Rozdział 5. Modele wektorowej autoregresji

MODELOWANIE POLSKIEJ GOSPODARKI z R

Wielowymiarowe modele szeregów czasowych

Modele VAR

- uwzględniają wzajemne powiązania między zmiennymi ekonomicznymi
- narzędzie statystyczne pomocne w opisie dynamicznych właściwości wielowymiarowej zmiennej
- VAR jako model ekonomiczny (SVAR)

Specyfikacja modelu VAR

Model VAR jest postaci dla wektora $y_t = [y_{1t}, y_{2t}, \ldots, y_{nt}]'$:

$$y_t = A_0 + A_1 y_{t-1} + A_2 y_{t-2} + \ldots + A_p y_{t-p} + \epsilon_t, \epsilon_t \sim N(0, \Sigma)$$

lub szybciej:

$$(I - A_1 L - \ldots - A_P L^P) y_t = A(L) y_t = A_0 + \epsilon_t$$

Liczba parametrów wynosi:

A_0: $N \times 1$

A_p: $N \times N$

Σ: $\frac{N(N+1)}{2}$

Stacjonarność modelu VAR

Ważną cechą modelu VAR jest stacjonarność, którą można zdefiniować w kategoriach wygasania wpływu szoku ϵ_t na wartość wektora y_t:

$$\lim_{k\to\infty} \frac{\partial y_{t+k}}{\partial \epsilon_t} = 0.$$

W takim przypadku istnieje długookresowa wartość μ do której powraca proces y_t:

$$\mu = A(1)^{-1} A_0,$$

Tempo powrotu do równowagi jest dane przez pierwiastki równania:

$$|A(z)| = 0,$$

Dla stacjonarnego modelu VAR pierwiastki (których liczba wynosi $P \times N$) znajdują się poza kołem jednostkowym, tj.: $\{|z_i| > 1 : i = 1, 2, \ldots, PN\}$

Stacjonarność modelu VAR

Stacjonarny model VAR ma reprezentację nieskończonej wektorowej średniej ruchomej (vector moving vaerage, VMA):

$$y_t = A(L)^{-1}(A_0 + \epsilon_t) = \mu + \sum_{i=0}^{\infty} \Theta_i \epsilon_{t-i},$$

gdzie:

$$\Theta_i = \frac{\partial y_{t+i}}{\partial \epsilon_t}.$$

Stacjonarność modelu VAR - przykład VAR(1)

Dla modelu VAR(1) postaci $y_t = A_0 + Ay_t + \epsilon_t$:

$$\Theta_k = \frac{\partial y_{t+k}}{\partial \epsilon_t} = A^k.$$

Czyli model VAR(1) jest stacjonarny jeżeli:

$$\lim_{k \to \infty} A^k = 0.$$

Postać nieskończonej średniej ruchomej jest następująca:

$$y_t = (I - A)^{-1} A_0 + \sum_{i=0}^{\infty} (A)^i \epsilon_{t-i}.$$

Stacjonarność modelu VAR - przykład VAR(1) cd

Przy liczeniu granicy ciągu A^k przydatna jest dekompozycja Jordanowska:

$$A = V\Lambda V^{-1},$$

gdzie Λ i V są macierzami zawierającymi wartości własne oraz odpowiadające im wektory własne macierzy A:

$$\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$$
$$V = [v_1, \ v_2, \ldots, v_3],$$

Zdefiniujmy $\tilde{y}_t = V^{-1} y_t$, $\tilde{A}_0 = V^{-1} A_0$ oraz $\tilde{\epsilon}_t = V^{-1}\epsilon_t$ i zapiszmy model VAR(1) jako:

$$\tilde{y}_t = \tilde{A}_0 + \Lambda \tilde{y}_{t-1} + \tilde{\epsilon}_t,$$

Macierz Λ jest diagonalna o elementach na przekątnej $\{\lambda_n : n = 1, 2, \ldots, N\}$, zaś równanie dla n-tej składowej wektora $\tilde{y}_t$ jest następujące:

$$\tilde{y}_{nt} = \tilde{\alpha}_{0n} + \lambda_n \tilde{y}_{n,t-1} + \tilde{\epsilon}_{nt}.$$

Zmienna $\tilde{y}_{nt}$ jest stacjonarna, jeżeli $|\lambda_n| < 1$. Model VAR(1) jest stacjonarny jeżeli $|\lambda_n| < 1$ dla $n = 1, 2, \ldots, N$.

