

Maciej Oesterreich

WYKORZYSTANIE PROGRAMU R W PROGNOZOWANIU NA PODSTAWIE MODELI PRZYCZYNOWO-OPISOWYCH W WARUNKACH BRAKU PEŁNEJ INFORMACJI

APPLICATION OF R-ENVIRONMENT AND DESCRIPTIVE MODELS IN FORECASTING IN CONDITIONS OF THE LACK OF FULL INFORMATION

Katedra Zastosowań Matematyki w Ekonomii, Zachodniopomorski Uniwersytet Technologiczny w Szczecinie, ul. Klemensa Janickiego 31, 71-270 Szczecin, moesterreich@zut.edu.pl

Summary. This paper presents application of procedure in R-environment to forecasting missing data in seasonal time series. In forecasts construction was used descriptive model. The six variants of systematic gaps in the data were analyzed. The goal was to choose the best predictors for construction of *ex ante* forecasts.

Słowa kluczowe: luki systematyczne, prognozowanie brakujących danych, szeregi czasowe.
Key words: electronic reporting, reporting, taxes.

WSTĘP

Jednym z podstawowych warunków modelowania, a następnie prognozowania w klasycznej ekonometrii jest kompletność danych. Niespełnienie tego warunku w wielu przypadkach nie musi być jednoznaczne z niemożnością przeprowadzenia analiz ekonometrycznych. Od pewnego czasu w literaturze przedmiotu pojawiają się propozycje metod, których celem jest nie tylko „odtworzenie” brakujących informacji w szeregu czasowym, ale także budowa prognoz ekstrapolacyjnych. Stosowanie niektórych metod jest związane ze spełnieniem specyficznych warunków dotyczących: dostępności określonej liczby wyrazów początkowych (modele adaptacyjne), eliminacji sezonowości (metody numeryczne) lub występowania tylko luk niesystematycznych. Tego rodzaju ograniczenia nie występują w modelowaniu i prognozowaniu na podstawie równań przyczynowo-opisowych z parametrami zmieniającymi się sezonowo, w których sezonowość jest opisana za pomocą wielomianu trygonometrycznego.

Model przyczynowo-opisowy ze stałymi parametrami przy zmiennej objaśniającej oraz zmieniającymi się sezonowo parametrami opisującymi periodyczne wahania sezonowe za pomocą wielomianu trygonometrycznego przedstawiono poniżej:

$$Y_t = \beta_0 + \beta_1 t + \sum_{j=2}^p \beta_j X_{jt} + \sum_{i=1}^{\frac{m}{2}} (a_{ji} \sin \omega_i t + b_{ji} \cos \omega_i t)$$

gdzie:

$$\omega = \frac{2\pi \cdot i}{m}$$

W modelach tych występuje także zmienna czasowa t . Szacowanie parametrów równania na podstawie zarówno kompletnych, jak i niekompletnych danych nie wymaga spełnienia warunku sumowalności do zera. W przypadku wystąpienia luk systematycznych mogą się pojawić problemy wpływające na komplikację procesu modelowania. Należą do nich (Zastosowanie hierarchicznych modeli szeregów czasowych... 2003):

1. Stałość niektórych składowych.
2. Bardzo silne skorelowanie składowych harmonicznych przyjmujące niekiedy postać dokładnej współliniowości.

