ΧΡΟΝΙΚΕΣ ΣΕΙΡΕΣ

7ο μάθημα: Πολυμεταβλητή παλινδρόμηση (ΕΠΑΝΑΛΗΨΗ)

Παπάνα Αγγελική

Μεταδιδακτορική ερευνήτρια, ΑΠΘ & ΠΑΜΑΚ
E-mail: angeliki.papana@gmail.com, agpapana@auth.gr
Webpage: http://users.auth.gr/agpapana

Εισαγωγή

Τα υποδείγματα πολυμεταβλητής παλινδρόμησης του πληθυσμού είναι επεκτάσεις των υποδειγμάτων απλής παλινδρόμησης, όπου η εξαρτημένη μεταβλητή Y επηρεάζεται από ένα σύνολο ερμηνευτικών (ανεξάρτητων) μεταβλητών $\mathbf{X}_1, \mathbf{X}_2, .., \mathbf{X}_k$.

Η γραμμική μορφή του πολυμεταβλητού υποδείγματος του πληθυσμού είναι:

$$\mathbf{Y}_t = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \mathbf{X}_{1t} + \boldsymbol{\beta}_2 \mathbf{X}_{2t} + \cdots + \boldsymbol{\beta}_k \mathbf{X}_{kt} + \mathbf{e}_t$$

- Για $k = 1$, προκύπτει το απλό γραμμικό υπόδειγμα παλινδρόμησης.
- Οι μέθοδοι εκτίμησης και ελέγχου του πολλαπλού γραμμικού υποδείγματος είναι μια γενίκευση της απλής περίπτωσης για $k = 1$.

Το γραμμικό υπόδειγμα παλινδρόμησης με τρεις μεταβλητές

Αν στο απλό γραμμικό υπόδειγμα προσθέσουμε μια επιπλέον ανεξάρτητη μεταβλητή, τότε προκύπτει το πολλαπλό γραμμικό υπόδειγμα παλινδρόμησης του πληθυσμού με τρεις μεταβλητές, που είναι:

$$\mathbf{Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + e_t}$$

όπου $\mathbf{Y_t}$ είναι η εξαρτημένη μεταβλητή, $\mathbf{X_{1t}, X_{2t}}$ είναι οι ανεξάρτητες μεταβλητές, $\mathbf{e_t}$ ο διαταρακτικός όρος και $\boldsymbol{\beta_0, \beta_1, \beta_2}$ οι συντελεστές της πολλαπλής γραμμικής παλινδρόμησης του πληθυσμού.

- **Ο συντελεστής $\boldsymbol{\beta_0}$ είναι η σταθερά και προσδιορίζει το σημείο τομής του άξονα των $\mathbf{Y}$ και του επιπέδου της παλινδρόμησης του πληθυσμού.**
- **Οι συντελεστές $\boldsymbol{\beta_1, \beta_2}$ λέγονται συντελεστές μερικής παλινδρόμησης, επειδή μετρούν μερικές μεταβολές του $\mathbf{E(Y_t | X_{1t}, X_{2t})}$.**

Έστω $\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2$ είναι οι εκτιμήσεις των $\beta_0, \beta_1, \beta_2$ που προέρχονται από ένα δείγμα που πάρθηκε τυχαία από έναν πληθυσμό. Τότε, η συνάρτηση του πολλαπλού γραμμικού υποδείγματος της παλινδρόμησης του δείγματος είναι:

$$\hat{Y}_t = \widehat{\beta}_0 + \widehat{\beta}_1 X_{1t} + \widehat{\beta}_2 X_{2t} + u_t$$

όπου u_t είναι η εκτίμηση του διαταρακτικού όρου.

- **Οι συντελεστές $\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2$ λέγονται μερικοί συντελεστές της παλινδρόμησης (partial regression coefficients) και έχουν τη μικρότερη διακύμανση από κάθε άλλον αμερόληπτο εκτιμητή.**
- **Ο εκτιμητής $\widehat{\beta}_1$ μετράει την μεταβολή της εξαρτημένης μεταβολής Y_t λόγω μιας μεταβολής της ανεξάρτητης μεταβλητής X_{1t}, εφόσον η ανεξάρτητη μεταβλητή X_{2t} παραμένει σταθερή.**

Περιγραφή του υποδείγματος τριών μεταβλητών με μήτρες

Η σχέση $\mathbf{Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + e_t}$ μπορεί να γραφεί ως:

$$\mathbf{Y_t = \beta_0 X_{0t} + \beta_1 X_{1t} + \beta_2 X_{2t} + e_t}$$

όπου $\mathbf{X_{0t} = 1}$ για κάθε $\boldsymbol{t} = \mathbf{1, .., T}$.

Για δείγμα $\mathbf{T}$ παρατηρήσεων, θα έχουμε το παρακάτω σύστημα εξισώσεων:

$$\mathbf{Y_1 = \beta_0 X_{01} + \beta_1 X_{11} + \beta_2 X_{21} + e_1}$$

$$\mathbf{Y_2 = \beta_0 X_{02} + \beta_1 X_{12} + \beta_2 X_{22} + e_2}$$

…………………………

$$\mathbf{Y_T = \beta_0 X_{0T} + \beta_1 X_{1T} + \beta_2 X_{2T} + e_T}$$

Το σύστημα αυτό μπορεί να γραφεί με μορφή μητρών ως:

$$\begin{bmatrix} \Upsilon_1 \\ \vdots \\ \Upsilon_T \end{bmatrix} = \begin{bmatrix} X_{10} & X_{11} & X_{12} \\ & \vdots & \\ X_{T0} & X_{T1} & X_{T2} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ \vdots \\ e_T \end{bmatrix}$$

ή

$$\boldsymbol{\Upsilon} = \boldsymbol{X\beta} + \boldsymbol{e}$$

όπου $\boldsymbol{\Upsilon}, \boldsymbol{X}, \boldsymbol{\beta}, \boldsymbol{e}$ είναι μήτρες.

