**DYNAMICZNE MODELE EKONOMETRYCZNE**
X Ogólnopolskie Seminarium Naukowe, 4–6 września 2007 w Toruniu
Katedra Ekonometrii i Statystyki, Uniwersytet Mikołaja Kopernika w Toruniu

*Jacek Kwiatkowski*
*Uniwersytet Mikołaja Kopernika w Toruniu*

# Modele RCA w bayesowskim modelowaniu i prognozowaniu indeksu WIG20

## 1. Wprowadzenie

Modele autoregresyjne z losowymi parametrami (RCA, ang. Random Coefficient Autoregressive) są uogólnieniem klasycznych modeli autoregresyjnych. Ze względu na nieliniową postać warunkowej wariancji, wpisują się w popularny w ekonometrii finansowej nurt modeli nieliniowych. Można również dopatrzyć się związków między modelami częściowo stacjonarnymi, a modelami RCA. Modele ze stochastycznym pierwiastkiem jednostkowym (STUR), służące do opisu procesów częściowo zintegrowanych, są szczególnym przypadkiem modeli RCA.

Własności modeli RCA szczegółowo omawiają Nicholls i Quinn (1982). Ich bayesowska analiza została wstępnie przedstawiona przez Morretin i Safádi (2003). W pracy tej autorzy omawiają warunkowe gęstości *a posteriori* oraz przedstawiają pobieżnie gęstości predyktywne. Przedstawiony artykuł jest istotnym rozszerzeniem bayesowskiej analizy tej klasy modeli. Po pierwsze uwzględniono restrykcje nałożone na parametry, tak aby zapewnić stacjonarność modelu. Po drugie zaprezentowano, poprzez bayesowskie testowanie konkurencyjnych modeli, testowanie stałości parametrów autoregresyjnych. Dodatkowo, w artykule omówiono również bayesowskie prognozowanie modeli RCA.

Badania dotyczą dziennych logarytmicznych stóp zmian indeksu WIG20 w okresie od 2 stycznia 2003 do 28 lutego 2007.

Układ artykułu jest następujący: w części drugiej przedstawiono postać oraz warunki stacjonarności modelu RCA. W części trzeciej omówiono bayesowską estymację, prognozowanie i testowanie modeli RCA. Część czwarta zawiera badania empiryczne. Część piąta zawiera wnioski.

## 2. Postać i własności modelu RCA

Dla procesu $y_t$, model z losowymi współczynnikami autoregresyjnymi rzędu *p* możemy przedstawić w następującej postaci (Nicholls, Quinn, 1982):

$$y_t = \phi_0 + \sum_{i=1}^{p} (\phi_i + \beta_{it}) y_{t-i} + \varepsilon_t . \tag{1}$$

Przyjmijmy następujące założenia dotyczące modelu (1) (por. Górka, 2007):

a. $\{\varepsilon_t, t = 1, 2, ..., N\}$ jest ciągiem zmiennych losowych o niezależnych i identycznych rozkładach normalnych z zerową wartością oczekiwaną $E(\varepsilon_t) = 0$ i wariancją $E(\varepsilon_t^2) = \sigma^2$,
b. współczynniki autoregresji $\phi_i$ dla $i = 1, ..., p$ są stałe w czasie,
c. $\beta_t = (\beta_{1t}, ..., \beta_{pt})^T$ jest losowym wektorem, którego każda współrzędna jest zmienną losową o średniej zero $E(\beta_{it}) = 0$ i wariancji $E(\beta_{it}^2) = \omega_i^2$ dla $i = 1, ..., p$ i $t = 1, ..., N$,
d. zmienne losowe $\beta_{1t}, ..., \beta_{pt}$ są wzajemnie niezależne,
e. proces resztowy $\varepsilon_t$ i losowe parametry są względem siebie niezależne; $\varepsilon_t \perp \beta_{is}, t \neq s, i = 1, ..., p$.

Dla modelu (1) Morretin i Safádi (2003) wyprowadzają warunkowe gęstości *a posteriori*. W celu znalezienia brzegowych gęstości *a posteriori* i ich charakterystyk stosują algorytm Gibbsa.

