Zeszyty Naukowe
Wyższej Szkoły Bankowej we Wrocławiu
Nr 20/2011

Tadeusz Kufel

Wyższa Szkoła Bankowa w Toruniu

Ocena efektywności procedury *Congruent Specyfication* dla małych prób

Streszczenie. Procedura specyfikacji dynamicznego modelu zgodnego zapewnia, że każdy estymowany model (GUM) będzie posiadał proces resztowy o własnościach białego szumu. Celem artykułu jest przedstawienie procedury automatycznej specyfikacji modelu według koncepcji modelowania zgodnego, dostępnej pod nazwą *Congruent Specyfication*, jako pakiet funkcji programu GRETL, oraz ocena jego efektywności w warunkach zmniejszającej się liczebności próby. Empiryczny przykład prezentuje różne aspekty skracania próby ze 120 do 24 obserwacji.

Słowa kluczowe: modelowanie zgodne, automatyczne procedury modelowania, GRETL

Wprowadzenie

Celem artykułu jest przedstawienie procedury automatycznej specyfikacji modelu[1] według koncepcji modelowania zgodnego, dostępnej jako pakiet funkcji

[1] Szerszy opis algorytmu automatyzującego proces modelowania i prognozowania pod nazwą AUTOMETRICS, autorstwa J.Doornika, można znaleźć w pracach: J.A. Doornik, *Encompassing and Automatic Model Selection*, „Oxford Bulletin of Economics and Statistics" 2008, tom 70,

programu GRETL[2], oraz ocena jego efektywności dla zmniejszającej się liczebności próby.

Autorem koncepcji dynamicznego modelowania zgodnego[3] jest Zygmunt Zieliński. Zgodność w nazwie modelu rozumiana jest jako zgodność harmonicznej struktury procesu objaśnianego z łączną harmoniczną strukturą procesów objaśniających oraz procesu resztowego. Proces resztowy jest niezależny od procesów objaśniających. Najprostszym przykładem modelu, który jest zawsze zgodny, jest model zbudowany dla procesów białoszumowych o postaci:

$$\varepsilon_{yt} = \sum_{i=1}^{k} \rho_i \varepsilon_{x_i t} + \varepsilon_t . \tag{1}$$

Powyższy model jest zawsze zgodny, gdyż struktura harmoniczna procesu ε_{yt} jest równa (lub inaczej spektrum tego procesu jest równoległe względem osi częstości) łącznej strukturze procesów $\varepsilon_{x_i t}$ oraz procesu ε_t, ponieważ funkcja gęstości spektralnej dla procesu o własnościach białego szumu jest stała.

Zgodny dynamiczny model ekonometryczny wykorzystuje przy budowie modelu informację na temat wewnętrznej struktury każdego uwzględnionego w badaniu procesu. Przyjmując za Y_t proces objaśniany oraz za X_{it} ($i = 1, 2, ..., k$) procesy objaśniające, wewnętrzna struktura opisywana jest za pomocą modeli podstawowych i zawiera:

– modele opisujące składniki niestacjonarne:

$$Y_t = P_{yt} + S_{yt} + \eta_{yt}, \qquad X_{it} = P_{x_i t} + S_{x_i t} + \eta_{x_i t}, \tag{2}$$

gdzie: P_{yt}, $P_{x_i t}$ – wielomianowe funkcje zmiennej czasowej t dla odpowiednich procesów, S_{yt}, $S_{x_i t}$ – składniki sezonowe o stałej lub zmiennej amplitudzie wahań dla odpowiednich procesów, η_{yt}, $\eta_{x_i t}$ – stacjonarne autoregresyjne procesy odnoszące się do odpowiednich procesów oraz

– modele autoregresyjne:

$$B(u)\eta_{yt} = \varepsilon_{yt}, \qquad A_i(u)\eta_{x_i t} = \varepsilon_{x_i t}, \tag{3}$$

gdzie: $B(u)$, $A_i(u)$ – stacjonarne autoregresyjne operatory, dla których wszystkie pierwiastki równania $|B(u)| = 0$ i $|A_i(u)| = 0$ leżą poza okręgiem jednostkowym, ε_{yt}, $\varepsilon_{x_i t}$ – białe szumy dla odpowiednich procesów.

