Weryfikacja modelu ekonometrycznego – teoria

1. Testowanie indywidualnej istotności parametrów strukturalnych – test t-Studenta

Hipotezy: $H_0: \quad \beta_i = 0 \quad H_1: \quad \beta_i \neq 0$

Statystyka testowa: $t_i = \dfrac{\hat{\beta}_i}{\hat{\sigma}\left(\hat{\beta}_i\right)}$ Wartość krytyczna: $t\left(\frac{\alpha}{2}, T-K-1\right)$

Jeśli zachodzi nierówność $|t_i| \geq t\left(\frac{\alpha}{2}, T-K-1\right)$, to odrzucamy H_0 na rzecz H_1.

Hipoteza zerowa H_0: Parametr β_i nieistotnie różni się od zera, tj. zmienna objaśniająca x_i statystycznie nieistotnie wpływa na zmienną objaśnianą y.

Hipoteza alternatywna H_1: Parametr β_i istotnie różni się od zera, tj. zmienna objaśniająca x_i statystycznie istotnie wpływa na zmienną objaśnianą y.

2. Przedziały ufności dla parametrów strukturalnych

$$P\left\{\hat{\beta}_i - t_{\frac{\alpha}{2},T-K-1}\hat{\sigma}\left(\hat{\beta}_i\right) \leq \beta_i \leq \hat{\beta}_i - t_{\frac{\alpha}{2},T-K-1}\hat{\sigma}\left(\hat{\beta}_i\right)\right\} = 1 - \alpha$$

Z prawdopodobieństwem $1 - \alpha$ rzeczywista wartość parametru β_i zawiera się przedziale $\langle \hat{\beta}_i - t_{\frac{\alpha}{2},T-K-1}\hat{\sigma}\left(\hat{\beta}_i\right), \hat{\beta}_i - t_{\frac{\alpha}{2},T-K-1}\hat{\sigma}\left(\hat{\beta}_i\right) \rangle$.

3. Testowanie łącznej istotności parametrów strukturalnych – test Fishera- Snedecora

Hipotezy:

$$H_0: \quad \beta_1 = 0 \wedge \beta_2 = 0 \wedge \ldots \wedge \beta_K = 0$$

$$H_1: \quad \beta_1 \neq 0 \vee \beta_2 \neq 0 \vee \ldots \vee \beta_K \neq 0$$

Statystyka testowa:

$$F = \frac{\dfrac{R^2}{K}}{\dfrac{1-R^2}{T-K-1}} = \frac{R^2}{1-R^2} \cdot \frac{T-K-1}{K}$$

Wartość krytyczna: $F_{T-K-1}^{K}(\alpha)$

Jeśli zachodzi nierówność $F \geq F_{T-K-1}^{K}(\alpha)$, to odrzucamy H_0 na rzecz H_1.

4. Miary dopasowania

Wariancja resztowa	$\hat{\sigma}_{\xi}^2 = \dfrac{\sum_{t=1}^{T}(y_t - \hat{y}_t)^2}{T-K-1} = \dfrac{\sum_{t=1}^{T}\hat{\xi}_t^2}{T-K-1}$	
Średni błąd resztowy	$\hat{\sigma}_{\xi} = \sqrt{\hat{\sigma}_{\xi}^2}$	Interpretacja: Przeciętna wartość zmiennej objaśnianej różni się od wartości teoretycznej średnio o $\pm\hat{\sigma}_{\xi}$.
Współczynnik zmienności losowej	$V = \dfrac{\hat{\sigma}_{\xi}}{\bar{y}} \cdot 100\%$	Interpretacja: Udział średniego błędu resztowego w średniej wartości zmiennej objaśnianej wynosi V.
Współczynnik determinacji	$R^2 = \dfrac{\sum_{t=1}^{T}(\hat{y}_t - \bar{y})^2}{\sum_{t=1}^{T}(y_t - \bar{y})^2}$	Interpretacja: Zmienność zmiennej objaśnianej kształtuje się w R^2 pod wpływem zmienności zmiennych objaśniających modelu.

Współczynnik indeterminacji	$\phi^2 = \dfrac{\sum_{t=1}^{T}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{T}(y_t - \bar{y})^2}$ $R^2 + \phi^2 = 1$	Interpretacja: Zmienność zmiennej objaśnianej kształtuje się w ϕ^2 pod wpływem zmienności zmiennych nieujętych modelu.
Skorygowany współczynnik determinacji	$\bar{R}^2 = R^2 - \dfrac{K}{T-K-1}(1-R^2)$	Interpretacja: Zmienność zmiennej objaśnianej kształtuje się w $\bar{R}^2$ pod wpływem zmienności zmiennych objaśniających modelu, po uwzględnieniu liczby stopni swobody
Skorygowany współczynnik indeterminacji	$\bar{\phi}^2 = \dfrac{T-1}{T-K-1}\phi^2$ $\bar{R}^2 + \bar{\phi}^2 = 1$	Interpretacja: Zmienność zmiennej objaśnianej kształtuje się w $\bar{\phi}^2$ pod wpływem zmienności zmiennych nieujętych modelu, po uwzględnieniu liczby stopni swobody.

