

* Wyższa Szkoła Bankowa w Toruniu
Katedra Metod Ilościowych

** Uniwersytet Mikołaja Kopernika w Toruniu
Katedra Ekonometrii i Statystyki

Marcin Błażejowski^{}, Paweł Kufel^{**}, Tadeusz Kufel^{**}*

AUTOMATYCZNA PROCEDURA BUDOWY SPECYFIKACJI ZGODNEGO DYNAMICZNEGO MODELU EKONOMETRYCZNEGO W OPROGRAMOWANIU GRETL

Z a r y s t r e ś c i. W artykule podjęto próbę budowy automatycznej procedury specyfikacji pełnego zgodnego dynamicznego modelu ekonometrycznego w implementacji dla oprogramowania GRETL. Wykorzystując koncepcje modelowania zgodnego autorstwa prof. Z. Zielińskiego zbudowano algorytm postępowania w badaniu parametrów elementów wewnętrznej struktury wykonywanych procesów.

S ł o w a k l u c z o w e: koncepcja modelowania zgodnego, model pełny, GUM, procedury automatycznej selekcji, PcGets, RETINA, GRETL.

1. WPROWADZENIE

Punktem wyjścia koncepcji budowy wstępnej specyfikacji dynamicznych modeli ekonometrycznych były różne założenia. Jedne dotyczyły związków przyczynowych – struktur współzależnościowych, a inne struktur wewnętrznych procesów z pominięciem przyczynowości, a jeszcze inne zakładały uwzględnienie jednego i drugiego założenia.

Autorem koncepcji dynamicznego modelowania zgodnego¹, które uwzględnia jedno i drugie założenie jest Profesor Zygmunt Zieliński. Przez zgodność modelu w sensie Zielińskiego rozumie się zgodność harmonicznej struktury procesu objaśnianego z łączną harmoniczną strukturą procesów objaśniających oraz procesu resztowego, który jest niezależny od procesów objaśniających.

Znajomość wewnętrznej struktury wszystkich badanych procesów umożliwia budowę dynamicznego modelu zgodnego.

¹ Por. Zieliński (1984, 1993, 1994, 1995), Talaga, Zieliński (1986).

Model zgodny² dla rzeczywistych procesów Y_t i X_{it} ma następującą postać:

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

Koncepcja dynamicznego modelowania zgodnego zwraca uwagę na konieczność uwzględnienia już na etapie specyfikacji modelu informacji o wewnętrznej strukturze zastosowanych procesów, do których zaliczamy:

- składnik trendowy (P_t) i sezonowy (S_t),
- wielomiany autoregresyjne $B(u)$ i $A_i(u)$.

Celem artykułu jest zaprezentowanie elementów automatycznej procedury specyfikacji zgodnego dynamicznego modelu ekonometrycznego, wyznaczającej deterministyczny składnik trendowo-sezonowy oraz autoregresyjne struktury wykorzystywanych procesów egzogenicznych i endogenicznego. Ponadto, przedstawienie analizy porównawczej z algorytmami PcGets i RETINA.

2. ALGORYTM AUTOMATYCZNEJ SPECYFIKACJI ZGODNEGO DYNAMICZNEGO MODELU EKONOMETRYCZNEGO

Poniżej znajduje się ogólny opis algorytmu automatycznej specyfikacji zgodnego dynamicznego modelu ekonometrycznego dostępnego w programie Gretl. W przedstawionym opisie nie są omówione wewnętrzne mechanizmy całej procedury, wykorzystywane zmienne lokalne ani wewnętrzne i zewnętrzne funkcje. W schemacie pokazano jedynie ogólną ideę proponowanej procedury oraz jej zgodność z koncepcją zgodnych, dynamicznych modeli ekonometrycznych Profesora Zygmunta Zielińskiego.