---

s. 915-925; J.A. Doornik, *Autometrics*, w: J.L. Castle, N. Shephard, *The Methodology and Practice of Econometrics*, Oxford University Press 2009; J.A. Doornik, D.F. Hendry, *Modelling Dynamic Systems Using PcGive 11*, Vol. II, wyd. 4, Timberlake Consultants Press, London 2009.

[2] Pakiet funkcji *Congruent Specification* jest autorstwa: Marcina Błażejowskiego, Pawła Kufla oraz Tadeusza Kufla.

[3] Por. Z. Zieliński, *Zmienność w czasie strukturalnych parametrów modelu ekonometrycznego*, „Przegląd Statystyczny" 1984, R. XXXI, z. 1/2, s. 135-148; L. Talaga, Z. Zieliński, *Metody spektralne w modelowaniu ekonometrycznym*, PWN, Warszawa 1986.

Budowa zgodnego dynamicznego modelu ekonometrycznego odbywa się przez podstawienie do zależności (1) procesów ε_{yt}, ε_{x_jt} o własnościach białych szumów otrzymanych z równań (3), a następnie na podstawie równań (2) wyznaczane są procesy autoregresyjne η_{yt}, η_{x_jt} i podstawiane do otrzymanego wcześniej równania. Po przekształceniu otrzymuje się zgodny dynamiczny model ekonometryczny dla rzeczywistych procesów Y_t i X_{it}:

$$B(u)Y_t = \sum_{i=1}^{k} A_i'(u)X_{it} + P_t + S_t + \varepsilon_t. \tag{4}$$

Proces resztowy w powyższym modelu ma własności takie same jak w modelu (1), a więc posiada własności białego szumu, zatem warunek zgodności struktur harmonicznych został spełniony.

Zgodny dynamiczny model ekonometryczny uwzględnia informację o wewnętrznej strukturze (składniki trendowe i sezonowe oraz autoregresyjne) oraz zależności przyczynowo-skutkowe wykorzystanych procesów. Informacje te wykorzystywane są na etapie specyfikacji modelu.

1. Opis algorytmu specyfikacji zgodnego dynamicznego modelu ekonometrycznego

Zaimplementowany w oprogramowaniu GRETL algorytm specyfikacji zgodnego dynamicznego modelu ekonometrycznego został podzielony na trzy etapy[4].

Etap 1. Wczytanie danych do procedury:

1. Wskazanie zmiennej endogenicznej Y, listy k potencjalnych zmiennych objaśniających X_j ($j = 1, 2, \ldots, k$) oraz listy deterministycznych zmiennych zero-jedynkowych.
2. Określenie długości próby n i ustalenie minimalnej liczby stopni swobody *dfmin* nieograniczonego ogólnego modelu:

$$dfmin = \begin{cases} round(0.1 * n) & \text{dla } n \le 200, \\ 20 & \text{dla } n > 200. \end{cases}$$

3. Sprawdzenie częstotliwości modelowanych procesów i ustalenie:
   1) bloku periodycznych zmiennych zero-jedynkowych m,
   2) maksymalnego rzędu autoregresji *pmax* dla wykorzystywanych procesów.

[4] Opracowano na podstawie: M. Błażejowski, P. Kufel, T. Kufel, *Automatyczna procedura budowy specyfikacji zgodnego dynamicznego modelu ekonometrycznego w oprogramowaniu GRETL*, Acta Universitatis Copernici, Ekonomia XXXIX, Wydawnictwo Naukowe UMK, Toruń 2009, s. 83-92.

Etap 2. Analiza wewnętrznej struktury wszystkich procesów:

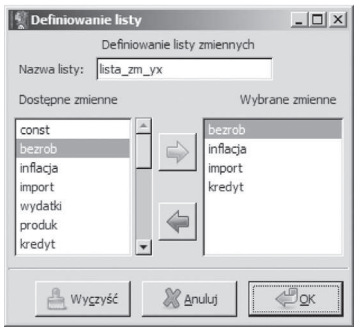
1. Sprawdzenie, czy analizowane procesy posiadają deterministyczny komponent (trend i/lub sezonowość).
2. Sprawdzenie, czy reszty po odjęciu bloku deterministycznych komponentów wykazują stacjonarną strukturę autoregresyjną.
3. Określenie rzędów funkcji autokorelacji cząstkowej dla procesów resztowych po odjęciu komponentów deterministycznych oraz różnicowaniu (w przypadku stwierdzenia pierwiastka jednostkowego), uwzględniając maksymalny rząd autoregresji. Wartości funkcji PACF sprowadzane są do wartości *max_lag*, określonej następująco:

$$max_lag = \begin{cases} round(10 * \log_{10}(n)) & \text{dla } n > 100, \\ round(0.2 * n) & \text{dla } n \leq 100. \end{cases}$$