Inną alternatywną postać modelu RCA przedstawił Tsay (1987). Ma ona kilka zalet. Umożliwia stosunkowo proste całkowania numeryczne metodą Monte Carlo z funkcją ważności. Dodatkowo posiada ciekawą interpretację ze względu na postać warunkowej wariancji i możliwości dokonywania porównań z powszechnie stosowanymi modelami GARCH.

Oznaczmy przez $\psi_{t-1}$ zbiór informacji dostępnych w okresie $t-1$. Model RCA (*p*) możemy przedstawić w następującej postaci (Tsay, 1987):

$$y_t = \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t , \tag{2}$$

$$\varepsilon_t \mid \psi_{t-1} \sim N(0, h_t), \tag{3}$$

$$h_t = \sigma^2 + \sum_{i=1}^{p} \omega_i^2 y_{t-i}^2 . \tag{4}$$

W modelu RCA dynamikę warunkowej wariancji $h_t$ opisuje wariancja procesu $\varepsilon_t$ oraz iloczyny wariancji parametrów $\beta_{it}$ i opóźnionych kwadratów obserwacji.

Warunki konieczne i dostateczne stacjonarności procesu RCA($p$) przedstawiają Andel (1976) oraz Nicholls i Quinn (1982). Oznaczmy przez $M$ macierz kwadratową stopnia $p$:

$$M_p = \begin{bmatrix} 0_{p-1} & I_{p-1} \\ \phi_p & \phi_{-p}^T \end{bmatrix}, \tag{5}$$

gdzie $0_{p-1}$ jest $(p-1)$ - wymiarowym wektorem zerowym, $I_{p-1}$ jest macierzą jednostkową stopnia $(p-1)$ oraz $\phi_{-p} = (\phi_{p-1}, ..., \phi_1)^T$. Proces autoregresyjny z losowymi parametrami jest stacjonarny w szerszym sensie jeżeli:

1. wartości własne macierzy $M$ leżą wewnątrz koła jednostkowego,
2. $\omega^T a < 1$, gdzie $\omega = (\omega_1^2, ..., \omega_p^2)$ jest wektorem, którego współrzędne są wariancjami losowych współczynników autoregresji $\beta_{it}$, natomiast $a$ jest ostatnią kolumną macierzy $(I - M \otimes M)^{-1}$. Symbol $\otimes$ oznacza iloczyn Kroneckera.

Przykładowo, proces RCA(1), czyli taki w którym występuje tylko jeden losowy parametr, jest stacjonarny w szerszym sensie jeżeli $\phi_1 \in (-1,1)$ oraz $\omega_1^2 + \phi_1^2 < 1$.

## 3. Bayesowska analiza modeli RCA

Modelowaniem objęto szereg $\{y_t, t = 0,1,..., N+k\}$, otrzymany według formuły $y_t = 100\ln(x_t / x_{t-1})$, gdzie $N$ oznacza liczbę obserwacji, $k$ jest horyzontem prognozy, natomiast $x_t$ oznacza dzienne notowania indeksu WIG20 w chwili $t$.

Do analizy przyjmijmy najbardziej ogólną postać modelu, mianowicie RCA(3). Pozostałe modele otrzymujemy poprzez nałożenie restrykcji na poszczególne parametry. Dzięki temu możemy testować rząd autoregresji oraz badać stałość parametrów. Model RCA(3) zakłada dla $y_t$ warunkowy (względem całej przeszłości procesu) rozkład normalny o średniej $\mu_t = \phi_0 + \sum_{i=1}^{3} \phi_i y_{t-i}$ i wariancji $h_t = \sigma^2 + \sum_{i=1}^{3} \omega_i^2 y_{t-i}^2$.