Stacjonarność modelu VAR - przykład VAR(1) cd

Zauważmy, że:

$$A^k = V\Lambda^k V^{-1},$$

a zatem $\lim_{k\to\infty} \Lambda^k = 0$ implikuje, że $\lim_{k\to\infty} A^k = 0$.

Dla modelu VAR(1) wartości własne λ_n macierzy A są odwrotnością pierwiastków równania $|A(z)| = 0$,tj. $z_n = \frac{1}{\lambda_n}$

Wartości własne: $|A - \lambda I| = 0$

Pierwiastki równania: $|I - Az| = 0$

Stacjonarność modelu VAR(P)

Stacjonarność modelu VAR(P) można ustalić poprzez zapisanie go w postaci kanonicznej (tj. modelu VAR(1)):

$$\begin{bmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ \dots \\ y_{t-P+1} \end{bmatrix} = \begin{bmatrix} A_0 \\ 0 \\ 0 \\ \dots \\ 0 \end{bmatrix} + \begin{bmatrix} A_1 & A_2 & A_3 & \dots & A_P \\ I & 0 & 0 & \dots & 0 \\ 0 & I & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ y_{t-3} \\ \dots \\ y_P \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ 0 \\ 0 \\ \dots \\ 0 \end{bmatrix}.$$

oraz obliczenie $N \times P$ wartości własnych dla macierzy kanonicznej.

Estymacja modelu VAR

- Najczęściej stosowana metoda: MNK oddzielnie dla wszystkich równań
- Stosowanie MNK jest uzasadnione bo zmienne objaśniające są z góry ustalone i tym samym niezależne względem składnika losowego ϵ_t (brak problemu endogeniczności zmiennych objaśniających)

Estymacja modelu VAR

Niech $x_t = [1\ y'_{t-1}\ \dots\ y'_{t-P}]'$ oraz $A = [A_0\ A_1\ \dots\ A_P]$, zaś model VAR będzie postaci:

$$y_t = Ax_t + \epsilon_t,$$

Estymator MNK dla parametrów n-tego równania modelu VAR:

$$y_{nt} = a_n x_t + \epsilon_{nt},$$

gdzie a_n jest n-tym wierszem macierzy A, wynosi:

$$\hat{a}_n = \left(\sum_{t=1}^{T} y_{nt} x'_t\right) \left(\sum_{t=1}^{T} x_t x'_t\right)^{-1}.$$

Estymator MNK dla całej macierzy parametrów:

$$\hat{A} = \left(\sum_{t=1}^{T} y_t x'_t\right) \left(\sum_{t=1}^{T} x_t x'_t\right)^{-1}.$$

Estymacja modelu VAR

Dla $\epsilon_t \sim N(0, \Sigma)$, tj. $y_t \sim N(Ax_t, \Sigma)$, wartość funkcji wiarygodności wynosi:

$$\mathcal{L}(A, \Sigma) = (2\pi)^{-TN/2} |\Sigma|^{-T/2} \prod_{t=1}^{T} \exp\left\{ -\frac{1}{2}(y_t - Ax_t)' \Sigma^{-1} (y_t - Ax_t) \right\}.$$

Estymator MNW macierzy A pokrywa się z estymatorem MNK, zaś estymator macierzy kowariancji składnika losowego wynosi:

$$\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^{T} e_t^2,$$

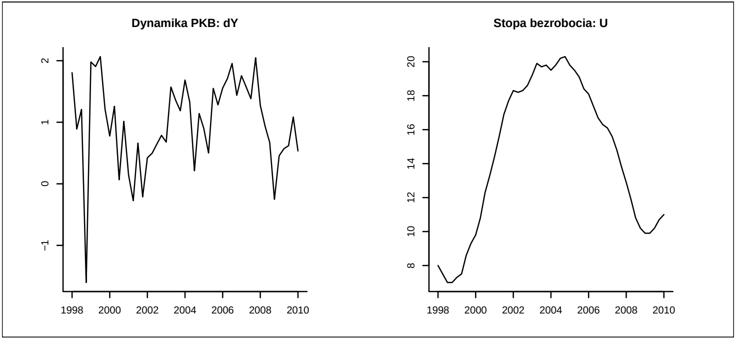
gdzie $e_t = y_t - \hat{A}x_t$.

