

MODELOWANIE SEZONOWOŚCI W KMNRL

Łukasz Kwiatkowski

Katedra Ekonometrii i Badań Operacyjnych

Plan wykładu

- 1) Wprowadzenie
- 2) Terminologia
- 3) Zmienne sezonowe w KMNRL
- 4) Parametryzacje struktury sezonowej w KMNRL
 - a) Parametryzacja bez wyrazu wolnego ← **UWAGA: to zagadnienie proszę pominąć**
 - b) Parametryzacja z wyrazem wolnym
- 5) Znaczenie parametrów struktury sezonowej
- 6) Przykład
- 7) Ocena dopasowania modelu z sezonowością do danych
- 8) Predykcja w modelu ze strukturą sezonową
- 9) Testowanie sezonowości

Wprowadzenie

- W tym materiale dowiemy się, jak uwzględnić w modelu regresji liniowej możliwe **wahania sezonowe zmiennej objaśnianej**. Ponieważ zjawisko sezonowości może dotyczyć, oczywiście, tylko danych w postaci szeregów czasowych (a nie danych przekrojowych), w całej tej prezentacji zakładamy, że modelowaniu podlegają właśnie **dane czasowe**, a tym samym indeks obserwacji, $t = 1, 2, \dots, T$, oznacza numery kolejnych okresów (dni, miesiący, kwartałów itp.), z których pochodzą obserwacje.
- Wiele szeregów czasowych danych społeczno-gospodarczych odznacza się występowaniem **regularnych wahań sezonowych** (ang. *seasonal fluctuations*) – wystarczy wpisać hasło “seasonal data” w oknie przeglądarki... W zakładce “grafika” zobaczą Państwo “całą masę” przykładowych szeregów czasowych (zwykle miesięcznych lub kwartalnych) z wyraźnymi wzorcami sezonowymi (ang. *seasonal patterns*), w tym m.in. dane dotyczące:
 - miesięcznej lub kwartalnej wielkości produkcji / sprzedaży / ceny danego dobra
 - stopy inflacji (liczonej “miesiąc do miesiąca” czy też “kwartał do kwartału”)
 - stopy bezrobocia rejestrowanego
 - PKB (kwartalnego)
 - innych różnych wskaźników makroekonomicznych
 - wielkości ruchu pasażerskiego na lotniskach (i podobnie, dane dotyczące sektora turystycznego)
 - zużycia i ceny energii elektrycznej – te aż z trzema poziomami sezonowości jednocześnie: miesięczną (w obrębie roku), dzienną (w obrębie tygodnia) i godzinową (w obrębie doby)

Terminologia

- Zanim przejdziemy dalej, potrzebujemy usystematyzować terminologię związaną z sezonowością
- W konkretnym zbiorze danych czasowych – o konkretnej częstotliwości obserwacji (godzinowej, dziennej, miesięcznej, kwartalnej) – możemy wyróżnić m **sezonów** (inaczej: **faz**) w obrębie pojedynczego **cyklu** – przykładowo:
 - W przypadku danych o częstotliwości kwartalnej (w skrócie: danych kwartalnych) wyróżniamy $m = 4$ **sezony/fazy** (kwartały), składające się na pojedynczy **cykl** (rok), zatem możemy tu mówić o cyklu rocznym
 - W przypadku danych miesięcznych wyróżniamy $m = 12$ **sezonów/faz** (miesiące) w obrębie pojedynczego **cyklu**, którym jest tu rok (cykl roczny)
 - W przypadku danych dziennych wyróżniamy $m = 7$ **sezonów/faz** (dni) w obrębie pojedynczego **cyklu**, którym jest tu tydzień (cykl tygodniowy)
 - W przypadku danych godzinowych wyróżniamy $m = 24$ **sezony/fazy** (godziny) w obrębie pojedynczego **cyklu**, którym jest tu doba (cykl dobowy)

UWAGA: Zauważmy, że w powyższym zestawieniu **nie** zawarliśmy danych tygodniowych ($m = 52$ w cyklu rocznym). Wynika to zwyczajnie z faktu, że nie spotykamy zjawisk charakteryzujących się sezonowością tygodniową. Owszem jednak, dysponując danymi tygodniowymi w dalszym ciągu możemy badać występowanie w nich sezonowości miesięcznej czy kwartalnej. Jednak takim przypadkiem (niespójność częstotliwości danych z typem sezonowości) nie będziemy się tu zajmować (choć od strony technicznej to nic trudnego :)