Etap 1: wczytanie danych do procedury:

1. Wskazanie zmiennej endogenicznej Y , listy potencjalnych zmiennych objaśniających X oraz listy deterministycznych zmiennych zero-jedynkowych.
2. Określenie długości próby i ustalenie minimalnej liczby stopni swobody $dfmin$ nieograniczonego ogólnego modelu:
 - jeśli $n < 200$, wówczas $dfmin = round(0.1 \times n)$,
 - jeśli $n \geq 200$, wówczas $dfmin = 20$.
3. Sprawdzenie częstotliwości analizowanych procesów i ustalenie:
 - bloku periodycznych zmiennych zero-jedynkowych, który będzie wykorzystany w kolejnym etapie algorytmu,

² Szerszy opis koncepcji modelowania zgodnego przedstawiają prace: Zieliński, Kufel (1995), Piłatowska (2008), Kufel, Piłatowska (1997).

- maksymalnego rzędu modeli autoregresyjnych $pmax$ szacowanych w kolejnym etapie algorytmu.

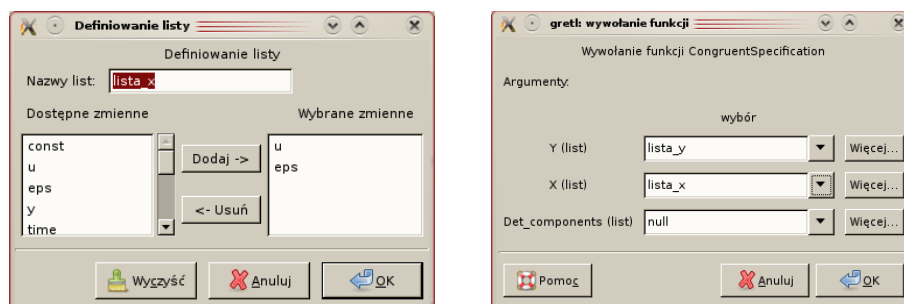
Etap 2: analiza wewnętrznej struktury wszystkich procesów:

1. Sprawdzenie, czy analizowane procesy posiadają deterministyczny komponent (trend i/lub sezonowość),
2. Sprawdzenie, czy reszty po odjęciu bloku deterministycznych komponentów wykazują integrację.
3. Określenie rzędów modeli autoregresyjnych modelowanych na podstawie wyników korelogramu dla reszt po odjęciu komponentów deterministycznych oraz różnicowaniu (w przypadku stwierdzenia pierwiastka jednostkowego), uwzględniając maksymalny rząd modeli autoregresyjnych $pmax$.

Etap 3: zbudowanie startowej specyfikacji nieograniczonego ogólnego (pełnego) modelu:

1. Wyznaczenie liczby stopni swobody modelu pełnego $dfstart$ po uwzględnieniu wszystkich składników: rozkładu opóźnień zmiennej endogenicznej Y , potencjalnych zmiennych objaśniających X , deterministycznego trendu i/lub sezonowości oraz deterministycznych zmiennych zero-jedynkowych:
 - jeżeli $dfstart < dfmin$, wówczas rzędy modeli autoregresyjnych o najwyższej wartości są zmniejszane o 1,
 - jeżeli $dfstart \geq dfmin$, wówczas startowa specyfikacja (model pełny) jest drukowana w formie skryptu wewnętrznego języka programu Gretl.

W **etapie 1** importowane są wszystkie szeregi czasowe oraz – jeśli tak wynika ze specyfikacji – zestaw dodatkowych zmiennych zero-jedynkowych do opisu charakterystycznych okresów w próbie. Na wykresie 1 po lewej stronie przedstawione jest okno dialogowe służące do wyboru procesów z załadowanego do Gretla zbioru danych, natomiast po prawej stronie widoczne jest okno dialogowe specyfikacji całego modelu (proces endogeniczny, lista potencjalnych zmiennych objaśniających oraz deterministycznych zmiennych zero-jedynkowych). Ponadto na tym etapie ustawiane są pewne dodatkowe zmienne wewnętrzne, takie jak $dfmin$ (minimalna liczba stopni swobody modelu pełnego) oraz określana jest częstotliwość obserwowania procesów.



Wykres 1. Okna dialogowe wyboru zmiennych w automatycznej procedurze budowy zgodnego dynamicznego modelu ekonometrycznego w programie Gretl

Źródło: opracowanie własne.

W **etapie 2** przeprowadzana jest analiza wewnętrznej struktury modelowanych procesów. Sprawdzane jest występowanie liniowego trendu deterministycznego oraz deterministycznej cykliczności. W zależności od częstotliwości danych uwzględniane są następujące bloki periodycznych zmiennych zero-jedynkowych:

- sezonowość dla danych miesięcznych i kwartalnych,
- sezonowość dla danych tygodniowych (52 tygodni),
- cykliczność o okresie 1 tygodnia dla danych dziennych.