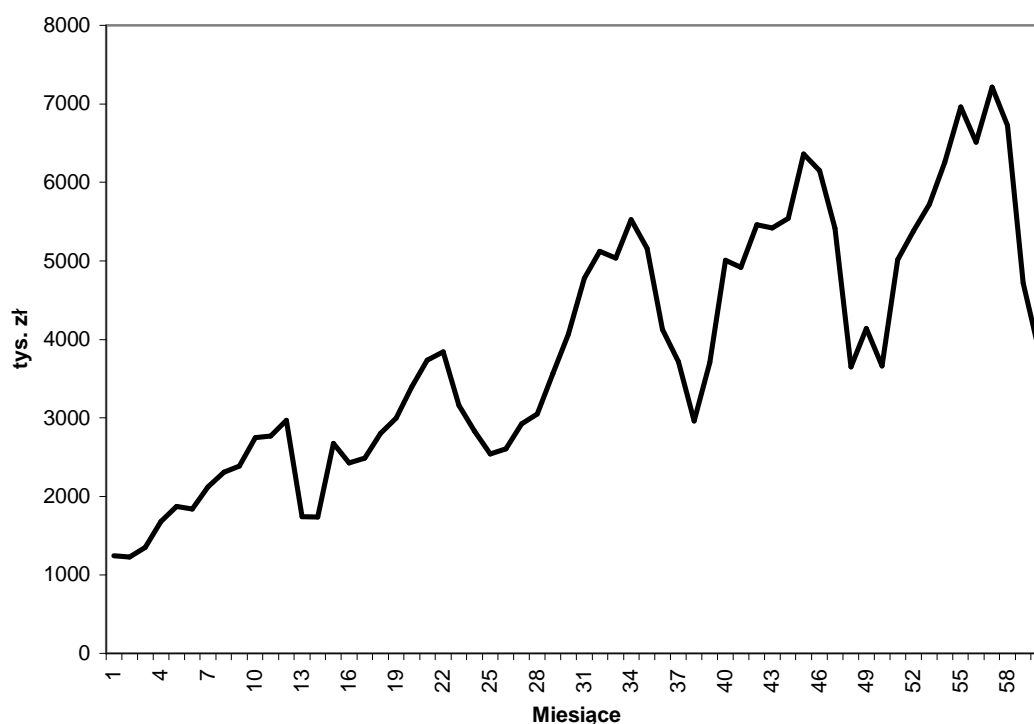
Występowanie następstwa wymienionego w punkcie 1 będzie jednoznaczne z pominięciem składowych o zerowych wariancjach i współliniowych. Natomiast w przypadku drugim konieczne będzie szacowanie wielu wersji modeli o zestawach składowych niebędących kombinacjami liniowymi innych składowych. Z punktu widzenia własności predyktywnych wersje te będą nierozróżnialne, ponieważ charakteryzować się będą identycznymi ocenami parametrów struktury stochastycznej. Natomiast będą się one różnić dokładnością prognoz inter- i ekstrapolacyjnych. Oznacza to, że dla celów prognozowania *ex-ante* należy wybierać te wersje, które będą się charakteryzować minimalnymi ocenami błędów prognoz *ex-post*. W przypadku danych miesięcznych liczba wszystkich możliwych kombinacji spośród 13 zmiennych objaśniających, składających się z 11 składowych harmonicznych oraz zmiennej czasowej t i jednej zmiennej objaśniającej ($X_{1,t}$), równa jest $2^{13}-1 = 8191$. Dlatego też do wykonania obliczeń w niniejszej pracy skonstruowano program dla środowiska R, który w sposób „automatyczny” dokonuje wyszukiwania modeli zawierających najwięcej zmiennych niezależnych o niezerowych wariancjach, niewspółliniowych oraz niebędących kombinacjami liniowymi innych zmiennych, znacznie skracając ich czas.

Algorytm składa się z następujących etapów:

1. Wygenerowanie macierzy wszystkich możliwych kombinacji zmiennych objaśniających.
2. Wyeliminowanie tych kombinacji, w których występują zmienne objaśniające z zerową wariancją.
3. Proces modelowania. Podczas tego etapu są usuwane takie zestawy kombinacji, w których wykryto kombinacje liniowe oraz współliniowość składowych. Dla pozostałych wersji szacuje się parametry i wyznacza się prognozy interpolacyjne i ekstrapolacyjne.
4. Procedurę kończy wylistowanie kombinacji o maksymalnej liczbie zmiennych objaśniających, uwzględniających dodatkowe kryteria, takie jak poziom współczynnika determinacji, poziom *AIC*, czy poziom statystyki *F* (Anderson i in. 2011) dla ostatniej zmiennej dodanej do modelu. Obok wyżej wymienionych charakterystyk dodatkowo są obliczane średnie względne błędy prognoz interpolacyjnych, a także ekstrapolacyjnych dla horyzontu wynoszącego 3, 6, 9 i 12 miesięcy.