Zmienne sezonowe w KMNRL

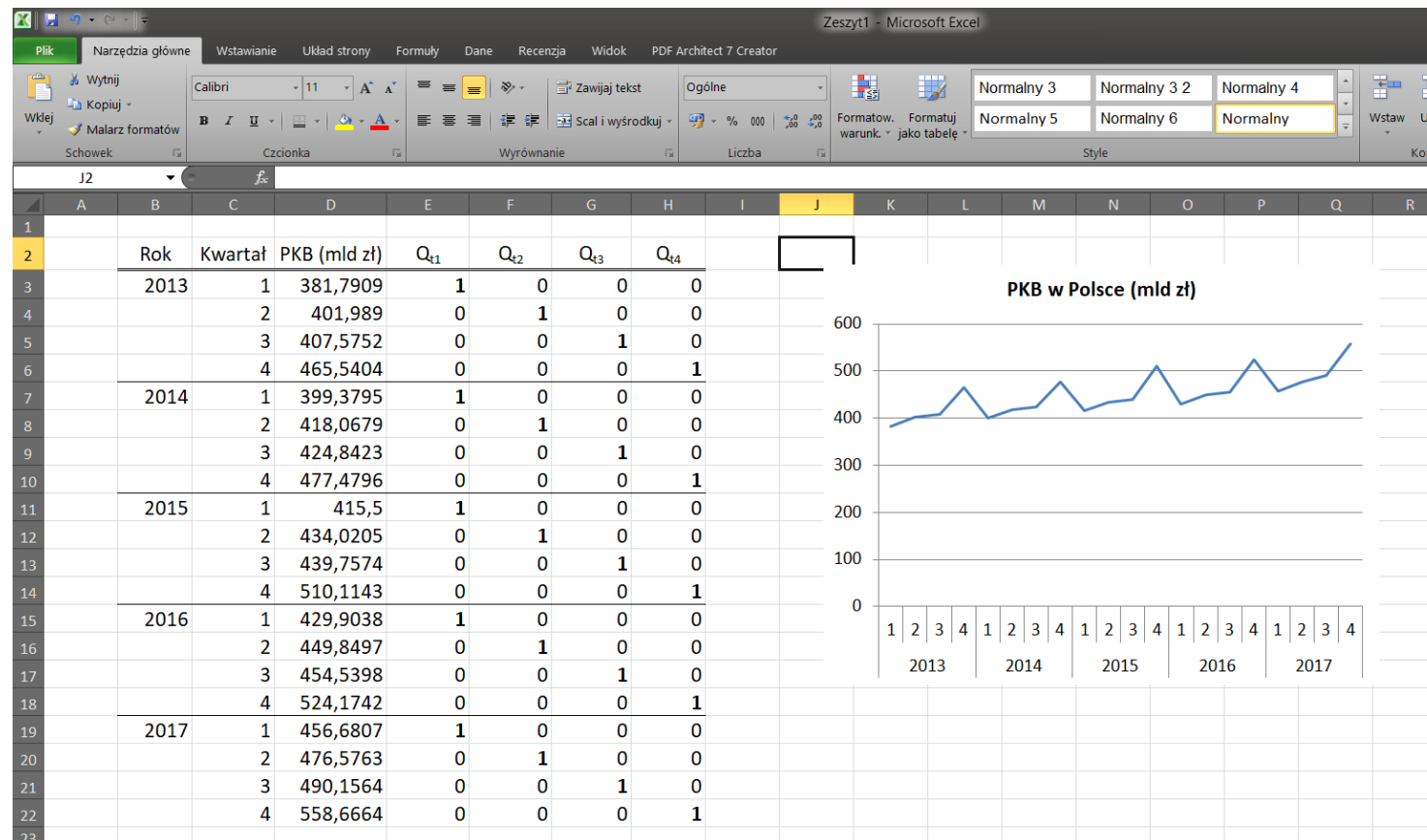
- Do modelowania sezonowości w ramach modelu regresji liniowej potrzebujemy wprowadzić doń tzw. **sztuczne zmienne sezonowe** (ang. *seasonal dummy variables, seasonal dummies*), w skrócie, po prostu, **zmienne sezonowe**.
- Modelując dane sezonowe o częstotliwości m sezonów w obrębie cyklu (czyli np. $m = 4$ dla danych kwartalnych), tworzymy m **zmiennych sezonowych**, przy czym każda z nich jest “zwykłą” zmienną typu 0-1 taką, że wartość “1” przyjmuje tylko dla tych obserwacji, które pochodzą z określonego sezonu (temu, któremu odpowiada dana zmienna), zaś “0” – dla obserwacji pochodzących z wszystkich pozostałych sezonów.
- Zmienne te możemy ogólnie oznaczyć symbolem D_{ti} (od ang. *dummy*), gdzie $t = 1, 2, \dots, T$ – jak dotychczas – jest numerem kolejnych obserwacji, zaś $i = 1, 2, \dots, m$ jest numerem danej zmiennej sezonowej, a zarazem numerem sezonu/fazy cyklu, któremu ta zmienna odpowiada:

$$D_{ti} = \begin{cases} 1 & \text{gdy obserwacja o nr } t \text{ pochodzi z } i - \text{tego sezonu} \\ 0 & \text{w przeciwnym razie} \end{cases}$$

- Dla danych o konkretnej częstotliwości możemy przyjąć bardziej swoiste oznaczenia zmiennych sezonowych:
 - Q_{ti} – dla danych kwartalnych (ang. *quarterly data*), $i = 1, 2, \dots, 4 = m$
 - M_{ti} – dla danych miesięcznych (ang. *monthly data*), $i = 1, 2, \dots, 12 = m$
 - D_{ti} – dla danych dziennych (ang. *daily data*), $i = 1, 2, \dots, 7 = m$
 - H_{ti} – dla danych godzinowych (ang. *hourly data*), $i = 1, 2, \dots, 24 = m$

Zmienne sezonowe w KMNRL

- **Przykład:** analizujemy kwartalną wysokość PKB w Polsce w latach 2013-2017 → dane są kwartalne, zatem tworzymy $m = 4$ zmienne sezonowe: $Q_{t1}, Q_{t2}, Q_{t3}, Q_{t4}$
- **Zauważmy,** że gdyby te cztery kolumny widoczne na poniższym zrzucie ekranu potraktować jako kolumny macierzy X , to w każdym wierszu występuje “1” dokładnie tylko raz, zaś na pozostałych współrzędnych występują “0”. Tak więc dla każdego $t = 1, 2, \dots, T$ zachodzi: $\sum_{i=1}^m D_{ti} = 1$
- **Pytanie:** jeśli każdą z tych kolumn potraktować jako wektor, to do jakiego wektora się one sumują?