Przyjmijmy za $\theta = (\phi_0, \phi_1, \phi_2, \phi_3, \omega_1^2, \omega_2^2, \omega_3^2, \sigma^2)$ wektor nieznanych parametrów. Dodatkowo niech $y = (y_1, ..., y_N)$ oznacza obserwowany fragment szeregu czasowego do dnia $N$, natomiast $y_f = (y_{N+1}, ..., y_{N+k})$ oznacza fragment szeregu podlegający prognozowaniu. Gęstość rozkładu prawdopodobieństwa wektora obserwacji przy ustalonych parametrach i warunkach początkowych (które nie są uwzględnione w notacji) ma postać:

$$p(y, y_f \mid \theta) = \prod_{i=1}^{N+k} p(y_t \mid \psi_{t-1}, \theta) = \prod_{i=1}^{N+k} f_N(y_t \mid \mu_t, h_t), \tag{6}$$

gdzie $f_N(z \mid c, w)$ oznacza gęstość rozkładu normalnego o średniej $c$ i wariancji $w$. Przyjmujemy, że łączny rozkład *a priori* $p(\theta)$ jest iloczynem niezależnych rozkładów *a priori* jego współrzędnych.

Dla wariancji $\omega_i^2$ losowych parametrów $\beta_{it}$ oraz dla wariancji resztowej $\sigma^2$ przyjęto rozkłady wykładnicze o średniej i odchyleniu standardowym równym odpowiednio 1 i 10. Dla współczynników autoregresji założono wielowymiarowy rozkład normalny o zerowym wektorze średnich i diagonalnej macierzy kowariancji, z elementami na głównej przekątnej równymi 10. Rozkład ten określony jest na zbiorze $\Phi = (\phi_1, \phi_2, \phi_3)^T \in C_p$, gdzie $C_p$ oznacza obszar stacjonarności procesu AR(3). Dla parametru $\phi_0$ przyjęto rozkład normalny o średniej równej zero i wariancji 10. Przestrzeń parametrów dodatkowo ucięta jest przez warunek ograniczający $\omega^T a < 1$.

Tabela 1. Bayesowskie modele autoregresyjne ze stałymi i losowymi parametrami

| Model | Estymowane parametry | Model | Estymowane parametry |
|---|---|---|---|
| $M_1$ | $(\phi_0, \sigma^2)$ | $M_5$ | $(\phi_0, \phi_1, \omega_1^2, \sigma^2)$ |
| $M_2$ | $(\phi_0, \phi_1, \sigma^2)$ | $M_6$ | $(\phi_0, \phi_1, \phi_2, \omega_1^2, \omega_2^2, \sigma^2)$ |
| $M_3$ | $(\phi_0, \phi_1, \phi_2, \sigma^2)$ | $M_7$ | $(\phi_0, \phi_1, \phi_2, \phi_3, \omega_1^2, \omega_2^2, \omega_3^2, \sigma^2)$ |
| $M_4$ | $(\phi_0, \phi_1, \phi_2, \phi_3, \sigma^2)$ | | |

Przyjęte rozkłady *a priori* są rozkładami właściwymi, bardzo rozproszonymi i odzwierciedlają nieprecyzyjną wiedzę badacza na temat ich prawdziwych wartości.

Rozkład *a posteriori* wektora parametrów $\theta$ nie jest żadnym znanym rozkładem. Dlatego też jego charakterystyki, a także momenty rozkładu predyktywnego dla $(y_{t+1,\ldots,} y_{t+N})$ oraz histogramy otrzymane są za pomocą metody Monte Carlo z funkcją ważności.

W tabeli 1 przedstawiono specyfikacje wszystkich modeli opisujących dynamikę stóp zwrotu. Model $M_1$ opisuje brak jakiejkolwiek zależności autoregresyjnej. Modele $M_2, M_3, M_4$ to standardowe modele autoregresji od rzędu pierwszego, aż do trzeciego. Modele $M_5, M_6, M_7$ to modele RCA, w których struktura autoregresyjna jest opisana zarówno przez parametry losowe oraz nielosowe.

W celu porównania mocy objaśniającej konkurujących modeli, należy określić ich prawdopodobieństwa *a priori.* W naszym przypadku mamy siedem konkurujących ze sobą modeli próbkowych. Prawdopodobieństwa *a priori* definiuje się tak aby były jednakowe dla każdego modelu. Na temat możliwości i stosowalności bayesowskiego testowania modeli pisze m.in. Osiewalski (2001).