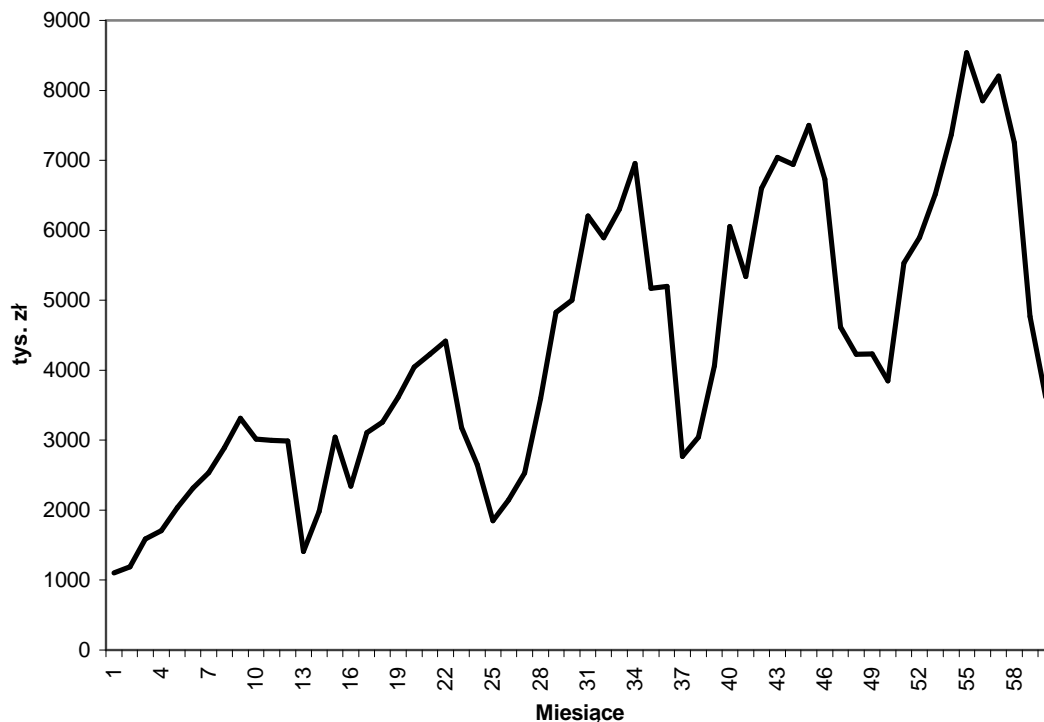
PRZYKŁAD EMPIRYCZNY

Wykorzystanie wyżej wymienionej procedury zostanie przedstawione na przykładzie modelowania i prognozowania kosztów całkowitych produkcji materiałów budowlanych w zależności od wielkości sprzedaży w pewnym przedsiębiorstwie w okresie pięciu kolejnych lat. Na podstawie dostępnych danych statystycznych, pochodzących z pierwszych czterech lat, za pomocą proponowanej procedury zostanie przeprowadzony wybór kombinacji składowych, a następnie proces prognozowania inter- i ekstrapolacyjnego. Zostanie także przeprowadzona analiza dokładności obu rodzajów prognoz. W jej wyniku zostaną wybrane modele dla celów prognozowania *ex-ante*. W procesie analizy dokładności prognoz ekstrapolacyjnych zostaną wykorzystane dane z piątego roku. Kształtowanie się zmiennej prognozowanej przedstawiono na rys. 1.



Rys. 1. Kształtowanie się całkowitych kosztów produkcji betonu komórkowego w pewnym przedsiębiorstwie produkcyjnym

Jak wspomniano wcześniej, zmienną objaśniającą była sprzedaż materiałów budowlanych w tym samym przedsiębiorstwie i w analogicznym okresie. Jej kształtowanie przedstawiono na rys. 2.



Rys. 2. Kształtowanie się całkowitej wartości sprzedaży materiałów budowlanych w pewnym przedsiębiorstwie produkcyjnym

Rozpatrywanych będzie sześć przedstawionych w tabeli 1 wariantów systematycznych luk w danych.

Tabela 1. Warianty luk systematycznych

| | | | | | | | | | | | | | |
|-------------|------|---|----|-----|----|---|----|-----|------|----|---|----|-----|
| Wariant I | Lata | I | II | III | IV | V | VI | VII | VIII | IX | X | XI | XII |
| | I | X | | X | | X | | X | | X | | X | |
| | II | X | | X | | X | | X | | X | | X | |
| | III | X | | X | | X | | X | | X | | X | |
| | IV | X | | X | | X | | X | | X | | X | |
| Wariant II | Lata | I | II | III | IV | V | VI | VII | VIII | IX | X | XI | XII |
| | I | | X | | X | | X | | X | | X | | X |
| | II | | X | | X | | X | | X | | X | | X |
| | III | | X | | X | | X | | X | | X | | X |
| | IV | | X | | X | | X | | X | | X | | X |
| Wariant III | Lata | I | II | III | IV | V | VI | VII | VIII | IX | X | XI | XII |
| | I | X | X | | | X | X | | | X | X | | |
| | II | X | X | | | X | X | | | X | X | | |
| | III | X | X | | | X | X | | | X | X | | |
| | IV | X | X | | | X | X | | | X | X | | |

Tabela 1 (cd.)

| | | | | | | | | | | | | | |
|------------|------|---|----|-----|----|---|----|-----|------|----|---|----|-----|
| Wariant IV | Lata | I | II | III | IV | V | VI | VII | VIII | IX | X | XI | XII |
| | I | | | X | X | | | X | X | | | X | X |
| | II | | | X | X | | | X | X | | | X | X |
| | III | | | X | X | | | X | X | | | X | X |
| | IV | | | X | X | | | X | X | | | X | X |
| Wariant V | Lata | I | II | III | IV | V | VI | VII | VIII | IX | X | XI | XII |
| | I | X | X | X | | | | X | X | X | | | |
| | II | X | X | X | | | | X | X | X | | | |
| | III | X | X | X | | | | X | X | X | | | |
| | IV | X | X | X | | | | X | X | X | | | |
| Wariant VI | Lata | I | II | III | IV | V | VI | VII | VIII | IX | X | XI | XII |
| | I | | | | X | X | X | | | | X | X | X |
| | II | | | | X | X | X | | | | X | X | X |
| | III | | | | X | X | X | | | | X | X | X |
| | IV | | | | X | X | X | | | | X | X | X |

X – dane.