Parametryzacja struktury sezonowej w KMNRL

- W praktyce rozważa się **dwa alternatywne (ale w pełni równoważne!)** sposoby uwzględnienia sezonowości zmiennej objaśnianej w modelu regresji, czyli innymi słowy **dwie alternatywne parametryzacje struktury sezonowej**:

- Parametryzacja BEZ wyrazu wolnego ← **UWAGA: to zagadnienie proszę pominąć**
- Parametryzacja Z wyrazem wolnym

- Poniżej omówimy je obydwie. Na początek przypomnijmy sobie ogólny skalarny zapis równania regresji, póki co bez struktury sezonowej:

$$y_t = \underbrace{x_t \beta}_{\text{część systematyczna modelu}} + \varepsilon_t$$

- Strukturę sezonową włączamy – za pomocą zmiennych sezonowych D_{ti} (ogólne oznaczenie) – do *części systematycznej* równania regresji na dwa sposoby. Pierwszym z nich jest... (kolejny slajd)

Parametryzacja struktury sezonowej w KMNRL: wariant BEZ wyrazu wolnego

➤ Parametryzacja BEZ wyrazu wolnego:

$$y_t = \underbrace{\sum_{i=1}^m \theta_i D_{ti}}_{\substack{\text{struktura sezonowa} \\ \text{część systematyczna modelu}}} + \tilde{x}_t \tilde{\beta} + \varepsilon_t$$

— a rozpisawszy wyrażenie z operatorem sumowania:

$$y_t = \theta_1 D_{t1} + \theta_2 D_{t2} + \dots + \theta_m D_{tm} + \tilde{x}_t \tilde{\beta} + \varepsilon_t$$

→ **Wielkości z falką:** $\tilde{x}_t$ i $\tilde{\beta}$ to po prostu pozostałe (tj. niezwiązane z sezonowością) zmienne objaśniające i odpowiadające im parametry (**ALE UWAGA BEZ wyrazu wolnego, czyli BEZ β_0**)

→ **Parametry struktury sezonowej** oznaczamy tu małą literą θ_i (czyt. „teta”), $i = 1, 2, \dots, m$

→ Całą **część systematyczną** modelu możemy bez problemu oznaczyć jako $x_t \beta$:

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

gdzie $x_t = [D_{t1} \ D_{t2} \ \dots \ D_{tm} \ \tilde{x}_t]$, zaś $\beta = [\theta_1 \ \theta_2 \ \dots \ \theta_m \ \tilde{\beta}']'$ (← obie transpozycje konieczne!)

→ Zauważmy, że w tej parametryzacji **uwzględniamy WSZYSTKIE m zmiennych sezonowych** (w naturalnej ich kolejności), ALE za to:

→ **NIE występuje tu wyraz wolny** → **Dlaczego???** (wyjaśniamy na kolejnym slajdzie)

Parametryzacja struktury sezonowej w KMNRL: wariant BEZ wyrazu wolnego

- Zastanówmy się, co by się stało, gdyby do równania z poprzedniego slajdu dołożyć jeszcze wyraz wolny, uzyskując tym samym równanie:

$$y_t = \beta_0 + \theta_1 D_{t1} + \theta_2 D_{t2} + \dots + \theta_m D_{tm} + \tilde{x}_t \tilde{\beta} + \varepsilon_t$$

Dla takiej specyfikacji pierwsza kolumna w macierzy X stanowiłaby, oczywiście, kolumnę jedynek, natomiast kolejne m kolumn — analogicznie do sytuacji przedstawionej na wcześniejszym zrzucie z Excela — składałyby się z zer i jedynek — tak, że suma w każdym wierszu tych m kolumn wynosiłaby dokładnie 1. Zatem te m kolumn z wartościami 0 i 1 zmiennych sezonowych w sumie dają kolumnę jedynek. Jeśli przez $X^{(i)}$ oznaczyć i -tą kolumnę macierzy X dla modelu o powyższym równaniu (z wyrazem wolnym), wówczas powyższe stwierdzenie można zapisać w ten sposób:

$$X^{(1)} = X^{(2)} + X^{(3)} + \dots + X^{(m+1)}$$

Ponieważ suma wektorów jest przykładem ich kombinacji liniowej (o wszystkich współczynnikach równych 1 — przy każdym wektorze po prawej stronie w powyższym zapisie stoi współczynnik =1), okazuje się, że te pierwsze $m+1$ kolumn macierzy X są wektorami liniowo współzależnymi (czyli takimi, że jeden z nich daje się wyrazić jako kombinacja liniowa pozostałych). To z kolei, przeczy założeniu nr 3 Klasycznego Modelu (Normalnej) Regresji Liniowej i powoduje, że macierz X nie jest pełnego rzędu kolumnowego (tj. $\text{rz}(X) < k$, gdzie k to w dalszym ciągu liczba wszystkich kolumn macierzy X). W końcu, jest to równoważne z tym, że wyznacznik macierzy $X'X$ — tej odwracanej we wzorze na $\hat{\beta}$ — wynosi dokładnie 0, przez co nie istnieje $(X'X)^{-1}$, a tym samym nie istnieje $\hat{\beta}$!!!