4. Test stałości wariancji składników zakłócających – test White'a

H_0: składniki zakłócające mają stałe wariancje;

H_A:, wariancja składników zakłócających zmienia się w czasie.

$$\begin{cases} H_0: \; E\xi_t^2 = \sigma_\xi^2 = const. \\ H_A: \; E\xi_t^2 \neq const. \end{cases}$$

Model podstawowy: $y_t = \beta_0 + \beta_1 x_{t1} + ... + \beta_K x_{tK} + \xi_t; \; (t = 1,...,T)$.

Model pomocniczy:

$$\hat{\xi}_t^2 = \gamma_0 + \gamma_1 x_{t1} + ... + \gamma_K x_{tK} + \gamma_1^* x_{t1}^2 + ... + \gamma_k^* x_{tK}^2 + \gamma_{12} x_{t1} x_{t2} + ... + \gamma_{K-1,K} x_{t,K-1} x_{tK} + \varepsilon_t,$$

gdzie: $\hat{\xi}_t = y_t - \hat{y}_t$

Oznaczmy liczbę parametrów w relacji pomocniczej jako $N > K$. Statystyką o asymptotycznym rozkładzie χ^2 jest:

$$\chi_N^2(\gamma) = TR^2 \overset{asymp.}{\sim} \chi_N^2$$

Reguły podejmowania decyzji:

- jeżeli $\chi_N^2(\gamma) \leq \chi_N^2(\alpha)$, to nie ma podstaw do odrzucenia hipotezy zerowej, że współczynniki strukturalne w relacji pomocniczej są równe zeru; możemy powiedzieć, że nie mamy podstaw do odrzucenia hipotezy, iż wariancja składników zakłócających jest stała w czasie,
- jeżeli $\chi_N^2(\gamma) > \chi_N^2(\alpha)$, hipoteza zerowa jest odrzucana na rzecz hipotezy, że współczynniki strukturalne w relacji pomocniczej istotnie różnią się od zera, zatem wariancja składników zakłócających zmienia się wraz ze zmianami poziomów zmiennej endogenicznej.

Program GRETL wyznacza wartość statystyki testu White'a wraz odpowiadającym tej wartości poziomem $[p]$ - minimalnego ryzyka koniecznego do przyjęcia, by odrzucić hipotezę zerową.

5. Badanie autokorelacji rzędu 1-go składnika losowego

Test Durbina-Watsona

Hipotezy: $H_0: \; \rho_1 = 0$ lub $H_0: \; \rho_1 = 0$

$H_1: \; \rho_1 > 0$ $H_1: \; \rho_1 < 0$

Statystyka testowa: $$DW = \frac{\sum_{t=2}^{T}\left(\hat{\xi}_t - \hat{\xi}_{t-1}\right)^2}{\sum_{t=1}^{T}\hat{\xi}_t^2}$$

Statystyka DW może przyjmować wartości z przedziału $(0,4)$. Ponadto $DW \approx 2\cdot(1-\hat{\rho}_1)$.

Wartości krytyczne d_L oraz d_U odczytujemy z tablic rozkładu DW, które zależą od liczby obserwacji i od liczby zmiennych objaśniających modelu.

Możemy rozpatrywać dwa przypadki:

- Jeśli $DW \in (0,2)$, to podejrzewamy dodatnie skorelowanie składników losowych $(\hat{\rho}_1 > 0)$.
  Wtedy: jeśli $DW < d_L$, to odrzucamy H_0 na rzecz H_1,

  jeśli $DW > d_U$, to brak jest podstaw do odrzucenia H_0 na rzecz H_1,

  jeśli $DW \in \langle d_L, d_U \rangle$, to test nie rozstrzyga.

- Jeśli $DW \in (2,4)$, to podejrzewamy ujemne skorelowanie składników losowych $(\hat{\rho}_1 < 0)$.
  Obliczamy $DW^* = 4 - DW$. Wtedy:

  jeśli $DW^* < d_L$, to odrzucamy H_0 na rzecz H_1,

  jeśli $DW^* > d_U$, to brak jest podstaw do odrzucenia H_0 na rzecz H_1,

  jeśli $DW^* \in \langle d_L, d_U \rangle$, to test nie rozstrzyga.

Uwagi: Statystyka DW nie ma zastosowania w przypadku modeli dynamicznych, w których występują opóźnione zmienne endogeniczne.