Wartość funkcji wiarygodności w maksimum jest proporcjonalna do $|\hat{\Sigma}|$:

$$\mathcal{L}(\hat{A}, \hat{\Sigma}) = (2\pi)^{-TN/2} |\hat{\Sigma}|^{-T/2} \exp\left\{ -\frac{TN}{2} \right\}.$$

Estymacja modelu VAR - Przykład

Model VAR(2) dla tempa wzrostu PKB oraz stopy bezrobocia w Polsce



Estymacja modelu VAR - Przykład

```
Odwrotnosc modulu pierwiastków rownania charakterystycznego:
0.916  0.916 0.554 0.386

Oszacowania równania tempa wzrostu PKB (dY):
dY = dY.l1 + U.l1 + dY.l2 + U.l2 + const

      Estimate Std. Error t value Pr(>|t|)
dY.l1   0.1718     0.1525    1.13     0.27
U.l1   -0.0826     0.1672   -0.49     0.62
dY.l2   0.2097     0.1521    1.38     0.18
U.l2    0.1036     0.1656    0.63     0.53
const   0.2877     0.3981    0.72     0.47
```

Estymacja modelu VAR - Przykład

```
Oszacowania równania bezrobocia (U):
U = dY.l1 + U.l1 + dY.l2 + U.l2 + const

      Estimate Std. Error t value Pr(>|t|)
dY.l1  -0.0648     0.0746   -0.87    0.390
U.l1    1.8056     0.0818   22.08  < 2e-16 ***
dY.l2   0.0412     0.0744    0.55    0.582
U.l2   -0.8340     0.0810  -10.30  4.6e-13 ***
const   0.4645     0.1947    2.39    0.022 *


Macierz korelacji dla reszt
         dY       U
dY    1.000  -0.162
U    -0.162   1.000
```

Estymacja modelu VAR - Przykład

Postać VMA(∞)

```
     Theta0       Theta1       Theta2
dY   1.00 0.00    0.17 -0.08    0.24 -0.06
U    0.00 1.00   -0.06  1.80   -0.09  2.43

      Theta3        Theta4              Theta100
dY    0.08 -0.04    0.06 -0.01   ...    0.00 0.00
U    -0.11  2.89   -0.12  3.18   ...    0.00 0.00
```

Dobór opóźnienia modelu VAR

1. Kryteria informacyjne:

$$AIC = \ln(|\hat{\Sigma}|) + 2\frac{K}{T}$$

$$BIC = \ln(|\hat{\Sigma}|) + \frac{K}{T}\ln T$$

$$HQIC = \ln(|\hat{\Sigma}|) + 2\frac{K}{T}\ln(\ln T)$$

   gdzie $K = N(1 + NP)$ określa liczbę parametrów.
2. Testy na autokorelacje oraz istotność opóźnień

Dobór opóźnienia modelu VAR

```
Wartosci kryteriow informacyjnych
        P = 1    P = 2    P = 3    P = 4     Wybor

AIC(n) -1.815  -2.976  -3.017  -2.947    P = 3
HQ(n)  -1.725  -2.827  -2.807  -2.677    P = 2
SC(n)  -1.574  -2.574  -2.455  -2.224    P = 2
```

Test ilorazu wiarygodności na istotność opóźnień

Zespół hipotez:

$$H_0 : A_P = 0$$
$$H_1 : A_P \neq 0.$$

Statystyka testu ilorazu funkcji wiarygodności:

$$LR = T(\ln|\hat{\Sigma}_r| - \ln|\hat{\Sigma}_u|),$$

ma asymptotyczny rozkład $\chi^2(m)$, gdzie $m = N^2$ jest równe liczbie restrykcji zerowych.

Testy na autokorelację reszt

1. Autokorelacja:
   - estymator MNK jest obciążony
   - oznaka złej specyfikacji modelu VAR (zbyt niskie opóźnienie)
2. Podstawowe testy
   - Ljung-Box
   - Breusch-Godfrey

Test Ljunga-Boxa

Zespół hipotez:

$$H_0 : \forall_{1 \le j \le J} \; \Gamma_j = 0$$
$$H_1 : \exists_{1 \le j \le J} \; \Gamma_j \neq 0,$$

gdzie $\Gamma_j = Cov(\epsilon_t, \epsilon_{t-j})$ jest macierzą autokowariancji rzędu j dla składnika losowego modelu VAR.
Statystyka:

$$LB = T^2 \sum_{j=1}^{J} \frac{1}{T-j} tr(\hat{\Gamma}_j' \hat{\Gamma}_0^{-1} \hat{\Gamma}_i \hat{\Gamma}_0^{-1}) \tag{1}$$

gdzie $\hat{\Gamma}_s = \frac{1}{T} \sum_{t=s+1}^{T} y_t y_{t-s}'$ przy prawdziwości H_0 ma rozkład χ^2 o $N^2 \times (J - P)$ stopniach swobody.

