)\right)$$

→ Można pokazać, że taki przedział:

- faktycznie kumuluje w sobie masę prawdopodobieństwa $1-\alpha$ (a nie mniej czy więcej), choć...
- **nie** jest już przedziałem najkrótszym z możliwych, spełniających ten wymóg (wyznaczenie takiego najkrótszego przedziału to już nie lada "czelendż" ;)
- **nie** jest już przedziałem symetrycznym względem prognozy punktowej (co łatwo pokazać: średnia arytmetyczna z jego końców **nie** jest równa $\hat{y}_t^{*P}$)

- **Interpretacja** takiego przedziału brzmi tak samo, jak zaprezentowano to we wcześniejszych rozważaniach

Prognozowanie probabilistyczne

Na zakończenie krótka refleksja rozszerzająca nieco nasze spojrzenie na zagadnienie prognozowania. Całość tego materiału została poświęcona predykcji punktowej i przedziałowej. W pierwszym podejściu prognoza ma charakter pojedynczej liczby (z całego "spektrum" wszystkich, mniej lub bardziej, ale w dalszym ciągu możliwych wartości), której towarzyszy ocena stopnia niepewności, ujęta za pomocą błędów predykcji *ex ante*: średniego i względnego. Natomiast w drugim podejściu (w predykcji przedziałowej) otrzymujemy nieco bogatszą informację, w postaci pewnego przedziału liczbowego, ukazującego cały zbiór najbardziej wiarygodnych/prawdopodobnych wartości prognozowanej zmiennej (przy zadanym poziomie ufności). W ciągu ostatniej dekady na znaczeniu zyskało jeszcze szersze – i zarazem na dzień dzisiejszy najpełniejsze – podejście do predykcji ekonometrycznej, a mianowicie tzw. prognozowanie probabilistyczne (ang. *probabilistic forecasting*; *density forecasting*), które polega na wyznaczeniu całego rozkładu wszystkich możliwych wartości prognozy, reprezentowanego za pomocą funkcji gęstości. Istotnie, dysponując całym rozkładem wartości prognozowanej zmiennej (zwykle określonym na całym zbiorze liczb rzeczywistych lub rzeczywistych dodatnich) "widzimy", które wartości charakteryzują się większym, a które mniejszym prawdopodobieństwem, możemy wyznaczać dla niego różne dodatkowe mierniki/wielkości liczbowe charakteryzujące położenie rozkładu (wartość oczekiwana, modalna, mediana), rozrzut wartości, skośność czy "spiczastość" rozkładu, możemy wyznaczać miary pozycyjne (kwantyle) czy też badać liczbę modalnych itp. (bywa bowiem nierzadko, że takie rozkłady są "mniej grzeczne" niż rozkłady normalny czy *t*-Studenta – obydwa charakteryzujące się jednomodalnością i symetrią). Jak widać, informacja zawarta w "całym rozkładzie" prognozy jest znacznie bogatsza, aniżeli ta ujęta tylko w prognozie punktowej czy nawet przedziałowej. Predykcją probabilistyczną, choć obecnie jest ona wiodącym podejściem do prognozowania (nie tylko w nauce, ale przede wszystkim – praktyce), nie będziemy się tu jednak zajmować, ale warto o niej wiedzieć... ;)