W następnym kroku formułowany jest wspólny blok wszystkich komponentów deterministycznych w ten sposób, że jeżeli dany składnik (trend, blok periodycznych zmiennych zero-jedynkowych, dodatkowe zmienne zero-jedynkowe) okazał się statystycznie istotny w przynajmniej jednym procesie, to zostaje on uwzględniony w bloku wspólnych komponentów deterministycznych. Następnie w resztach – po odjęciu wspólnych składników deterministycznych – sprawdzane jest występowanie integracji za pomocą testu pierwiastka jednostkowego ADF, przy czym realizowany jest wariant testowania „w dół” na poziomie istotności $\alpha = 0.1$. Jeżeli stwierdzono występowanie pierwiastka jednostkowego proces jest różnicowany. Następnie określone są rzędy modeli autoregresyjnych dla wszystkich analizowanych procesów na podstawie wyników korelogramu w ten sposób, że rząd modelu autoregresyjnego danego procesu ustalany jest jako maksymalna istotna, na poziomie istotności $\alpha = 0.01$, wartość funkcji PACF. Korelogram wyznaczany jest do rzędu opóźnień max_lag wyznaczonego na podstawie następującej formuły:

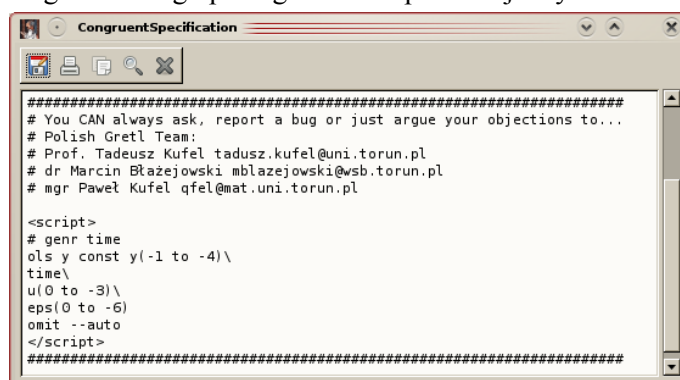
$$max_lag = round(10 \times \log_{10}(N)), \quad (2)$$

gdzie N jest liczbą obserwacji, oraz przy spełnieniu następującego dodatkowego warunku: jeżeli

$$round(10 \times \log_{10}(N)) > round(0.2 \times N), \quad (3)$$

wtedy $max_lag = round(0.2 \times N)$.

W **etapie 3** formułowana jest startowa specyfikacja modelu pełnego, uwzględniająca wszystkie wspólne komponenty deterministyczne, stopień integracji poszczególnych procesów oraz rzędy ich modeli autoregresyjnych, dla której sprawdzana jest liczba stopni swobody. W przypadku, kiedy jest ona mniejsza niż ustalona w etapie 1 minimalna liczba stopni swobody df_{min} , w kolejnych krokach maksymalne rzędy modeli autoregresyjnych redukowane są o 1 tak, aż zostanie spełniony warunek minimalnej liczby stopni swobody. Ostatecznie zostaje wydrukowany skrypt wewnętrznego języka programu Gretl, który pozwala na oszacowanie modelu pełnego. Przykładowe okno zawierające startową specyfikację nieograniczonego pełnego modelu prezentuje wykres 2.



```
#####
# You CAN always ask, report a bug or just argue your objections to...
# Polish Gretl Team:
# Prof. Tadeusz Kufel tadusz.kufel@uni.torun.pl
# dr Marcin Błazejowski mblazejowski@wsb.torun.pl
# mgr Paweł Kufel qfel@mat.uni.torun.pl

<script>
# genr time
ols y const y(-1 to -4)\
time\
u(0 to -3)\
eps(0 to -6)
omit --auto
</script>
#####
```

Wykres 2. Przykładowe okno wynikowe automatycznej procedury budowy zgodnego dynamicznego modelu ekonometrycznego w programie Gretl zawierające skrypt startowej specyfikacji nieograniczonego modelu pełnego

Źródło: opracowanie własne.