## 4. Wyniki estymacji i rozkłady predyktywne

Modele autoregresyjne z losowymi parametrami zastosowano do opisu zmienności indeksu giełdowego 20 największych spółek notowanych na warszawskiej Giełdzie Papierów Wartościowych (WIG20). Celem badania był szereg logarytmicznych stóp zmian (w procentach). Badaniem objęto szereg z okresu od 02.01.2003 do 28.02.2007. składający się z $N = 1050$ obserwacji. Prognozowaniem objęto stopy zmian obliczone na podstawie trzech kolejnych dziennych notowań w okresie od 01.03-05.03.2007. Wyniki dla modeli $M_1 - M_7$ otrzymano wykorzystując 500 tys. losowań Monte Carlo z funkcją ważności równą gęstości (maksymalnie) 8-wymiarowego rozkładu *t*-Studenta o 3 stopniach swobody. Tabela 2 zawiera wartości oczekiwane i odchylenia standardowe *a posteriori* parametrów wszystkich modeli. W modelach $M_2 - M_7$, współczynniki autoregresji szacowane są ze stosunkowo małą precyzją, co wskazuje na brak istotnej autokorelacji stóp zmian. Wniosek ten znajduje potwierdzenie w wynikach zamieszczonych w tabeli 3, która zawiera prawdopodobieństwa *a posteriori.* Najbardziej prawdopodobnym modelem jest model $M_1$, który nie uwzględnia struktury autoregresyjnej. Skupia on ponad 64 procent całej masy prawdopodobieństwa. Drugim modelem, z prawdopodobieństwem równym 0.197, jest AR(1). Trzeci model to RCA(3). Jego prawdopodobieństwo to 0.106. Czynnik Bayesa, obliczony między AR(1) i RCA(3) jest równy 1.85. Oznacza to, że – przy jednakowych prawdopodobieństwach *a priori* obu modeli – standardowy model autoregresyjny rzędu pierwszego jest prawie dwa razy bardziej prawdopodobny *a posteriori* niż model RCA(3). Stosunkowo dobry wynik modelu RCA(3) może wynikać z tego, że przy mało istotnych współczynnikach autoregresji, jego postać zbliżona jest do modelu ARCH(3). Wyniki bayesowskiego porównania są mało wrażliwe na zmianę wartości parametrów w rozkładzie *a priori*.

Tabela 2. Wartości oczekiwane i odchylenia standardowe (podane w nawiasie) parametrów w modelach AR $(M_2, M_3, M_4)$ i RCA $(M_5, M_6, M_7)$

| | Parametry | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $\phi_0$ | $\phi_1$ | $\phi_2$ | $\phi_3$ | $\omega_1^2$ | $\omega_2^2$ | $\omega_3^2$ | $\sigma^2$ |
| $M_1$ | 0.093 (0.040) | - | - | - | - | - | - | 1.716 (0.072) |
| $M_2$ | 0.086 (0.041) | 0.061 (0.031) | - | - | - | - | - | 1.711 (0.073) |
| $M_3$ | 0.086 (0.038) | 0.060 (0.030) | 0.003 (0.031) | - | - | - | - | 1.711 (0.068) |
| $M_4$ | 0.084 (0.042) | 0.060 (0.032) | 0.001 (0.030) | 0.020 (0.030) | - | - | - | 1.713 (0.069) |
| $M_5$ | 0.081 (0.040) | 0.060 (0.030) | - | - | 0.043 (0.027) | - | - | 1.654 (0.079) |
| $M_6$ | 0.090 (0.039) | 0.060 (0.034) | -0.004 (0.033) | - | 0.042 (0.027) | 0.037 (0.022) | - | 1.600 (0.088) |
| $M_7$ | 0.102 (0.102) | 0.064 (0.032) | -0.014 (0.032) | 0.025 (0.034) | 0.047 (0.029) | 0.042 (0.024) | 0.125 (0.042) | 1.393 (0.096) |

*Źródło*: obliczenia własne.