Test Ljunga-Boxa - Przykład

Dla modelu VAR(2)

Wyniki testu Ljunga-Boxa

J	stat.	df	prawd.
3	12.1	4	0.017
4	14.9	8	0.061
5	16.7	12	0.156
6	23.2	16	0.107
7	24.6	20	0.218
8	30.1	24	0.183

Test Breuscha-Godfrey

Test Breuscha-Godfrey'a = wielowymiarowy test mnożników Lagrange'a.

Dla regresji postaci:

$$e_t = F_0 + F_1 y_{t-1} + \ldots + F_P y_{t-p} + G_1 e_{t-1} + \ldots + G_J e_{t-J} + \nu_t,$$

weryfikowany jest zespół hipotez:

$$\begin{aligned} H_0 :& \forall_{1 \leq j \leq J} \; G_j = 0 \\ H_1 :& \exists_{1 \leq j \leq J} \; G_j \neq 0. \end{aligned}$$

Statystyka testu:

$$BG = T(N - tr(\hat{\Omega}_r^{-1} \hat{\Omega}_u)),$$

gdzie $\hat{\Omega}_r$ i $\hat{\Omega}_u$ oznaczają oszacowanie macierzy kowariancji składnika losowego dla modelu z restykacjami oraz bez restrykcji, przy prawdziwości hipotezy zerowej ma rozkład χ^2 o JN^2 stopniach swobody.

Test Breusch-Godfrey'a - Przykład

Dla modelu VAR(2)

`Breusch-Godfrey'a`

J	stat.	df1	prawd.
1	8.09	4	0.088
2	10.10	8	0.258
3	18.57	12	0.099
4	20.53	16	0.197

Normalność rozkładu składnika losowego

Zespół hipotez testu Dornika-Hansena (wielowymiarowe uogólnienie testu Jarque-Bera):

$$H_0 : \epsilon_t \sim N(0, \Sigma)$$
$$H_1 : \epsilon_t \nsim N(0, \Sigma)$$

Dla standaryzowanych reszty modelu VAR: $u_t = \hat{\Sigma}^{-1/2} e_t$, liczone są skośność s_n i kurtoza k_n dla $n = 1, 2, \ldots, N$. Statystka testu:

$$DH = \frac{T}{6} \sum_{n=1}^{N} s_n^2 + \frac{T}{24} \sum_{n=1}^{N} (k_n - 3)^2$$

przy prawdziwości H_0 ma rozkład χ^2 o $2N$ stopniach swobody.

Stabilność modelu VAR

Metody testowania stabilności modelu VAR:

1. rekursywna estymacja parametrów
2. formalne testy na zmianę postaci strukturalnej, np. wielowymiarowe testy Chowa
3. analizie błędów prognozy *ex-post* o horyzoncie jednego okresu

Stabilność modelu VAR

Etapy testu Chowa (typu *break-point*):

1. ustal obserwację T_1 zmiany strukturalnej modelu
2. oszacuj model VAR na podstawie dwóch podprób oraz pełnej próby:
   model 1: $t \in \{1, 2, \ldots, T_1\}$, reszty $e_{t(1)}$, macierz $\hat{\Sigma}_{(1)} = T_1^{-1} \sum_{t=1}^{T_1} e_{(1)} e'_{(1)t}$

   model 2: $t \in \{T_1 + 1, \ldots, T\}$, reszty $e_{t(2)}$, macierz $\hat{\Sigma}_{(2)} = T_2^{-1} \sum_{t=1}^{T_2} e_{(2)t} e'_{(2)}$

   model 3: $t \in \{1, 2, \ldots, T\}$, reszty e_t, macierz $\hat{\Sigma} = T^{-1} \sum_{t=1}^{T} e_t e'_t$
3. oblicz statystykę:

$$LR_{BP} = T \ln |\hat{\Sigma}| - (T_1 \ln |\hat{\Sigma}_{(1)}| + T_2 \ln |\hat{\Sigma}_{(2)}|)$$

która przy prawdziwości H_0 o stabilności modelu ma rozkład χ^2 o $N(PN + 1) + \frac{(N+1)N}{2}$ stopniach swobody (całkowita liczba parametrów modelu VAR, wliczając parametry macierzy Σ).