Etap 3. Zbudowanie startowej specyfikacji nieograniczonego ogólnego (GUM), tzw. pełnego modelu zgodnego:

1. Wyznaczenie liczby stopni swobody modelu pełnego *dfstart* po uwzględnieniu wszystkich składników: rozkładu opóźnień zmiennej endogenicznej Y, potencjalnych zmiennych objaśniających X_j, deterministycznego trendu i/lub sezonowości oraz deterministycznych zmiennych zero-jedynkowych:
   1) jeżeli *dfstart* < *dfmin*, to rzędy autoregresji o najwyższych wartościach są zmniejszane o 1,
   2) jeżeli *dfstart* ≥ *dfmin*, to startowa specyfikacja (model pełny) jest drukowana w formie skryptu wewnętrznego języka programu Gretl.

Etap 1 jest przygotowaniem danych do wykonania procedury. Importowane są wszystkie szeregi czasowe i na ich podstawie, jeśli konieczne, dodatkowe zmienne zero-jedynkowe opisujące pewne okresy w próbie. W tym etapie określana jest częstotliwość obserwowania procesów oraz inne dodatkowe zmienne wewnętrzne, jak np. minimalna liczba stopni swobody dla modelu pełnego – *dfmin*.

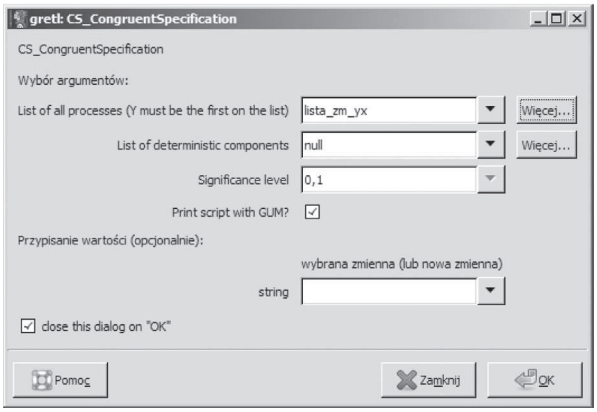


Rys. 1. Definiowanie listy zmiennych dla procedury *Congruent Specification*

Ź r ó d ł o: opracowanie własne.

Rysunek 1 przedstawia definiowanie listy zmiennych wykorzystywanych w procedurze. Należy zwrócić uwagę, aby pierwszą zmienną na liście była zmienna objaśniana.

Rysunek 2 przedstawia okno dialogowe wyboru listy zmiennych, listy składników deterministycznych oraz wartości poziomu istotności.



Rys. 2. Główne okno procedury *Congruen Specification*

Ź r ó d ł o: opracowanie własne.

W drugim etapie algorytmu przeprowadzana jest analiza wewnętrznej struktury wykorzystywanych szeregów czasowych. W pierwszej kolejności sprawdza się, czy występują istotne składniki deterministyczne, takie jak trend liniowy oraz cykliczność deterministyczna. Określona w etapie pierwszym częstotliwość obserwowania determinuje bloki (m – 1) zmiennych zero-jedynkowych: dla danych miesięcznych (m = 12), kwartalnych (m = 4), tygodniowych (m = 52) lub danych dziennych (m = 5, 6, 7).

Jeśli w którymkolwiek z analizowanych szeregów wystąpił istotny składnik deterministyczny, to zostanie uwzględniony w bloku wspólnych komponentów deterministycznych. Następnie budowany jest model dla składników deterministycznych, a w jego resztach sprawdzany jest stopień integracji procesu za pomocą testu pierwiastka jednostkowego ADF-GLS na poziomie istotności $\alpha = 0{,}1$. Gdy występuje pierwiastek jednostkowy, wówczas proces jest różnicowany. W kolejnym kroku określane są rzędy autoregresji, co odbywa się za pomocą wartości funkcji PACF i ustalane jest jako maksymalna istotna wartość funkcji na poziomie $\alpha = 0{,}01$.