W tabeli 2 zestawiono wyniki modelowania dla sześciu wcześniej wymienionych wariantów luk w danych. Rezultatem zastosowania programu R jest otrzymanie 2–11 wersji zawierających maksymalną dla poszczególnych wariantów liczbę składowych. Dla każdego wariantu i każdej wersji podano oznaczone cyfrą jeden te składowe harmoniczne, które weszły do modelu. Zostały one oznaczone symbolem s_i i c_i . Symbol s_i oznacza $\sin \omega_i t$, a c_i – $\cos \omega_i t$. Dodatkowo w kolumnie t podano informacje o występowaniu lub też braku zmiennej czasowej. W kolejnej kolumnie została podana liczba składowych harmonicznych.

W tabeli tej zestawiono także oceny: kryterium informacyjnego Akaike (AIC), współczynników determinacji (R^2), poziomu odchyień standardowych składnika losowego (S_e), współczynników zmienności losowej (V_s) oraz statystyk F^* odpowiadających składowym, które jako ostatnie zostały wprowadzone do modelu.

Z informacji dotyczących kształtowania ocen wyżej wymienionych parametrów wynika, że są one identyczne w wariantach luk dla poszczególnych kombinacji składowych. Zatem są one nierozróżnialne z punktu widzenia własności predyktywnych. Liczby wersji w poszczególnych wariantach zawierały się od dwóch w drugim wariantie do dziewięciu w czwartym.

W pierwszym wariantie otrzymano osiem wersji zawierających składowe. Spełniały one warunek dotyczący minimalnego poziomu statystyki F ostatniej zmiennej wprowadzonej do modelu. Przeciętna różnica między wartościami rzeczywistymi a teoretycznymi wyniosła 226,64 tys. zł, a udział odchylenia standardowego składnika losowego w średnim poziomie zmiennej prognozowanej wyniósł 6,57%.

W drugim wariantie stwierdzono występowanie tylko dwóch dwuelementowych kombinacji składowych harmonicznych. Charakteryzowały się one nieznacznie wyższym poziomem zarówno S_e , jak i V_s w porównaniu z modelem z pierwszego wariantu.

Tabela 2. Zestawienie kombinacji składowych harmonicznych oraz informacji o własnościach modeli na nich zbudowanych

| Wariant | Lp. modelu | t | s_1 | s_2 | s_3 | s_4 | s_5 | c_1 | c_2 | c_3 | c_4 | c_5 | c_6 | Liczba składowych | AIC | R^2 | S_e | V_s (%) | F^* | |
|---------|------------|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------------------|---------------|---------------|---------------|---------------|-------------|-------------|
| I | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 4 | 335,10 | 0,98 | 224,64 | 6,57 | 1,71 | |
| | 2 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 4 | 335,10 | 0,98 | 224,64 | 6,57 | 1,71 | |
| | 3 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 4 | 335,10 | 0,98 | 224,64 | 6,57 | 1,71 | |
| | 4 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 4 | 335,10 | 0,98 | 224,64 | 6,57 | 1,71 | |
| | 5 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 4 | 335,10 | 0,98 | 224,64 | 6,57 | 1,71 | |
| | 6 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 4 | 335,10 | 0,98 | 224,64 | 6,57 | 1,71 | |
| | 7 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 4 | 335,10 | 0,98 | 224,64 | 6,57 | 1,71 |
| | 8 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 4 | 335,10 | 0,98 | 224,64 | 6,57 | 1,71 |
| II | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 2 | 335,85 | 0,98 | 235,26 | 6,80 | 1,77 | |
| | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 2 | 335,85 | 0,98 | 235,26 | 6,80 | 1,77 | |
| III | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 5 | 333,94 | 0,99 | 216,41 | 6,44 | 1,20 | |
| | 2 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 5 | 333,94 | 0,99 | 216,41 | 6,44 | 2,76 | |
| | 3 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 5 | 333,94 | 0,99 | 216,41 | 6,44 | 1,20 | |
| | 4 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 5 | 333,94 | 0,99 | 216,41 | 6,44 | 1,30 | |
| | 5 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 5 | 333,94 | 0,99 | 216,41 | 6,44 | 2,76 | |
| | 6 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 5 | 333,94 | 0,99 | 216,41 | 6,44 | 1,20 | |
| | 7 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 5 | 333,94 | 0,99 | 216,41 | 6,44 | 1,30 |
| | 8 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 5 | 333,94 | 0,99 | 216,41 | 6,44 | 1,30 |
| IV | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 6 | 349,93 | 0,96 | 298,59 | 8,47 | 1,34 | |
| | 2 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 6 | 349,93 | 0,96 | 298,59 | 8,47 | 1,34 | |
| | 3 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 6 | 349,93 | 0,96 | 298,59 | 8,47 | 1,34 | |
| | 4 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 6 | 349,93 | 0,96 | 298,59 | 8,47 | 1,34 | |
| | 5 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 6 | 349,93 | 0,96 | 298,59 | 8,47 | 1,31 | |
| | 6 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 6 | 349,93 | 0,96 | 298,59 | 8,47 | 1,31 | |
| | 7 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 6 | 349,93 | 0,96 | 298,59 | 8,47 | 1,34 | |
| | 8 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 6 | 349,93 | 0,96 | 298,59 | 8,47 | 1,34 | |
| | 9 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 6 | 349,93 | 0,96 | 298,59 | 8,47 | 1,31 |