Parametryzacja struktury sezonowej w KMNRL: wariant Z wyrazem wolnym

- Powyższe rozumowanie skłania ku rozważeniu alternatywnego sposobu uwzględniania struktury sezonowej w równaniu regresji:
- Parametryzacja Z wyrazem wolnym – w tym wariantcie, owszem, zachowujemy wyraz wolny (do którego obecności w dowolnym równaniu regresji możemy być po prostu przywiązani :), ale KOSZTEM usunięcia jednej – DOWOLNEJ – zmiennej sezonowej, przykładowo ostatniej (czyli D_{tm}):

$$y_t = \beta_0 + \beta_1 D_{t1} + \beta_2 D_{t2} + \cdots + \beta_{m-1} D_{t,m-1} + \tilde{x}_t \tilde{\beta} + \varepsilon_t$$

czyli

$$y_t = \beta_0 + \sum_{i=1}^{m-1} \beta_i D_{ti} + \tilde{x}_t \tilde{\beta} + \varepsilon_t$$

- Oznaczenia parametrów struktury sezonowej zostały tu zmienione „z powrotem” na „bety”, gdyż – jak się to później okaże – odpowiadające sobie współczynniki w obydwu parametryzacjach mają RÓŻNE znaczenie, więc nie są dokładnie tymi samymi parametrami, choć owszem, pomiędzy „betami” a „tetami” istnieją konkretne związki funkcyjne, dzięki którym obydwie parametryzacje są, w istocie, równoważne, a mianowicie:

a) $\beta_0 = \theta_m$

b) $\theta_i = \beta_0 + \beta_i$ lub równoważnie $\beta_i = \theta_i - \beta_0$, dla każdego $i = 1, 2, \dots, m-1$

→ Więcej o tych związkach pomiędzy parametrami powiemy sobie za chwilę

Parametryzacja struktury sezonowej w KMNRL: wariant Z wyrazem wolnym

- To, którą ze zmiennych sezonowych postanowimy usunąć w modelu Z wyrazem wolnym, to całkowicie nasza arbitralna i w pełni dowolna decyzja. Aby jednak uniknąć niepotrzebnych wątpliwości czy nieporozumień, w ramach naszych zajęć umawiamy się (przyjmujemy konwencję), że DOMYŚLNIE (czyli jeśli nie narzucono inaczej) zawsze będziemy pomijać ostatnią zmienną sezonową, czyli tę reprezentującą ostatnią fazę cyklu (czyli np. grudzień w przypadku danych miesięcznych, czwarty kwartał w przypadku danych kwartalnych, niedzielę w przypadku danych dziennych).
- Możemy być jednak spokojni, ponieważ usunięcie z modelu (z wyrazem wolnym) jednej ze zmiennych sezonowych NIE sprawia, że odpowiadający jej sezon/faza cyklu „ginie razem z tą zmienną” – jej rolę przejmuje bowiem wyraz wolny, który odpowiada teraz właśnie temu pominiętemu sezonowi. Możemy nieformalnie powiedzieć, że ten „wyrzucony” sezon „siedzi” teraz w wyrazie wolnym. Okaże się to bardziej oczywiste, gdy za chwilę omówimy znaczenie (interpretacje) poszczególnych parametrów w obydwu parametryzacjach.

Parametryzacja struktury sezonowej w KMNRL

➤ **Podsumujmy:** w celu uwzględnienia w modelu regresji struktury sezonowej (służącej wychwyceniu potencjalnych wahań sezonowych zmiennej objaśnianej) potrzebujemy wprowadzić do jego części systematycznej zestaw dodatkowych regresorów, będących pewnymi sztucznymi zmiennymi sezonowymi (typu 0-1). Możemy to zrobić na DWA alternatywne (ale w pełni równoważne) sposoby:

- w parametryzacji BEZ wyrazu wolnego – usuwamy wyraz wolny i wprowadzamy zestaw ~~wszystkich m zmiennych sezonowych (z parametrami struktury sezonowej oznaczonymi symbolem $\theta_i, i = 1, 2, \dots, m$):~~

$$y_t = \theta_1 D_{t1} + \theta_2 D_{t2} + \dots + \theta_m D_{tm} + \tilde{x}_t \tilde{\beta} + \varepsilon_t$$

- w parametryzacji Z wyrazem wolnym – zostawiamy wyraz wolny i wprowadzamy zestaw tylko $m - 1$ (spośród wszystkich m) zmiennych sezonowych – domyślnie, według przyjętej wyżej naszej „zajęciowej” konwencji, wykluczając tę zmienną, która odpowiada ostatniej fazie cyklu (tym sposobem tej fazie odpowiada teraz wyraz wolny); parametry struktury sezonowej oznaczamy symbolem $\beta_i, i = 1, 2, \dots, m - 1$:

$$y_t = \beta_0 + \beta_1 D_{t1} + \beta_2 D_{t2} + \dots + \beta_{m-1} D_{t,m-1} + \tilde{x}_t \tilde{\beta} + \varepsilon_t$$

Znaczenie parametrów struktury sezonowej: model BEZ wyrazu wolnego

➤ O czym informują nas parametry struktury sezonowej (a tym samym ich oceny)?

- W parametryzacji BEZ wyrazu wolnego: $y_t = \theta_1 D_{t1} + \theta_2 D_{t2} + \dots + \theta_m D_{tm} + \tilde{x}_t \tilde{\beta} + \varepsilon_t$

Parametr θ_i dla danej, i -tej fazy / sezonu cyklu ($i \in \{1, 2, \dots, m\}$) informuje nas, o ile (średnio rzecz biorąc) wartość zmiennej objaśnianej w i -tym sezonie jest wyższa ($\theta_i > 0$) / niższa ($\theta_i < 0$) na skutek wahań sezonowych, przy ustalonych wartościach pozostałych zmiennych objaśniających (tych znajdujących się w wektorze $\tilde{x}_t$).

W celu wyjaśnienia rozważmy przykładową sytuację, w której „znajdujemy się” w pierwszym sezonie. Wobec tego $D_{t1} = 1$, zaś $D_{ti} = 0$ dla $i > 1$. Wtedy równanie przyjmuje postać:

$$y_t = \theta_1 + \tilde{x}_t \tilde{\beta} + \varepsilon_t$$

Zaś gdyby nie występowały wahania sezonowe: $y_t = \tilde{x}_t \tilde{\beta} + \varepsilon_t$. Tak więc, ogólnie, θ_i jest tu tą różnicą w poziomie zmiennej y_t , która – przy ustalonych wartościach zmiennych objaśniających – wynika tylko z wahań sezonowych (w i -tym sezonie).