3. PORÓWNANIE MODELOWANIA ZGODNEGO Z ALGORYTMAMI PCGETS I RETINA

W poniższej części artykułu przedstawione są podobieństwa oraz różnice między algorytmem opartym na zgodnym dynamicznym modelowaniu ekonometrycznym, podejściu PcGets³ oraz RETINA. Porównanie PcGets oraz RETINA zostało przedstawione w Perez-Amaral, Gallo, White (2003, 2005). Na tej podstawie opracowano tabele 1–8. Ponadto porównanie modelowania zgodnego oraz PcGets można znaleźć w pracy Kufla (2004).

Modelowanie zgodne rozpoczyna się od początkowego ogólnego modelu, nieograniczonego, który bazuje na zestawie procesów – przyczyn i poprzez usunięcie procesów nieistotnych może być on odkryty. Takie podejście jest analogiczne do podejścia „od ogólnego do szczególnego”, jednakże formułowanie modelu początkowego jest różne.

³ Por. Krolzig, Hendry (2001), Hendry, Krolzig (2001), Doornik, Hendry (2007).

Tabela 1. Porównanie modelowania zgodnego z algorytmami PcGets i RETINA – cele

PcGets	1. Cele	
	RETINA	Modelowanie zgodne
1. Wybór oszczędnej reprezentacji nadmiernie z parametryzowanego modelu, ogólnego modelu bez restrykcji (GUM). 2. Najlepszy model w próbie 3. Model zgodny z teorią.	1. Identyfikacja oszczędnego zbioru atrybutów zapewniającego realizację prognozy poza badaną próbę.	1. Zgodny ogólny model redukowany jest do modelu zgodnego, w sensie Zielińskiego, zatem o białoszumowych resztach. 2. Dobre zachowanie w prognozie poza próbę. 3. Model zgodny z teorią.

Źródło: opracowanie własne.

Specyfikacja startowa w modelu zgodnym oparta jest na teorii oraz informacji o wewnętrznej strukturze badanych procesów, uwzględnia komponenty deterministyczne, integrację i autoregresję. Poszczególne modele autoregresyjne mogą mieć różny rząd, co jest największą różnicą w stosunku do modelowania „od ogólnego do szczególnego”.

Tabela 2. Porównanie modelowania zgodnego z algorytmami PcGets i RETINA – strategia

PcGets	2. Strategia	
	RETINA	Modelowanie zgodne
1. Od ogólnego do szczególnego	1. Od szczególnego do ogólnego. Rozpoczęcie od pojedynczej zmiennej. Dodawanie kolejnych zmiennych lub ich przekształceń, o ile poprawiają jakość prognozy poza próbę. 2. Elastyczny i oszczędny model. 3. Selektywne poszukiwanie przekształceń. 4. Kontrola współliniowości.	1. Zgodny ogólny model redukowany do zgodnego empirycznego modelu. 2. Eliminacja nieistotnych zmiennych procedurą <i>a posteriori</i> bazującej na statystyce <i>t</i> . 3. Zgodny empiryczny model jest oszczędny.

Źródło: opracowanie własne.

Tabela 2 przedstawia ogólne strategie algorytmów w poszukiwaniu ostatecznego modelu. Zgodny ogólny model jest redukowany krok po kroku do empirycznego modelu zgodnego z zachowaniem białoszumowości reszt. Eliminacja zmiennych na podstawie statystyki *t* posiada wystarczającą moc aby odkryć proces generujący dane.

Tabela 3. Porównanie modelowania zgodnego z algorytmami PcGets i RETINA – model bazowy

3. Model bazowy		
PcGets	RETINA	Modelowanie zgodne
1. GUM – specyfikacja badacza oparta na teorii. Możliwe użycie transformowanych zmiennych.	1. Oparty na oryginalnych zmiennych oraz zmiennych przekształconych.	1. Specyfikacja początkowego zgodnego modelu oparta na teorii oraz informacji o wewnętrznej strukturze procesów. 2. Model zgodny może zawierać przekształcone zmienne.

Źródło: opracowanie własne.