Tabela 2 (cd.)

| Wariant | Lp. modelu | t | s_1 | s_2 | s_3 | s_4 | s_5 | c_1 | c_2 | c_3 | c_4 | c_5 | c_6 | Liczba składowych | AIC | R^2 | S_e | V_s (%) | F^* |
|---------|------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-------------------|---------------|-------------|---------------|-------------|-------------|
| V | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 4 | 352,27 | 0,96 | 321,23 | 9,93 | 1,24 |
| | 2 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 4 | 352,27 | 0,96 | 321,23 | 9,93 | 1,24 |
| | 3 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 4 | 352,27 | 0,96 | 321,23 | 9,93 | 1,24 |
| | 4 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 4 | 352,27 | 0,96 | 321,23 | 9,93 | 1,24 |
| | 5 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 4 | 352,27 | 0,96 | 321,23 | 9,93 | 1,24 |
| | 6 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 4 | 352,27 | 0,96 | 321,23 | 9,93 | 1,24 |
| | 7 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 4 | 352,27 | 0,96 | 321,23 | 9,93 | 1,24 |
| | 8 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 4 | 352,27 | 0,96 | 321,23 | 9,93 | 1,24 |
| VI | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 6 | 352,79 | 0,96 | 316,89 | 8,69 | 1,01 |
| | 2 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 6 | 352,79 | 0,96 | 316,89 | 8,69 | 1,01 |
| | 3 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 6 | 352,79 | 0,96 | 316,89 | 8,69 | 1,50 |
| | 4 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 6 | 352,79 | 0,96 | 316,89 | 8,69 | 1,01 |
| | 5 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 6 | 352,79 | 0,96 | 316,89 | 8,69 | 1,01 |
| | 6 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 6 | 352,79 | 0,96 | 316,89 | 8,69 | 1,50 |

Pogrubioną czcionką wyróżniono modele charakteryzujące się najniższym poziomem średnich względnych błędów prognoz inter- lub/i ekstrapolacyjnych.

W trzecim wariancie, podobnie jak w pierwszym, zanotowano osiem kombinacji zmiennych niezależnych, które zawierały pięć składowych harmoniczných. Modele w tym wariancie charakteryzowały się najlepszym dopasowaniem wartości teoretycznych do wartości rzeczywistych, a także poziomem ocen S_e oraz V_s wśród wszystkich wariantów luk.

W czwartym wariancie odnotowano wystąpienie dziewięciu sześćoelementowych kombinacji składowych harmoniczných. Modele zbudowane z użyciem tych kombinacji charakteryzowały się S_e na poziomie 298,59 tys. zł przy V_s wynoszącym 8,47 punktów procentowych.

W piątym wariancie, podobnie jak w pierwszym i trzecim, odnotowano jedenaście czteroelementowych kombinacji składowych harmoniczných. Modele te charakteryzowały się odchyleniem standardowego składnika losowego na poziomie 321,23 tys. zł oraz współczynnikiem zmienności losowej wynoszącym 9,93%.

W szóstym wariancie odnotowano prawie o połowę mniej wersji niż w piątym. W odróżnieniu od piątego wariantu wersje te zawierały sześć zmiennych harmoniczných. Modele zbudowane na podstawie tych kombinacji charakteryzowały się poziomem ocen mierników S_e oraz V_s niższymi niż w przypadku modeli z piątego wariantu niezawierających zmiennej czasowej.