Tak więc, o parametrze θ_i ($i \in \{1, 2, \dots, m\}$) możemy powiedzieć, że określa kierunek i siłę efektu (inaczej: odchylenia) sezonowego w i -tej fazie cyklu.

Znaczenie parametrów struktury sezonowej: model Z wyrazem wolnym

- W parametryzacji Z wyrazem wolnym: $y_t = \beta_0 + \beta_1 D_{t1} + \beta_2 D_{t2} + \dots + \beta_{m-1} D_{t,m-1} + \tilde{x}_t \tilde{\beta} + \varepsilon_t$

W celu wyjaśnienia znaczenia parametrów struktury sezonowej w tej parametryzacji, przypomnijmy związki pomiędzy parametrami tej i tej BEZ wyrazu wolnego:

a) $\beta_0 = \theta_m$

b) $\theta_i = \beta_0 + \beta_i$ lub równoważnie $\beta_i = \theta_i - \beta_0$, dla każdego $i = 1, 2, \dots, m-1$

Ad a) Ze związku $\beta_0 = \theta_m$ wynika, że rolę ostatniego sezonu w cyklu przejmuje teraz wyraz wolny, który ma dokładnie takie samo znaczenie, co θ_m . Zatem:

Parametr wyrazu wolnego, β_0 , informuje nas, o ile (średnio rzecz biorąc) wartość zmiennej objaśnianej w ostatnim sezonie cyklu jest wyższa ($\beta_0 > 0$) / niższa ($\beta_0 < 0$) na skutek wahań sezonowych, przy ustalonych wartościach pozostałych zmiennych objaśniających (tych znajdujących się w wektorze $\tilde{x}_t$)

→ UWAGA: Należy PAMIĘTAĆ, że wyraz wolny zawsze odpowiada tej POMINIĘTEJ fazie spośród zmiennych sezonowych – tylko(!) domyślnie ostatniemu. Jeśli zatem w konstrukcji modelu pominięty byłby INNY niż ostatni sezon, wówczas należy **wziąć to pod uwagę** w formułowaniu interpretacji oceny parametru.

Znaczenie parametrów struktury sezonowej: model Z wyrazem wolnym – c.d.

- W parametryzacji Z wyrazem wolnym: $y_t = \beta_0 + \beta_1 D_{t1} + \beta_2 D_{t2} + \dots + \beta_{m-1} D_{t,m-1} + \tilde{x}_t \tilde{\beta} + \varepsilon_t$

Ad b) Natomiast ze związku $\beta_i = \theta_i - \beta_0$ wynika, że β_i jest **kontrastem** (różnicą) pomiędzy **efektem sezonowym** (inaczej: **odchyleniem sezonowym**) w danym, i -tym sezonie a **efektem sezonowym** w ostatniej fazie cyklu (której to odpowiada β_0). Zatem:

Parametr β_i ($i \in \{1, 2, \dots, m-1\}$) informuje nas, o ile (średnio rzecz biorąc) wartość **efektu** (in.: odchylenia) **sezonowego** w i -tej fazie cyklu jest wyższa ($\beta_i > 0$) / niższa ($\beta_i < 0$) w stosunku do **efektu** (in.: odchylenia) **sezonowego** w ostatniej fazie cyklu, przy ustalonych wartościach pozostałych zmiennych objaśniających (tych znajdujących się w wektorze $\tilde{x}_t$)

Równoważnie i w sposób zbliżony do interpretacji w modelu BEZ wyrazu wolnego:

Parametr β_i ($i \in \{1, 2, \dots, m-1\}$) informuje nas, o ile (średnio rzecz biorąc) wartość zmiennej objaśnianej w i -tym sezonie jest wyższa ($\beta_i > 0$) / niższa ($\beta_i < 0$) na skutek wahań sezonowych, **w stosunku do ostatniej fazy cyklu**, przy ustalonych wartościach pozostałych zmiennych objaśniających (tych znajdujących się w wektorze $\tilde{x}_t$).

→ UWAGA: Należy PAMIĘTAĆ, że wyraz wolny zawsze odpowiada tej POMINIĘTEJ fazie spośród zmiennych sezonowych – tylko(!) domyślnie ostatniemu. Jeśli zatem w konstrukcji modelu pominięty byłby INNY niż ostatni sezon, wówczas należy **wziąć to pod uwagę** w formułowaniu interpretacji oceny parametru.