Tabela 3. Prawdopodobieństwa *a posteriori* modeli AR i RCA obliczone dla logarytmicznych stóp zmian indeksu WIG20

| Model | $M_1$ | $M_2$ | $M_3$ | $M_4$ | $M_5$ | $M_6$ | $M_7$ |
|---|---|---|---|---|---|---|---|
| Prawdopodobieństwa *a posteriori* $p(M_i \mid y)$ | 0,6437 | 0,1976 | 0,0020 | 0,0008 | 0,0491 | 0.0000 | 0,1067 |

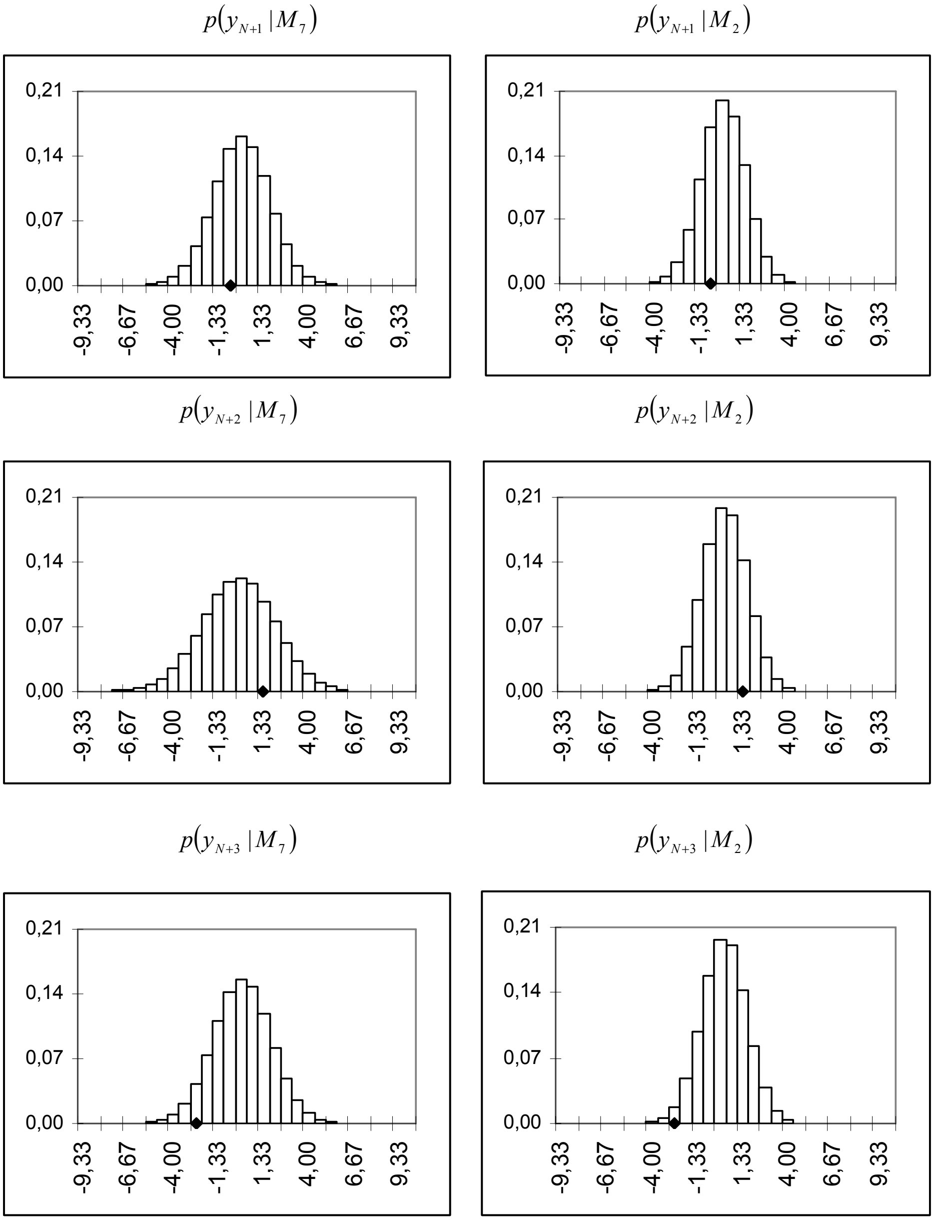
*Źródło*: obliczenia własne.