W tabeli 3 zestawiono oceny błędów prognoz interpolacyjnych dla miesięcy, w których wystąpiły luki w danych, oraz prognoz ekstrapolacyjnych dla horyzontu wynoszącego 12 miesięcy. Dodatkowo dla obu rodzajów prognoz podano informacje dotyczące kolejności ocen błędów w ramach poszczególnych wariantów luk w danych.

Najniższą ocenę błędu prognoz interpolacyjnych w pierwszym wariancie otrzymano dla kombinacji t , s_1 , s_2 , c_1 . Wyniósł on 7,78 punktu procentowego. Najmniej dokładne okazały się prognozy uzyskane z pomocą modelu zawierające składowe t , s_4 , s_5 , c_1 . Ocena ta jest wyższa o 5,37 punktów procentowych od oceny otrzymanej dla najlepszego modelu. Średnie względne błędy prognoz ekstrapolacyjnych zawierały się w przedziale 9,38–11,30%. Najlepszym predyktorem okazał się model zawierający składowe t , s_1 , s_2 , c_5 , przy czym ocena ta była niższa niż dla modelu o najniższym błędzie prognoz interpolacyjnych. Analiza rang wskazuje, że w przypadku tego wariantu luk w istotny sposób na dokładność prognoz wpłynęły składowe s_1 i s_5 . Wersje zawierające pierwszą z nich (rangi 1–4) charakteryzowały się błędami prognoz interpolacyjnych niższymi co najmniej o 2,43 punktu procentowego niż pozostałe. Jednocześnie oceny błędów prognoz ekstrapolacyjnych otrzymane dla wersji zawierających składową s_5 były o co najmniej 0,53 punktu procentowego wyższe od tych, które jej nie zawierały.

Tabela 3. Średnie względne błędy prognoz inter- (B.I.) i ekstrapolacyjnych (B.E.) (w %) oraz odpowiadające im rangi

| Wariant | Lp. modelu | B.I. | Rangi B.I. | B.E. h = 12 | Rangi B.E. h = 12 |
|---------|------------|-------|------------|-------------|-------------------|
| I | 1 | 12,97 | 7 | 11,18 | 6 |
| | 2 | 13,15 | 8 | 11,30 | 8 |
| | 3 | 12,46 | 6 | 10,68 | 5 |
| | 4 | 12,39 | 5 | 11,19 | 7 |
| | 5 | 9,96 | 4 | 9,88 | 3 |

Tabela 3 (cd.)

| Wariant | Lp. modelu | B.I. | Rangi B.I. | B.E. h = 12 | Rangi B.E. h = 12 |
|---------|------------|--------------|------------|-------------|-------------------|
| I | 6 | 9,41 | 3 | 10,15 | 4 |
| | 7 | 8,64 | 2 | 9,38 | 1 |
| | 8 | 7,78 | 1 | 9,85 | 2 |
| II | 1 | 11,93 | 2 | 4,70 | 2 |
| | 2 | 8,66 | 1 | 4,27 | 1 |
| III | 1 | 22,78 | 8 | 11,04 | 8 |
| | 2 | 18,88 | 6 | 7,91 | 1 |
| | 3 | 17,28 | 5 | 9,30 | 4 |
| | 4 | 21,99 | 7 | 9,13 | 3 |
| | 5 | 15,12 | 3 | 10,71 | 7 |
| | 6 | 14,95 | 2 | 10,08 | 6 |
| | 7 | 11,23 | 1 | 8,56 | 2 |
| | 8 | 16,58 | 4 | 9,34 | 5 |
| IV | 1 | 17,82 | 8 | 8,44 | 7 |
| | 2 | 19,69 | 9 | 8,64 | 9 |
| | 3 | 13,32 | 2 | 6,59 | 2 |
| | 4 | 14,98 | 4 | 7,10 | 5 |
| | 5 | 14,78 | 3 | 6,74 | 4 |
| | 6 | 17,16 | 6 | 6,66 | 3 |
| | 7 | 17,23 | 7 | 8,61 | 8 |
| | 8 | 11,99 | 1 | 6,45 | 1 |
| | 9 | 15,86 | 5 | 7,24 | 6 |
| V | 1 | 23,66 | 9 | 8,69 | 9 |
| | 2 | 27,69 | 10 | 11,18 | 11 |
| | 3 | 29,34 | 11 | 10,29 | 10 |
| | 4 | 15,42 | 6 | 6,94 | 6 |
| | 5 | 17,51 | 7 | 7,17 | 8 |
| | 6 | 14,22 | 4 | 6,13 | 2 |
| | 7 | 17,60 | 8 | 6,41 | 3 |
| | 8 | 15,03 | 5 | 6,09 | 1 |
| VI | 1 | 12,85 | 1 | 6,86 | 1 |
| | 2 | 14,39 | 2 | 7,59 | 5 |
| | 3 | 16,85 | 3 | 7,41 | 4 |
| | 4 | 17,64 | 4 | 7,30 | 3 |
| | 5 | 19,02 | 5 | 6,94 | 2 |
| | 6 | 20,13 | 6 | 8,35 | 6 |