Zgodny model bazowy formułowany jest na dwóch płaszczyznach: teorii, zapewniającej zależność przyczynową oraz informacji o wewnętrznej strukturze badanych procesów. To gwarantuje, że całkowita zmienność zmiennej endogenicznej jest zgodna i zawarta w zmienności zmiennych egzogenicznych (reszty posiadają własności białego szumu).

Tabela 4. Porównanie modelowania zgodnego z algorytmami PcGets i RETINA – elastyczność

4. Elastyczność		
PcGets	RETINA	Modelowanie zgodne
1. Ogólny model bez restrykcji zapewnia maksymalną elastyczność.	1. Dowolne przekształcenia zmiennych zapewniają maksymalną elastyczność.	1. Zgodny ogólny model jest pozbawiony restrykcji, zatem elastyczność jest maksymalna.

Źródło: opracowanie własne.

Model zgodny zakłada liniową zależność zmiennych modelu, jednakże zmienne mogą być transformowane (np. logarytmowane) i wówczas model może posiadać zależność nieliniową, zatem model zgodny jest elastyczny.

Tabela 5. Porównanie modelowania zgodnego z algorytmami PcGets i RETINA – współliniowość

5. Współliniowość		
PcGets	RETINA	Modelowanie zgodne
1. Formułowanie modelu bez restrykcji, aby otrzymać specyfikację ortogonalną. 2. Możliwe wykorzystanie opcji w PcGets dla użytkowników nie będących ekspertami.	1. Kontrola współliniowości poprzez dodawanie do listy dodatkowych transformacji tylko gdy współliniowość jest niższa od zadanej.	1. Współliniowość jest kontrolowana w oprogramowaniu GRETL.

Źródło: opracowanie własne.

Tabela 6. Porównanie modelowania zgodnego z algorytmami PcGets i RETINA – zmienne objaśniające

6. Zmienne objaśniające		
PcGets	RETINA	Modelowanie zgodne
1. Zmienne oryginalne lub przekształcone wyspecyfikowane w GUM.	1. Dozwolone są oryginalne zmienne oraz ich nieliniowe transformacje.	1. W początkowym modelu specyfikowane są zmienne oryginalne lub transformowane

Źródło: opracowanie własne.

Tabela 7. Porównanie modelowania zgodnego z algorytmami PcGets i RETINA – liniowość

7. Liniowość		
PcGets	RETINA	Modelowanie zgodne
1. Model liniowy lub nieliniowy ze względu na parametry wyspecyfikowane w GUM.	1. Liniowość ze względu na parametry. 2. Liniowość lub nieliniowość ze względu na transformacje parametrów.	1. Liniowość ze względu na parametry. 2. Liniowość lub nieliniowość ze względu na transformacje parametrów.

Źródło: opracowanie własne.

Model zgodny zakłada liniowość ze względu na parametry, więc estymacja jest prosta i wykorzystuje klasyczną metodę najmniejszych kwadratów. Przekształcenia zmiennych są dozwolone, zatem możliwe jest otrzymanie modelu nieliniowego ze względu na zmienne.

Tabela 8. Porównanie modelowania zgodnego z algorytmami PcGets i RETINA – typy danych

8. Typy danych		
PcGets	RETINA	Modelowanie zgodne
1. Szeregi czasowe i dane przekrojowe.	1. Głównie dane przekrojowe (brak przeciwwskazań dla szeregów czasowych).	1. Szeregi czasowe.

Źródło: opracowanie własne.

Modelowanie zgodne może być wykorzystane do szeregów czasowych, jak i danych przekrojowych, jednakże automatyczna procedura zakłada wykorzystanie tylko do szeregów czasowych. Istnieje możliwość aplikacji modelowania zgodnego do modeli wielorównaniowych, włączając równania współzależne, lecz automatyczna procedura byłaby bardzo skomplikowana (ale nie niemożliwa).

4. WNIOSKI

Budowa automatycznych algorytmów estymacji modeli ekonometrycznych jest nowym nurtem rozwoju metod ekonometrycznych (Granger, Hendry, 2005). Próba budowy automatycznego algorytmu specyfikacji dla modelowania

zgodnego jest początkiem badań do pełnej implementacji w oprogramowania GRETLL.