Tabela 4 zawiera wartości oczekiwane i błędy standardowe rozkładów predyktywnych stopy zmian indeksu WIG20, uzyskane w modelu RCA(3). Jak widać rozproszenie rozkładów predyktywnych rośnie wraz ze wzrostem horyzontu prognozy.

Tabela 4. Wartości oczekiwane i błędy standardowe gęstości predyktywnych w modelu RCA(3) dla logarytmicznych stóp zmian indeksu WIG20

| RCA(3) | $y_{N+1}$ | $y_{N+2}$ | $y_{N+3}$ |
|---|---|---|---|
| Wartość oczekiwana | -0.0150 | 0.0161 | 0.0323 |
| Odchylenie standardowe | 1.6272 | 2.0972 | 1.6784 |

*Źródło*: obliczenia własne.




Wykres 1. Histogramy brzegowych rozkładów predyktywnych logarytmicznych stóp zmian indeksu WIG20 obliczone w oparciu o model RCA(3) - $M_7$ i AR(1) - $M_2$

*Źródło*: opracowanie własne.

Na wykresie 1 przedstawiono histogramy oraz realizacje stopy zmian dla prognozy dokonanej na trzy okresy poza próbą, otrzymane w oparciu o model RCA(3) i AR(1). Jak widać, przy podobnej lokalizacji gęstości predyktywnych, rozproszenie jest większe w przypadku modelu RCA(3). Większe odchylenie standardowe rozkładów predyktywnych wynika z faktu przyjęcia warunkowej wariancji. Rzeczywiste wartości stóp zmian, zaznaczone są na rysunkach czarnymi punktami.

## 5. Podsumowanie

Przedstawione wyniki wskazują, że modele RCA nie wnoszą znaczących informacji do opisu szeregu WIG20. Stosunkowa duże prawdopodobieństwo *a posteriori* dla modelu RCA(3) może wynikać z faktu, że przy mało istotnych współczynnikach autoregresji jego postać jest bardzo zbliżona do klasycznego modelu ARCH(3). Ze względu na identyczną postać warunkowej średniej w modelach RCA i AR, gęstości predyktywne, uzyskane przy pomocy modeli autoregresyjnych z losowymi parametrami, charakteryzują się w porównaniu z klasycznymi modelami autoregresyjnymi większym rozproszeniem. Prezentowana publikacja zawiera analizę tylko jednego szeregu, trudno więc określić stopień przydatności modeli RCA. Uwzględnienie zależności pomiędzy losowymi procesami może jednak poprawić przydatność tego typu modeli do opisu zmienności finansowych stóp zwrotu.

## Literatura

Andel, J. (1976), Autoregressive Series with Random Parameters, Mathematische Operationsforschung und Statistics, *Series Statistics,* 7, 735–741.

Górka, J. (2007), Modele autoregresyjne z losowymi parametrami, w: Osińska M. (red.), *Procesy STUR. Modelowanie i zastosowanie do finansowych szeregów czasowych*, Wydawnictwo TNOiK, Toruń.

Morettin, P.A., Safádi, T. (2003), A Bayesian Analysis of Autoregressive Models with Random Normal Coefficients, *Journal of Statistical Computation and Simulation,* 73, 8, 563–574.

Nicholls, D.F., Quinn, B.G. (1982), *Random Coefficient Autoregressive Models: An Introduction*, Springer-Verlag, New York.

Osiewalski, J. (2001), *Ekonometria bayesowska w zastosowaniach*, Wyd. AE w Krakowie, Kraków.

Tsay, R.S. (1987), Conditional Heteroscedastic Time Series Models, *Journal of the American Statistical Association,* 82, 398.