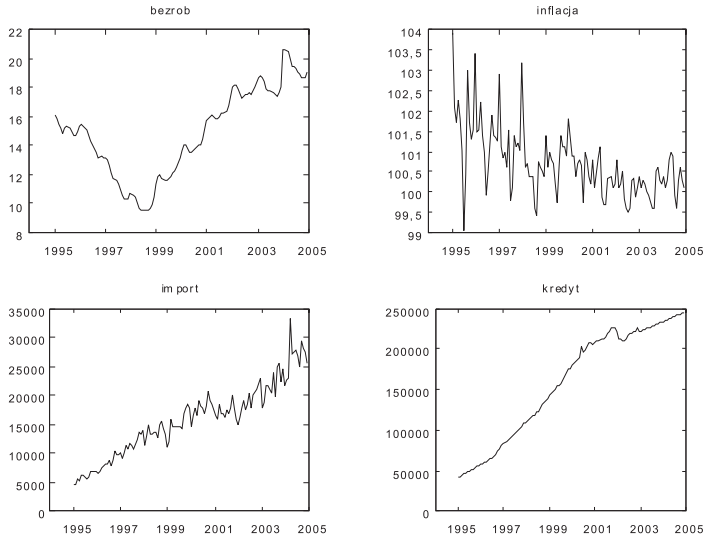
W trzecim etapie procedury budowana jest startowa specyfikacja zgodnego modelu ekonometrycznego, zawierająca informacje uzyskane we wcześniejszych etapach: komponenty deterministyczne, stopnie integracji procesów oraz rzędy autoregresji procesów. Dodatkowo sprawdzany jest warunek minimalnej liczby stopni swobody. W przypadku gdy liczba stopni swobody modelu startowego

jest mniejsza niż minimalna wartość stopni swobody *dfmin*, wówczas redukuje się maksymalne rzędy autoregresji o 1, aż do momentu spełnienia warunku o minimalnej liczbie stopni swobody. Rezultatem wywołania procedury jest skrypt wewnętrznego języka programu GRETL, który przedstawia specyfikację modelu zgodnego.

Przykłady działania algorytmu dla zmniejszającej się próby przedstawione zostały w następnej części artykułu.

2. Przykłady wykorzystania automatycznej procedury specyfikacji zgodnego dynamicznego modelu ekonometrycznego

Dla empirycznego zilustrowania sposobu wykorzystania automatycznej procedury *Congruent Specification* zostaną wykorzystane miesięczne dane makroekonomiczne dla Polski za okres od stycznia 1995 do grudnia 2004 (n = 120), które są dostępne na polskiej stronie internetowej programu GRETL znajdującej się pod adresem http://www.gretl.pl w pliku *macro_1993_2005.gdt*. Analizowany model dotyczyć będzie stopy bezrobocia opisywanej przez procesy: stopy inflacji, importu oraz udzielonych kredytów.



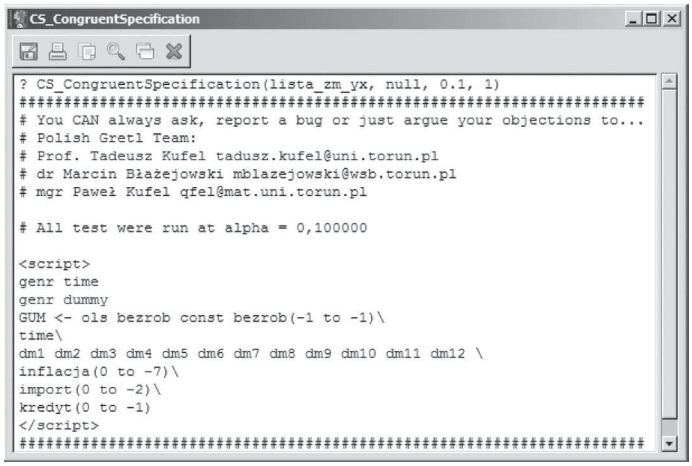
Rys. 3. Modelowane procesy w procedurze *Congruent Specification*

Ź r ó d ł o: opracowanie własne.