LITERATURA

- Cottrell A., Lucchetti R. (2009), *Gnu Regression, Econometrics and Time-series*, <http://gretl.sourceforge.net>.
- Doornik J. A., Hendry D. F. (2007), *Empirical Econometric Modelling - PcGive 12*, Timberlake Consultants Press, London.
- Granger C. W. J., Hendry D. F. (2005), *A dialogue concerning a new instrument for econometric modeling*, „Econometric Theory”, vol. 21, 278–297.
- Granger C. W. J., Morris M. J. (1976), *Time Series Modelling and Interpretation*, „Journal of the Royal Statistical Society”, Series A (General), 139(2), 246–257.
- Hendry D. F., Krolzig H.-M. (2001), *Automatic Econometric Model Selection*, Timberlake Consultants Press, London.
- Krolzig H.-M., Hendry D. F. (2001), *Computer automation of general-to-specific model selection procedures*, „Journal of Economic Dynamics and Control”, 25, 831–866.
- Kufel T., Piłatowska M. (1997), *The White Noise Properties in the Dynamic Congruent Modelling*, [w:] MACROMODELS '90, 163–178, Łódź University Press, Łódź.
- Kufel T. (2004), *General-to-Specific Modelling vs. Congruent Modelling in PcGets*, [w:] Zieliński Z. (red.), *Dynamic Econometric Models*, vol. 6, 83–92, Nicolaus Copernicus University Press, Toruń.
- Perez-Amaral T., Gallo G. M., White H. (2003), *Flexible Tool for Model Building: the Relevant Transformation of the Inputs Network Approach (RETINA)*, Technical report, Universidad Complutense de Madrid, Facultad de Ciencias Económicas y Empresariales.
- Perez-Amaral T., Gallo G. M., White H. (2005), *A Comparison of Complementary Automatic Modeling Methods: RETINA and PcGets*, „Econometric Theory”, Vol. 21, No. 1, 262–277.
- Piłatowska M. (2008), *The Econometric Models Satisfying the Congruence Postulate an Overview*, [w:] Zieliński Z. (red.) *Dynamic Econometric Models*, vol. 8, 53–60, Nicolaus Copernicus University Press, Toruń.
- Talaga L., Zieliński Z. (1986), *Analiza spektralna w modelowaniu ekonometrycznym*, PWN, Warszawa.
- Zieliński Z. (1984), *Zmienność w czasie strukturalnych parametrów modelu ekonometrycznego*, „Przegląd Statystyczny”, R. XXXI, z. 1/2, 135–148.
- Zieliński Z. (1993), *Dynamic Econometric Linear Models as a Tool of Description and Analysis of Causal Relationships in Economics*, [w:] Zieliński, Z. (red.) *Some Aspects of the Dynamic Econometric Modelling*, 7–58, Nicolaus Copernicus University Press, Toruń.
- Zieliński Z. (1994), *Linear Congruent Models Describing Relationships for Integrated Economic Processes*, [w:] Zieliński Z. (red.) *Dynamic Econometric Models*, vol. 1, 7–20, Nicolaus Copernicus University Press, Toruń.
- Zieliński Z. (1995), *Liniowe modele zgodne opisujące zależności sumacyjnych (zintegrowanych) procesów ekonomicznych*, [w:] Zeliaś, A. (red.) *Przestrzenno-czasowe modelowanie i prognozowanie zjawisk gospodarczych*, 77–87, AE Kraków, Kraków.
- Zieliński Z., Kufel T. (1995), *Specification of Dynamic Properties of the Econometric Models in the Light of Congruent Models Concept*, [w:] MACROMODELS '86, 25–52, Łódź University Press, Łódź.

AUTOMATIC PROCEDURE OF CONGRUENT DYNAMIC MODEL SPECIFICATION IN GRETL

A b s t r a c t. In the last years we can observe intensive development of automatic model selection procedures. Best known are PcGets and RETINA. Such intensive work encourage to work on a new procedures. The concept of Congruent Modelling, formulated by Prof. Zygmunt Zieliński, is a very good framework for such development, including programming work, as well as many theoretical considerations. In the paper we present our concept of algorithm for automatic congruent modelling procedure and propose it's implementation in Gretl.

K e y w o r d s: congruent dynamic modelling, GUM, automatic model selection, PcGets, RETINA, GRETL.