Stabilność modelu VAR

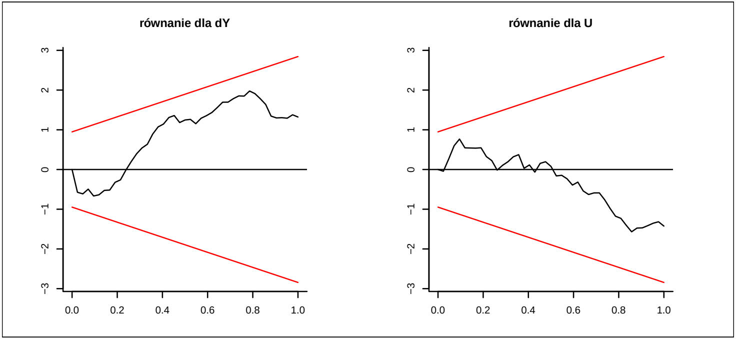
Procedura dla testowania stabilności na podstawie rekursywnych błędów prognozy o horyzoncie jednego okresu:

1. Ustalić moment początkowy liczenia rekursywanych reszt: T_1
2. Oszacować parametry $T - T1$ modeli VAR na podstawie obserwacji $t = 1, 2, \ldots, T^*$, gdzie $T^* = T_1, T_1 + 1, \ldots, T - 1$
3. Obliczyć prognozę punktową na okres $T^* + 1$
4. Obliczyć błąd prognozy na jeden okres do przodu oraz oraz średni błąd prognozy *ex-ante* ($e^P_{i\tau}$ oraz $S^P_{n\tau}$)
5. Skonstruować szereg oraz jego przedział ufności:

$$CUMSUM_{nT^\diamond} = \sum_{\tau=T_1+1}^{T^\diamond} \frac{e^P_{n\tau}}{S^P_{n\tau}}.$$

6. Jeżeli szereg znajduje sie w przedziale ufności: brak podstaw do odrzucenia hipotezy o stabilności modelu

Stabilność modelu VAR



Prognozowanie z modelem VAR

Dla modelu VAR(P):

$$y_t = A_0 + A_1 y_{t-1} + \ldots + A_P y_{t-P} + \epsilon_t.$$

Prognoza punktowa:

$$\begin{aligned}
y^f_{T+1} =& E(y_{T+1}) = A_0 + A_1 y_T + \ldots + A_P y_{T+1-P} \\
y^f_{T+2} =& A_0 + A_1 y^f_{T+1} + \ldots + A_P y_{T+2-P} \\
y^f_{T+3} =& A_0 + A_1 y^f_{T+2} + \ldots + A_P y_{T+3-P} \\
& \ldots \\
y^f_{T+k} =& A_0 + A_1 y^f_{T+k-1} + \ldots + A_P y^f_{T+k-P} \text{ dla } k > P
\end{aligned}$$

Prognozowanie z modelem VAR

Postać nieskończonej średniej ruchomej:

$$y_t = \mu + \sum_{i=0}^{\infty} \Theta_i \epsilon_{t-i}$$

Błąd prognozy wynikający z występowania składnika losowego wynosi:

$$\begin{aligned}
y_{T+1} - y^f_{T+1} &= \epsilon_{T+1} \\
y_{T+2} - y^f_{T+2} &= \epsilon_{T+2} + \Theta_1 \epsilon_{T+1} \\
&\vdots \\
y_{T+k} - y^f_{T+k} &= \sum_{i=0}^{k-1} \Theta_i \epsilon_{T+k-i}
\end{aligned}$$

Wariancja prognozy na k okresów:

$$\Sigma_k = Var(y_{T+k} - y^f_{T+k}) = \sum_{i=0}^{k-1} \Theta_i \Sigma \Theta_i'$$

Prognoza przedziałowa dla n-tej zmiennej:

$$P(y_{n,T+k} \in [y^f_{n,T+k} - c_\alpha \sqrt{\sigma_{k(nn)}}; y^f_{T+k} + c_\alpha \sqrt{\Sigma_{k(nn)}}]) = 1 - \alpha,$$

gdzie c_α jest wartością krytyczną rozkładu normalnego dla poziomu istotności 5%.

Prognozowanie z modelem VAR