Scenariusz badania efektywności automatycznej procedury specyfikacji zakładał, że ocenę liczby stopni swobody, liczby szacowanych parametrów, rzędów opóźnień procesów, a także efektywności otrzymanych 12 miesięcznych prognoz

za pomocą średnich miar *ex post*: RMSE i MAPE, przy założeniu zmniejszającej się liczby obserwacji o 1 miesiąc, to jest od liczebności startowej $n = 120$ (od stycznia 1995 do grudnia 2004 – 10 lat) do liczebności końcowej $n = 24$ (od stycznia 2003 do grudnia 2004 – 2 lata). Łącznie automatyczną procedurę specyfikacji wykorzystano 96 razy.

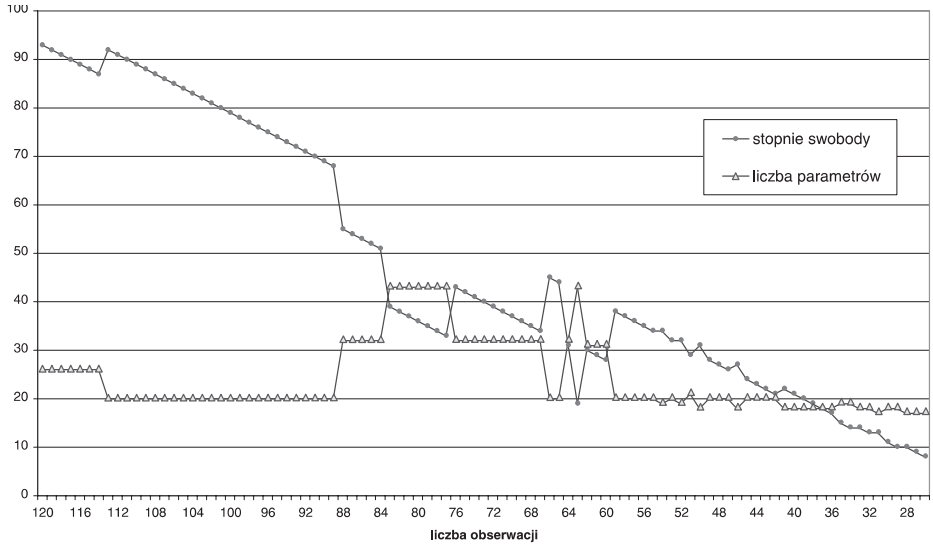
Zmienną objaśnianą była stopa bezrobocia, natomiast potencjalnymi zmiennymi objaśniającymi były: stopa inflacji, import oraz wartość udzielonych kredytów. Rysunek 4 prezentuje wynik działania procedury *Congruent Specification* zrealizowanej na wyżej wymienionych danych dla $n = 120$, przy czym weryfikacja testów statystycznych została przeprowadzona na poziomie istotności $\alpha = 0{,}1$.



Rys. 4. Specyfikacja modelu stopy bezrobocia w Polsce wg koncepcji modelowania zgodnego uzyskana w procedurze *Congruent Specification*

Ź r ó d ł o: opracowanie własne.

Nieograniczony model ogólny (GUM) powinien zawierać opóźnienie zmiennej bezrobocie o jeden okres, trend liniowy (time), zestaw zmiennych zero-jedynkowych do opisu sezonowości miesięcznej (dm1, dm2, dm3, …), inflację bieżącą oraz opóźnioną do 7 okresów włącznie, import bieżący oraz opóźniony do 2 okresów włącznie, bieżącą wartość kredytów oraz opóźnioną o 1 okres. Oznacza to, że oszacowany model GUM posiadał: $n = 120$, $k = 26$, a stopni swobody $s = 93$. Weryfikacja i eliminacja nieistotnych zmiennych przeprowadzona zgodnie z zasadą krokowej eliminacji *a posteriori* na poziomie istotności doprowadziła do uzyskania empirycznego modelu zgodnego, dla którego oszacowano prognozy i wyznaczono średnie błędy *ex post* dla 12 okresów (RMSE, MAPE). W kolejnych etapach, przy założeniu zmniejszającego się okresu modelowania o jeden miesiąc, liczbę stopni swobody i liczbę szacowanych parametrów przedstawia rysunek 5.