Test *h*-Durbina

$$y_t = \beta_0 + \beta_1 x_{t1} + \ldots + \beta_K x_{tK} + \gamma_1 y_{t-1} + \xi_t$$

Hipotezy: $H_0: \ \rho_1 = 0$

$H_1: \ \rho_1 \neq 0$

Statystyka testowa: $$h = \hat{\rho}_1 \cdot \sqrt{\frac{T}{1 - T\hat{\sigma}^2(\hat{\gamma}_1)}}$$

Gdzie:

$\hat{\rho}_1$ – współczynnik autokorelacji reszt MNK,

$\hat{\sigma}^2(\hat{\gamma}_1)$ – empiryczna wariancja błędu estymacji parametru γ_1,

T – liczebność próby.

Wartość krytyczna:

z_α – dystrybuanta standaryzowanego rozkładu normalnego.

Jeśli zachodzi nierówność: $|h| > z_\alpha$, to odrzucamy H_0 na rzecz H_1.

6. Test normalności rozkładu składnika losowego – test Jarque-Bery, test Doornika-Hansena.

$$\begin{cases} H_0: \ \xi_t \sim N(0,\sigma_\xi^2) \\ H_A: \ \xi_t \nsim N(0,\sigma_\xi^2) \end{cases}$$

Statystykami, którą wykorzystywać będziemy do sprawdzenia tej hipotezy są:

- statystyka Jarque-Bery (JB)[1],
- statystyka Doornika-Hansena (DH)[2].

$$JB = T\{\frac{1}{6}c_1^2 + \frac{1}{24}(c_2 - 3)^2\},$$

gdzie: $c_1 = \frac{m_3}{m_2^{(3/2)}}$ - współczynnik skośności, $c_2 = \frac{m_4}{m_2^2}$ - współczynnik kurtozy, $m_i = \frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t^i$, $(i = 1,2,3,4)$ - próbkowy moment centralny reszt z oszacowania MNK.

$$DH = z_1^2 + z_2^2$$

gdzie: z_1 oraz z_2 są odpowiednio: transformowanym współczynnikiem skośności oraz transformowanym współczynnikiem kurtozy3.

Jeśli składniki losowe mają rozkłady normalne, wtedy statystyki JB oraz DH mają rozkłady $\chi^2(2)$.

Reguła podejmowania decyzji jest następująca:

- jeżeli wartości statystyk JB lub DH nie większe od $\chi_\alpha^2(2)$, to nie ma podstaw do odrzucenia hipotezy zerowej, że składniki zakłócające modelu mają rozkłady normalne,
- jeżeli wartość statystyk JB lub DH są większe od $\chi_\alpha^2(2)$, to hipotezę zerową odrzucamy na korzyść alternatywnej, że składniki zakłócające modelu nie mają rozkładów normalnych.

Program GRETL oblicza wartość statystyki Doornika-Hansena wraz z odpowiadającym tej wartości poziomem $[p]$. Jeśli chcemy otrzymać wartość statystyki Jarque-Bery powinniśmy zapisać reszty z oszacowania modelu w arkuszu kalkulacyjnym programu GRETL i następnie skorzystać z zakładki *„zmienna''* $\rightarrow$ *„testy normalności rozkładu''*.

7. Test specyfikacji Ramsey'a – RESET

W Gretlu po oszacowaniu modelu skorzystaj z zakładki *Testy/test nieliniowości(kwadraty)*, *Testy/test nieliniowości(logarytmy)* oraz *Testy/test specyfikacji Ramsey'a RESET*. W zależności od wyboru testu i jego wariantu równanie pomocnicze przyjmuje różne postacie analityczne. Poniżej zapisano wersję testu RESET, w której w równaniu pomocniczym znajdują się kwadraty i sześciany zmiennej objaśnianej.

$$y_t = \beta_0 + \beta_1 x_{t1} + \ldots + \beta_K x_{tK} + \xi_t; (t = 1,\ldots,T)$$

$$y_t = \beta_0 + \beta_1 x_{t1} + \ldots + \beta_K x_{tK} + \gamma_1 \hat{y}^2{}_t + \gamma_2 \hat{y}^3{}_t + \eta_t; (t = 1,\ldots,T)$$

$H_0 : \gamma_1 + \gamma_2 = 0$ (Postać analityczna modelu jest poprawna)

$H_A : \gamma_1 + \gamma_2 \neq 0$ (Postać analityczna modelu jest niepoprawna)

$$F = \frac{\left(\sum_{t=1}^{T}\xi_t^2 - \sum_{t=1}^{T}\eta_t^2\right)/2}{\sum_{t=1}^{T}\eta_t^2/(T-k-3)} \qquad F \sim F(2, T-k-3)$$

Jeżeli $F \geq F(2, T-k-3)$ to odrzucamy hipotezę zerową na rzecz hipotezy alternatywnej.

---

[1] C. M. Jarque, A. K. Bera, Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals, Economic Letters, vol. 6, 1980, str. 255-259.

[2] J. A. Doornik, H. Hansen, An Omnibus Test for Univariate and Multivariate Normality, Nuffield College, Oxford, 1994.

[3] Z uwagi na dość skomplikowany zapis tych transformacji nie podajemy ich w tekście.