Pogrubioną czcionką wyróżniono modele charakteryzujące się najniższym poziomem średnich względnych błędów prognoz inter- lub/i ekstrapolacyjnych.

W drugim wariacie najlepsze prognozy interpolacyjne uzyskano za pomocą modelu charakteryzującego się oceną błędów prognoz wynoszącą 8,66%, zawierającego tylko składowe t , c_1 . Ocena ta była o 3,27 punktu procentowego niższa niż uzyskana dla modelu zawierającego składowe t , c_5 . Relacja predyktorów w przypadku błędów ekstrapolacyjnych jest taka sama, przy czym różnica jest znacznie niższa.

Najniższą ocenę błędu prognoz interpolacyjnych w trzecim wariacie, wynoszącą 11,23 punktu procentowego, otrzymano dla wersji zawierającej składowe t , s_1 , c_1 , c_2 , c_4 . Ocena

błędów ekstrapolacyjnych dla horyzontu wynoszącego 12 miesięcy dla tych składowych była o 0,65 punktu procentowego wyższa niż dla modelu o składowych t , s_5 , c_1 , c_2 , c_5 . Zwraca uwagę fakt, że dwukrotnie wyższe błędy w przypadku prognoz interpolacyjnych otrzymano dla modeli zawierających s_5 , c_4 , c_5

W czwartym wariacie najlepsze prognozy inter- oraz ekstrapolacyjne otrzymano dla wersji zawierającej składowe t , s_2 , s_3 , s_5 , c_1 , c_4 . Oceny średnich względnych błędów prognoz inter- i ekstrapolacyjnych wynosiły, odpowiednio, 11,99 i 6,45%. We wszystkich kombinacjach składowych w tym wariacie stwierdzono występowanie składowych t , s_2 , c_4 .

Najlepsze prognozy interpolacyjne w piątym wariacie otrzymano dla predyktora o czterech składowych: s_1 , s_3 , c_2 , c_5 . Należy jednak zauważyć, że oceny błędów prognoz interpolacyjnych w tym wariacie cechowały się dużym zróżnicowaniem i zawierały się w przedziale 14,22–29,34%. Oceny maksymalne były najwyższe wśród wszystkich analizowanych wariantów luk. W przypadku prognoz ekstrapolacyjnych różnice między najlepszymi a najgorszymi ocenami błędów były zdecydowanie niższe i przekraczały nieznacznie 5 punktów procentowych. Najlepsze prognozy ekstrapolacyjne otrzymano dla kombinacji s_1 , s_3 , s_5 , c_2 .

W szóstym wariacie, podobnie jak w drugim oraz czwartym, najlepsze prognozy interpolacyjne oraz ekstrapolacyjne otrzymano dla wersji zawierającej następujące składowe: t , c_1 , c_2 , c_3 , c_5 , c_6 . Średnie względne błędy prognoz interpolacyjnych wyniosły 12,85%, a ekstrapolacyjnych – 6,86%. We wszystkich kombinacjach w tym wariacie odnotowano występowanie zmiennych t , c_1 , c_6 . Jednakże na prognozy interpolacyjne wpływ miało występowanie lub też brak składowej c_2 (rangi B.I. 1–3), a w przypadku prognoz ekstrapolacyjnych składowej c_5 (rangi B.E. $h = 12$). Dokładna analiza powiązań między prognozami inter- a ekstrapolacyjnymi została przedstawiona poniżej.

Kształtowanie się najlepszych prognoz ekstrapolacyjnych dla poszczególnych wariantów luk przedstawiono na rys. 3.