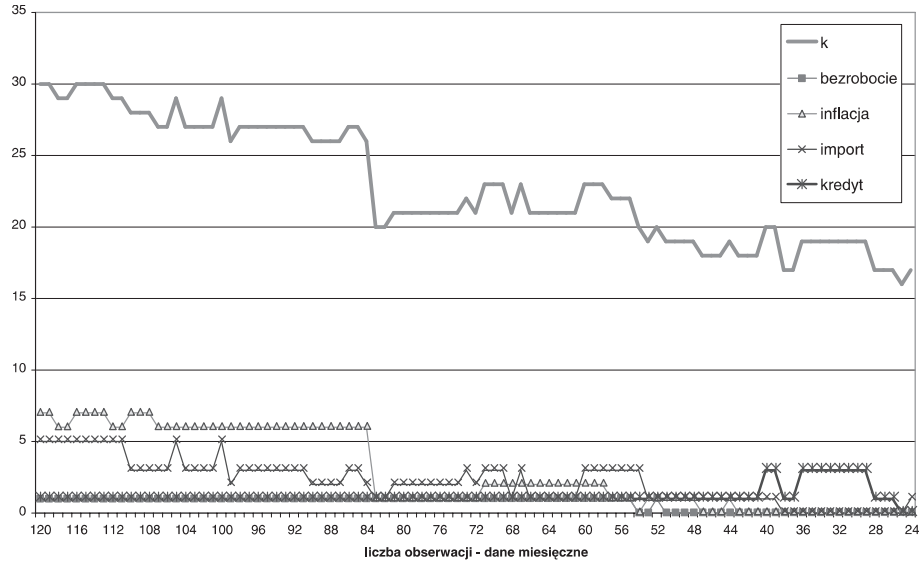
Rys. 5. Liczba stopni swobody i liczba szacowanych parametrów dla empirycznego modelu zgodnego stopy bezrobocia w Polsce według zmniejszającej się liczby obserwacji

Ź r ó d ł o: opracowanie własne.

Na szacowane parametry w powyższych modelach składa się: wyraz wolny, trend liniowy, zestaw 11 zmiennych sezonowych, opóźnione procesy endogeniczne oraz trzy procesy objaśniające wraz z własnymi opóźnieniami. Efekt zmniejszającej liczby obserwacji powoduje w pierwszej kolejności zmniejszenie się liczebności procesów opóźnionych, co wynika z dwóch powodów, że lag_max nie może być większy niż 20% liczebności próby, a przy zmniejszającym się n wartość krytyczna istotności dla funkcji PACF szybko rośnie ($1{,}96/n^{0{,}5}$), dlatego dla $n = 24$ wystąpił tylko jeden proces opóźniony w czasie (import_{t-1}).

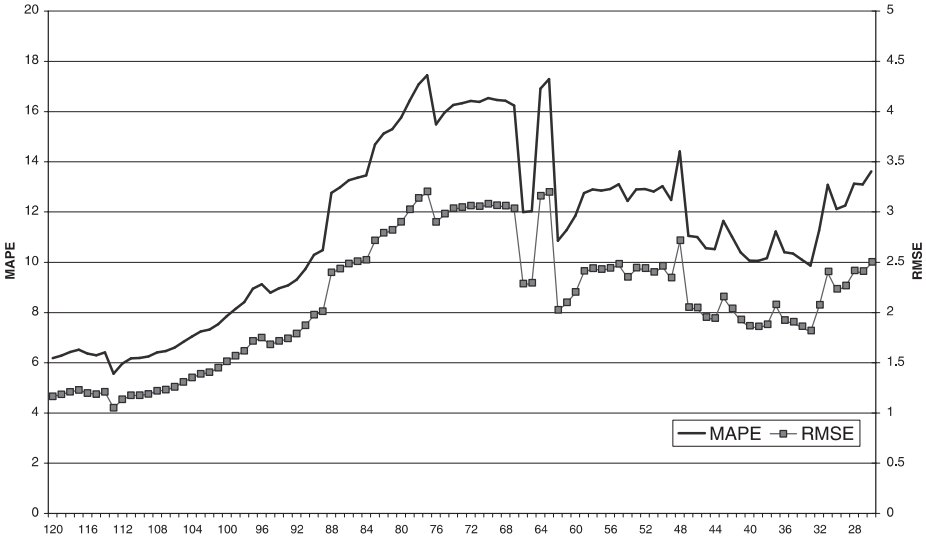
Liczbę procesów opóźnionych w modelach, to jest stopień istotnego rzędu funkcji PACF dla każdego procesu oddzielnie, a także liczbę stopni swobody dla oszacowanych modeli przy zmniejszającej się próbie od $n = 120$ do 24, przedstawia rysunek 6.