Znaczenie parametrów struktury sezonowej: logarytmiczna zmienna objaśniana

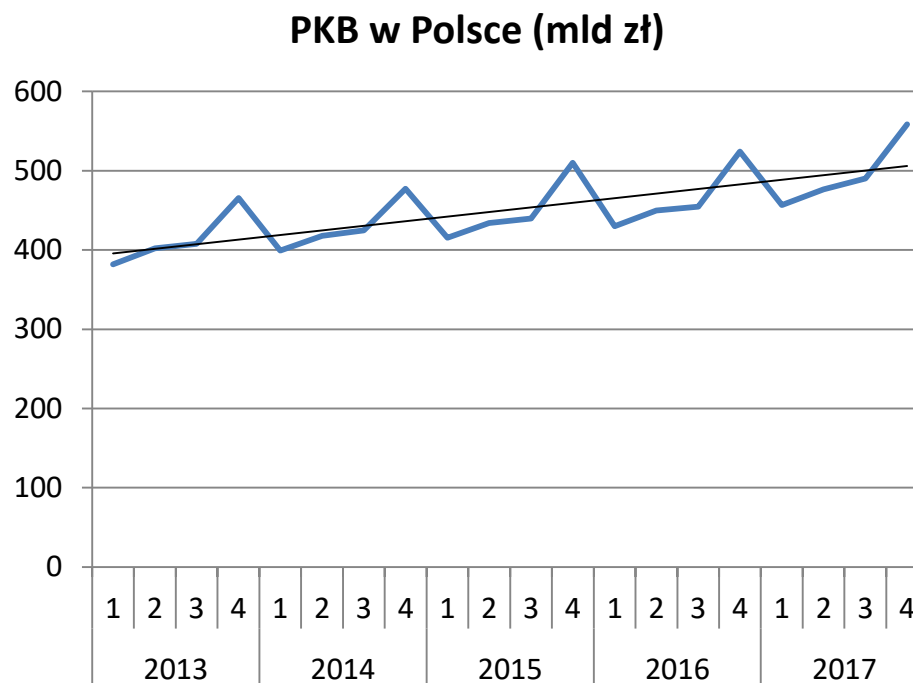
- Jeżeli zmienną objaśnianą jest w modelu bezpośrednio pewna zmienna ekonomiczna, wówczas we wszystkich przedstawionych wyżej interpretacjach ma zastosowanie jednostka tejże zmiennej.
- Jeśli jednak model zbudowano dla **logarytmu** pewnej zmiennej ekonomicznej (zatem $y_t = \ln y_t^*$, gdzie y_t^* oznacza zmienną ekonomiczną), wówczas na potrzeby interpretacji oceny danego parametru należy zastosować jego znaną nam już **transformację**:

$$(e^{ocena\ parametru} - 1) \cdot 100\%$$

co oznacza zarazem, że w interpretacjach, w miejsce jednostki zmiennej ekonomicznej używamy **procentów**.

Przykład

- Kontynuujemy zarysowany wcześniej przykład kwartalnej wysokości PKB w Polsce w latach 2013-2017



- Łatwo zauważyć, że oprócz wyraźnego wzorca sezonowości, w modelowanym szeregu czasowych zarysowuje się wyraźny **trend liniowy** – weźmiemy go zatem także pod uwagę w specyfikacji modelu:

- ~~Model BEZ wyrazu wolnego:~~ $y_t = \theta_1 Q_{t1} + \theta_2 Q_{t2} + \theta_3 Q_{t3} + \theta_4 Q_{t4} + \beta_1 t + \varepsilon_t$
- Model Z wyrazem wolnym: $y_t = \beta_0 + \beta_1 Q_{t1} + \beta_2 Q_{t2} + \beta_3 Q_{t3} + \beta_4 t + \varepsilon_t$

Przykład – wtręt terminologiczny

- **UWAGA!!!** W mowie potocznej istnieje możliwość błędnego zrozumienia tego, że zmienną czasową t – jako numer kolejnych, tu kwartalnych obserwacji – możemy określić jako „numer kwartału”, z którego pochodzi dana obserwacja. Ponieważ – faktycznie – rozważane tu dane są kwartalne, więc t jako numer kolejnych okresów, z których pochodzą obserwacje, jest *de facto* numerem kwartału. Oczywiście, zmienna czasowa t przyjmuje wartości naturalne od 1 do T , czyli do (dowolnie dużej) liczby wszystkich modelowanych obserwacji, co jak najbardziej kłóci się (i słusznie!) ze zwyczajowym (i jak najbardziej prawidłowym) rozumieniem wyrażenia „numer kwartału”, pod którym rozumiemy po prostu liczbę od 1 do 4 (a nie do jakiegoś dowolnie dużego T), określającą, owszem, numer kwartału, ale w rozumieniu sezonu (któregoś z czterech) w obrębie rocznego cyklu.
- Nie mniej jednak, należy **PAMIĘTAĆ**, że na poziomie modelu regresji, zmienna czasowa t ZAWSZE przyjmuje kolejne wartości liczb naturalnych od 1 do liczby wszystkich obserwacji, T .
- Zatem w przypadku rozważanych w tym przykładzie danych kwartalnych, t – jako numer okresu – przyjmuje (dokładnie tak samo, jak to robiliśmy dotychczas) **wartości od 1 do T** (tu akurat $T = 20 = 5$ lat $\times$ 4 kwartały), a **NIE** powtarzający się dla każdego kolejnego roku (2013-2017) ciąg czterech wartości od 1 do 4.
- Prawidłowa postać macierzy X w obydwu parametryzacjach – z poprawnie wprowadzonymi wartościami zmiennej t – została przedstawiona na kolejnym slajdzie

Przykład

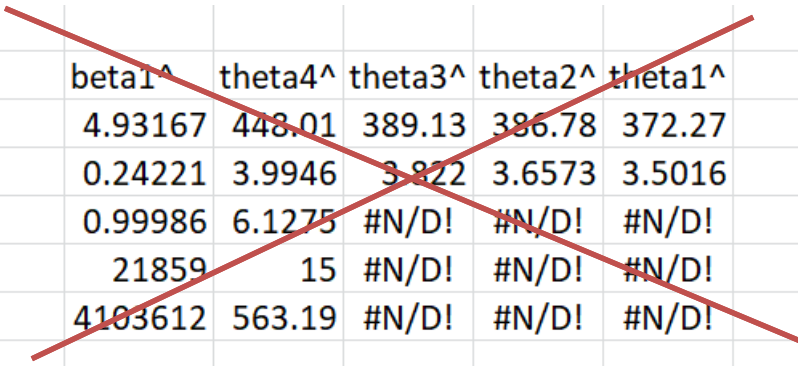
➤ Zbudujemy kolumnę y oraz macierze X w obydwu modelach:

C4		f _{sc}																					
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1																							
2									Model BEZ wyrazu wolnego							Model Z wyrazem wolnym							
3									$y_t = \theta_1 Q_{t1} + \theta_2 Q_{t2} + \theta_3 Q_{t3} + \theta_4 Q_{t4} + \beta_1 t + \varepsilon_t$							$y_t = \beta_0 + \beta_1 Q_{t1} + \beta_2 Q_{t2} + \beta_3 Q_{t3} + \beta_4 t + \varepsilon_t$							
4																							
5																							
6									X							X							
7	Rok	Kwartał	PKB (mld zł)	Q _{t1}	Q _{t2}	Q _{t3}	Q _{t4}		y	Q _{t1}	Q _{t2}	Q _{t3}	Q _{t4}	t		y	"1"	Q _{t1}	Q _{t2}	Q _{t3}	t		
8	2013	1	381.7909	1	0	0	0		381.79	1	0	0	0	1		381.79	1	1	0	0	1		
9		2	401.989	0	1	0	0		401.99	0	1	0	0	2		401.99	1	0	1	0	2		
10		3	407.5752	0	0	1	0		407.58	0	0	1	0	3		407.58	1	0	0	1	3		
11		4	465.5404	0	0	0	1		465.54	0	0	0	1	4		465.54	1	0	0	0	4		
12	2014	1	399.3795	1	0	0	0		399.38	1	0	0	0	5		399.38	1	1	0	0	5		
13		2	418.0679	0	1	0	0		418.07	0	1	0	0	6		418.07	1	0	1	0	6		
14		3	424.8423	0	0	1	0		424.84	0	0	1	0	7		424.84	1	0	0	1	7		
15		4	477.4796	0	0	0	1		477.48	0	0	0	1	8		477.48	1	0	0	0	8		
16	2015	1	415.5	1	0	0	0		415.5	1	0	0	0	9		415.5	1	1	0	0	9		
17		2	434.0205	0	1	0	0		434.02	0	1	0	0	10		434.02	1	0	1	0	10		
18		3	439.7574	0	0	1	0		439.76	0	0	1	0	11		439.76	1	0	0	1	11		
19		4	510.1143	0	0	0	1		510.11	0	0	0	1	12		510.11	1	0	0	0	12		
20	2016	1	429.9038	1	0	0	0		429.9	1	0	0	0	13		429.9	1	1	0	0	13		
21		2	449.8497	0	1	0	0		449.85	0	1	0	0	14		449.85	1	0	1	0	14		
22		3	454.5398	0	0	1	0		454.54	0	0	1	0	15		454.54	1	0	0	1	15		
23		4	524.1742	0	0	0	1		524.17	0	0	0	1	16		524.17	1	0	0	0	16		
24	2017	1	456.6807	1	0	0	0		456.68	1	0	0	0	17		456.68	1	1	0	0	17		
25		2	476.5763	0	1	0	0		476.58	0	1	0	0	18		476.58	1	0	1	0	18		
26		3	490.1564	0	0	1	0		490.16	0	0	1	0	19		490.16	1	0	0	1	19		
27		4	558.6664	0	0	0	1		558.67	0	0	0	1	20		558.67	1	0	0	0	20		
28																							
29																							

Przykład

➤ Obydwa modele szacujemy za pomocą funkcji REGLINP:

- ~~W parametryzacji BEZ wyrazu wolnego zaznaczamy CAŁĄ macierz X (kolumny „1” i tak w niej nie ma), natomiast w formule REGLINP, w miejsce trzeciego jej argumentu („stała”) wprowadzamy wartość „0” (a nie „1”, z przyzwyczajenia), ponieważ w tym modelu NIE występuje wyraz wolny~~



beta1^	theta4^	theta3^	theta2^	theta1^
4.93167	448.01	389.13	386.78	372.27
0.24221	3.9946	3.822	3.6573	3.5016
0.99986	6.1275	#N/D!	#N/D!	#N/D!
21859	15	#N/D!	#N/D!	#N/D!
4103612	563.19	#N/D!	#N/D!	#N/D!

- W parametryzacji Z wyrazem wolnym postępujemy tak, jak dotychczas: zaznaczamy macierz X , ale BEZ kolumny „1”, natomiast w miejsce trzeciego jej argumentu („stała”) wprowadzamy wartość „1”, ponieważ w tym modelu, owszem, wyraz wolny występuje

beta4^	beta3^	beta2^	beta1^	beta0^
4.9317	-58.89	-61.23	-75.75	448.01
0.2422	3.8829	3.9055	3.9429	3.9946
0.9856	6.1275	#N/D!	#N/D!	#N/D!
257.38	15	#N/D!	#N/D!	#N/D!
38654	563.19	#N/D!	#N/D!	#N/D!

Przykład

➤ Możemy teraz zapisać oszacowane równania obydwu modeli:

- ~~Model BEZ wyrazu wolnego:~~

$$\hat{y}_t = 372,27Q_{t1} + 386,78Q_{t2} + 389,13Q_{t3} + 448,01Q_{t4} + 4,93t$$

- Model Z wyrazem wolnym:

$$\hat{y}_t = 448,01 - 75,75Q_{t1} - 61,23Q_{t2} - 58,89Q_{t3} + 4,93t$$

➤ ~~Zauważmy, że wartość oceny współczynnika przy zmiennej czasowej jest dokładnie taka sama w obydwu parametryzacjach. Wynika to z faktu – powtórzmy – że tak naprawdę obydwie parametryzacje (modele) są sobie w pełni równoważne. Jeśli już o tej równoważności mowa, to zauważmy, że pomiędzy ocenami parametrów struktur sezonowych w obydwu specyfikacjach zachodzą omówione wcześniej związki:~~

a) ~~$\hat{\beta}_0 = 448,01 = \hat{\theta}_4$ (θ_4 stoi przy Q_{t4} w pierwszym zapisie)~~

b) ~~$\hat{\theta}_1 = 372,27 = \hat{\beta}_0 + \hat{\beta}_1 = 448,01 + (-75,75)$ (z dokładnością do zaokrągleń),~~