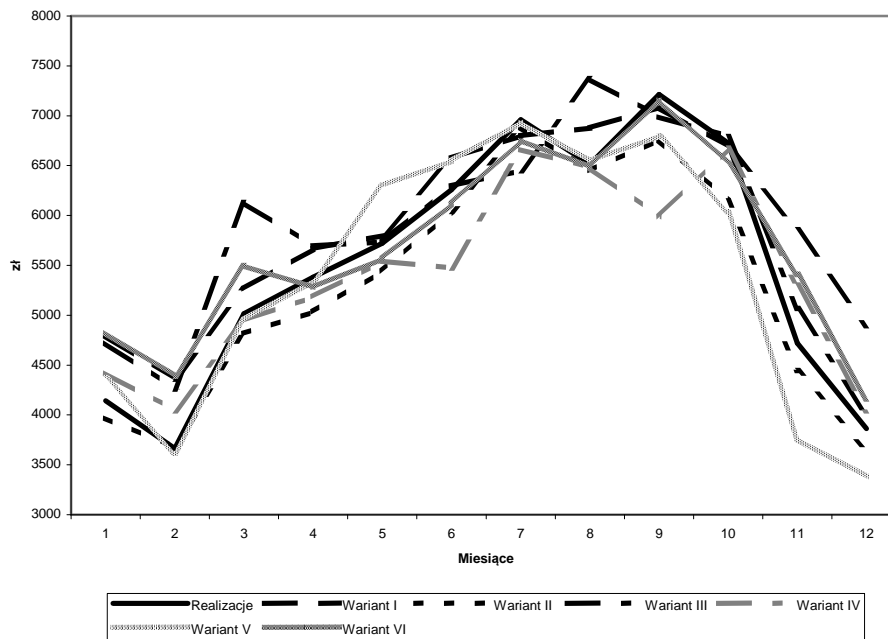
W tabeli 4 zestawiono współczynniki korelacji liniowej Pearsona pomiędzy ocenami błędów prognoz interpolacyjnych a ocenami błędów prognoz ekstrapolacyjnych w poszczególnych wariantach luk.

Tabela 4. Oceny współczynników korelacji liniowej pomiędzy średnimi względnymi błędami prognoz inter- i ekstrapolacyjnych

| Wariant | Ocena |
|---------|-------|
| I | 0,925 |
| III | 0,214 |
| IV | 0,811 |
| V | 0,829 |
| VI | 0,502 |

Z analizy informacji zawartych w tabeli 4 wynika, że w przypadku wariantów pierwszego, czwartego oraz piątego występowała silna zależność między ocenami obu rodzajów prognoz. Znacznie niższe oceny tych współczynników otrzymano dla wariantów trzeciego

i szóstego. Zróźnicowanie ocen współczynników oznacza, że układ luk w danych ma częściowy wpływ na relacje dokładności prognoz inter- i ekstrapolacyjnych.



Rys. 3. Najlepsze prognozy ekstrapolacyjne kształtowania się kosztów całkowitych produkcji betonu komórkowego dla poszczególnych wariantów luk

PODSUMOWANIE

W pracy wykazano przydatność zaprezentowanej procedury w prognozowaniu brakujących informacji w ekonomicznych szeregach czasowych z wahaniami sezonowymi dla luk systematycznych. Z badań przeprowadzonych w artykule wynika, że:

1. Poziom ocen błędów, zarówno inter-, jak i ekstrapolacyjnych, a także liczba kombinacji zmiennych w danym wariantcie zależy od rozkładu luk.
2. Dla każdego wariantu poszczególne wersje modeli są nierozróżnialne z punktu widzenia miar oceny stochastycznej struktury modelu. Dlatego konieczne jest przeprowadzenie analizy *ex-post* dokładności prognoz inter- i ekstrapolacyjnych.
3. W większości wariantów stwierdzono, że pojedyncze składowe lub grupy składowych w kombinacji mogą decydować o dokładności prognoz. Ich typy zależą głównie od rozkładu luk w danych.
4. Nie stwierdzono wyraźnego wpływu liczby składowych harmonicznych w modelu na jakość prognoz. W niektórych przypadkach modele z większą ich liczbą dawały gorsze prognozy niż w przypadku, gdy było ich mniej.
5. Oceny błędów prognoz interpolacyjnych były w większości przypadków wyższe niż oceny prognoz ekstrapolacyjnych we wszystkich badanych wariantach.
6. Z analizy współczynników korelacji liniowej Pearsona między ocenami błędów prognoz inter- i ekstrapolacyjnych uzupełnionej analizą rang wynika, że w trzech z pięciu anali-

zowanych wariantów modele o najlepszych prognozach interpolacyjnych charakteryzowały się względnie niskimi ocenami błędów prognoz ekstrapolacyjnych.

PIŚMIENNICTWO

Anderson D.R., Sweeney D.J., Williams T.A. 2011. *Statistics for Business and Economics*. South-Western College Pub., 739.

Zastosowanie hierarchicznych modeli szeregów czasowych w prognozowaniu zmiennych ekonomicznych z wahaniami sezonowymi. 2003. Red. J. Zawadzki. Szczecin, Wydaw. Akademii Rolniczej w Szczecinie, 26.