Rysunek 6 przedstawia liczebności procesów opóźnionych w nieograniczonym modelu zgodnym (GUM), które ulegają zmniejszeniu wraz ze spadkiem długości modelowanego szeregu czasowego ($n = 120, \ldots, 24$). Miesiące dla 1998 i 1999 mają najniższy poziom bezrobocia, a wzrost w następnych okresach (zmiana kierunku trendu na rosnący) dał efekt zaburzeń w ocenie rzędu autoregresji i liczby szacowanych parametrów.



Rys. 6. Liczba szacowanych parametrów wraz z rzędami opóźnień dla wszystkich procesów w empirycznym modelu zgodnym dla stopy bezrobocia w Polsce według zmniejszającej się liczby obserwacji

Ź r ó d ł o: opracowanie własne.



Rys. 7. Syntetyczne mierniki jakości prognoz

Ź r ó d ł o: opracowanie własne.

Rysunek 7 przedstawia syntetyczne mierniki jakości prognoz oszacowanych zawsze dla horyzontu 12 miesięcy 2005 r. przy zmniejszającej się długości próby. Ocenę trafności prognoz przedstawiają: pierwiastek średniokwadratowego błędu prognoz – RWSE oraz średni absolutny błąd procentowy – MAPE. Nietrafność uzyskanych prognoz wynikała z faktu wstąpienia Polski do struktur Unii Europejskiej i otwarcia europejskich rynków pracy dla Polaków, co zaowocowało znaczną emigracją, a prognozowana stopa bezrobocia miała tendencje do przeszacowania.

Powyższe scenariusz modelowania przy zmniejszającej się liczebności próby pozwalają wyciągnąć wniosek, że średnie błędy prognoz *ex post* świadczą o wysokiej jakości modelowania.

Podsumowanie

Procedura specyfikacji dynamicznego modelu zgodnego zapewnia, że każdy tak estymowany model (GUM) będzie posiadał proces resztowy o własnościach białego szumu, a jest to podstawowy warunek dobroci modelu wykorzystywanego do prognozowania.

Algorytm *Congruent Specyfication* – opracowany jako pakiet funkcji dla programu GRETL – wskazuje startową pełną specyfikację modelu, która zawiera informacje:

- jakie składniki będzie zawierał człon deterministyczny (trend/sezonowość) oraz
- jaka jest struktura autoregresyjna (opóźnień) wszystkich procesów wykorzystywanych w modelu.

Wykorzystanie funkcji estymacji i weryfikacji, a także predykcji w oprogramowaniu GRETL powoduje, że cała procedura modelowania i prognozowania jest wysoce efektywna pomimo zmniejszającej się liczebności próby.

Testowanie algorytmu *Congruent Specyfication* i włączanie kolejnych testów specyfikacji pozwoli uzyskać jeszcze doskonalszy algorytm automatyzujący proces ekonometrycznego modelowania i prognozowania.

Literatura

Błażejowski M., Kufel P., Kufel T., *Automatyczna procedura budowy specyfikacji zgodnego dynamicznego modelu ekonometrycznego w oprogramowaniu GRETL*, Acta Universitatis Nicolai Copernici, Ekonomia XXXIX, Wydawnictwo Naukowe UMK, Toruń 2009.

Doornik J.A., *Encompassing and Automatic Model Selection*, „Oxford Bulletin of Economics and Statistics" 2008, tom 70.

Doornik J.A., *Autometrics*, w: J.L. Castle, N. Shephard, *The Methodology and Practice of Econometrics*, Oxford University Press 2009.

Doornik J.A., Hendry D.F., *Modelling Dynamic Systems Using PcGive 11*, Vol. II, wyd. 4, Timberlake Consultants Press, London 2009.

Talaga L., Zieliński Z., *Metody spektralne w modelowaniu ekonometrycznym*, PWN, Warszawa 1986.

Zieliński Z., *Zmienność w czasie strukturalnych parametrów modelu ekonometrycznego*, „Przegląd Statystyczny" 1984, R. XXXI, z. 1/2.

Zieliński Z., *Liniowe modele zgodne opisujące zależności sumacyjnych (zintegrowanych) procesów ekonomicznych*. w: *Przestrzenno-czasowe modelowanie i prognozowanie zjawisk gospodarczych*, red. A. Zeliaś, AE Kraków, Kraków 1995.