~~podobnie dla $\hat{\theta}_2$ i $\hat{\theta}_3$~~

Przykład

➤ Interpretacje ocen parametrów struktury sezonowej

- **Model BEZ wyrazu wolnego**: $\hat{y}_t = 372,27Q_{t1} + 386,78Q_{t2} + 389,13Q_{t3} + 448,01Q_{t4} + 4,93t$

~~$\hat{\theta}_1 = 372,27$: Rokrocznie, na skutek wahań sezonowych, wartość PKB w Polsce w I kwartale jest wyższa ($\hat{\theta}_1 > 0$) przeciętnie o 372,27 mld zł. ($\leftarrow y_t$ jest wyrażone właśnie w mld zł)~~

- ~~○ Rokrocznie — podkreśla cykliczność, powtarzalność odnotowywanych wahań sezonowych~~
- ~~○ Oczywiście, jeśli cyklem modelowanych danych byłby tydzień (w przypadku częstotliwości dziennej), wówczas należałoby napisać: W każdym tygodniu~~
- ~~○ Gdyby w modelu występowały jeszcze inne zmienne objaśniające, wówczas należałoby jeszcze wpisać gdzieś w powyższej interpretacji tę „klauzulę”: „przy ustalonych wartościach pozostałych zmiennych objaśniających” — oczywiście, odpowiednio adaptując to sformułowanie do znaczenia tych pozostałych zmiennych~~

~~$\hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4$: — analogicznie~~

$\hat{\beta}_1 = 4,93$: [prawie tak samo, jak dotychczas, czyli:] Wyłączając wahania sezonowe, z kwartału na kwartał wartość PKB w Polsce wzrasta przeciętnie o 4,93 mld zł.

→ Zwróćmy uwagę na tę dodatkową „klauzulę” o pominięciu wahań sezonowych

Przykład

➤ Interpretacje ocen parametrów struktury sezonowej – c.d.

- Model Z wyrazem wolnym: $\hat{y}_t = 448,01 - 75,75Q_{t1} - 61,23Q_{t2} - 58,89Q_{t3} + 4,93t$

~~$\hat{\beta}_0 = 448,01$: Rokrocznie, na skutek wahań sezonowych, wartość PKB w Polsce w IV kwartale jest wyższa
($\hat{\beta}_0 > 0$) przeciętnie o 448,01 mld zł~~

—————> Interpretacja taka sama, jak dla $\hat{\theta}_4$ w modelu BEZ wyrazu wolnego

$\hat{\beta}_1 = -75,75$: Rokrocznie, efekt sezonowy (inaczej: odchylenie sezonowe) wartości PKB w Polsce
w I kwartale jest niższy ($\hat{\beta}_1 < 0$) w stosunku do efektu sezonowego w IV kwartale przeciętnie
o 75,75 mld zł

Równoważnie: Rokrocznie, na skutek wahań sezonowych, wartość PKB w Polsce w I kwartale jest niższa w
stosunku do IV kwartału przeciętnie o 75,75 mld zł.

$\hat{\beta}_2, \hat{\beta}_3$: analogicznie

$\hat{\beta}_4 = 4,93$: [tak samo jak w modelu bez wyrazu wolnego] Wyłączając wahania sezonowe, z kwartału
na kwartał wartość PKB w Polsce wzrasta przeciętnie o 4,93 mld zł.

→ Zwróćmy uwagę na tę dodatkową „klauzulę” o pominięciu wahań sezonowych

Ocena dopasowania modelu z sezonowością do danych

→ Rozważania tylko dla tych, którzy jednak dociekają także tej parametryzacji BEZ wyrazu wolnego ;P

- Do oceny dobroci dopasowania do danych modelu uwzględniającego strukturę sezonową wykorzystujemy dokładnie te same mierniki, które poznaliśmy dotychczas
 - Warto natomiast podkreślić, że w związku z równoważnością obydwu rozważanych parametryzacji struktury sezonowości, uzyskane w nich **wartości odpowiadających sobie mierników dokładnie się pokrywają.**
 - W takim razie jak wytłumaczyć tę widoczną w uzyskanych powyżej wynikach funkcji REGLINP różnicę w wartościach współczynnika determinacji???? → zob. kilka slajdów wcześniej:
 - w modelu BEZ wyrazu wolnego: $R^2 = 0,99986$
 - w modelu Z wyrazem wolnym: $R^2 = 0,9856$
- Sam też musiałem trochę poszukać i domyślić się kilku rzeczy, żeby uświadomić sobie, że gdy używamy funkcji REGLINP do estymacji modelu regresji BEZ wyrazu wolnego, wówczas, niestety, funkcja ta zwraca niepoprawną wartość współczynnika determinacji (oblicza go trochę inaczej niż powinien, ale – bez wchodzenia w szczegółowe wyjaśnienia – błędnie!!!)
- W związku z powyższym, korzystając z wyników generowanych przez funkcję REGLINP w ocenie dobroci dopasowania modelu ze strukturą sezonową do danych należy PAMIĘTAĆ, żeby skorzystać z wyników otrzymanych dla parametryzacji **z wyrazem wolnym**. Ewentualnie, korzystając z tej drugiej, bez wyrazu wolnego, możemy po prostu obliczyć R^2 „na piechotę” ($\hat{y} \rightarrow \hat{\varepsilon} \rightarrow SSE \rightarrow \varphi^2 \rightarrow R^2 = 1 - \varphi^2$)