$\boldsymbol{\Upsilon}$: **διάνυσμα** $T \times 1$

$\boldsymbol{X}$: **μήτρα** $T \times (k+1)$ **δηλ.** $T \times 3$

$\boldsymbol{\beta}$: **διάνυσμα** $(k+1) \times 1$ **δηλ.** 3×1

$\boldsymbol{e}$: **διάνυσμα** $T \times 1$

Περιγραφή του υποδείγματος k μεταβλητών με μήτρες

Αν στο προηγούμενο υπόδειγμα προσθέσουμε και άλλες ανεξάρτητες μεταβλητές, θα πάρουμε το παρακάτω πολυμεταβλητό υπόδειγμα:

$$\mathbf{Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + e_t} \quad \text{ή}$$

$$\mathbf{Y_t = \beta_0 X_{0t} + \beta_1 X_{1t} + \beta_2 X_{2t} \ldots + \beta_k X_{kt} + e_t,} \; t = 1, \ldots, \mathrm{T}.$$

Για δείγμα T παρατηρήσεων, θα έχουμε το παρακάτω σύστημα εξισώσεων:

$$\mathbf{Y_1 = \beta_0 X_{01} + \beta_1 X_{11} + \beta_2 X_{21} + \cdots + \beta_k X_{k1} + e_1}$$

$$\mathbf{Y_2 = \beta_0 X_{02} + \beta_1 X_{12} + \beta_2 X_{22} + \cdots + \beta_k X_{k2} + e_2}$$

....

$$\mathbf{Y_T = \beta_0 X_{0T} + \beta_1 X_{1T} + \beta_2 X_{2T} + \cdots + \beta_k X_{kT} + e_T}$$

Το σύστημα αυτό μπορεί να γραφεί με μορφή μητρών ως:

$$\begin{bmatrix} \Upsilon_1 \\ \vdots \\ \Upsilon_T \end{bmatrix} = \begin{bmatrix} X_{10} & X_{11} & X_{12} & \cdots & X_{1k} \\ & & \vdots & & \\ X_{T0} & X_{T1} & X_{T2} & \cdots & X_{Tk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} e_1 \\ \vdots \\ e_T \end{bmatrix}$$

ή

$$\boldsymbol{\Upsilon} = \boldsymbol{X\beta} + \boldsymbol{e}$$

όπου $\boldsymbol{\Upsilon}, \boldsymbol{X}, \boldsymbol{\beta}, \boldsymbol{e}$ είναι μήτρες.

$\boldsymbol{\Upsilon}$: διάνυσμα στήλης $\mathbf{T \times 1}$

$\boldsymbol{X}$: μήτρα $\mathbf{T \times (\boldsymbol{k} + 1)}$

$\boldsymbol{\beta}$: διάνυσμα στήλης $\mathbf{(\boldsymbol{k} + 1) \times 1}$

$\boldsymbol{e}$: διάνυσμα στήλης $\mathbf{T \times 1}$

Το $\boldsymbol{e}$ είναι το διάνυσμα στήλης $(\mathrm{T} \times 1)$ του διαταρακτικού όρου του υποδείγματος πληθυσμού.

Το $\mathrm{E}(\boldsymbol{e})$ είναι η προσδοκώμενη τιμή του διανύσματος του διαταρακτικού όρου.

Το $\mathbf{0}$ είναι το μηδενικό διάνυσμα διαστάσεων $\mathrm{T} \times 1$.

Το $\mathbf{I}$ είναι η μοναδιαία ή ταυτοτική μήτρα διαστάσεων $\mathrm{T} \times \mathrm{T}$.

Το $\boldsymbol{ee}'$ είναι μια συμμετρική μήτρα διαστάσεων $\mathrm{T} \times \mathrm{T}$:

$$\boldsymbol{ee}' = \begin{bmatrix} \boldsymbol{e}_1 \\ \vdots \\ \boldsymbol{e}_\mathrm{T} \end{bmatrix} \begin{bmatrix} \boldsymbol{e}_1 & \cdots & \boldsymbol{e}_\mathrm{T} \end{bmatrix} = \begin{bmatrix} \boldsymbol{e}_1^2 & \dots & \boldsymbol{e}_1\boldsymbol{e}_\mathrm{T} \\ \boldsymbol{e}_2\boldsymbol{e}_1 & \dots & \boldsymbol{e}_2\boldsymbol{e}_\mathrm{T} \\ \vdots & \vdots & \vdots \\ \boldsymbol{e}_\mathrm{T}\boldsymbol{e}_1 & \dots & \boldsymbol{e}_\mathrm{T}^2 \end{bmatrix}$$

Το $\mathrm{E}(\boldsymbol{ee'})$ είναι μήτρα διαστάσεων $\mathrm{T} \times \mathrm{T}$ και ονομάζεται μήτρα διακύμανσης - συνδιακύμανσης:

$$\boldsymbol{E}(\boldsymbol{ee'}) = \begin{bmatrix} \mathbf{E}\boldsymbol{e}_1^2 & \dots & \mathbf{E}\boldsymbol{e}_1\boldsymbol{e}_\mathrm{T} \\ \mathbf{E}\boldsymbol{e}_2\boldsymbol{e}_1 & \dots & \mathbf{E}\boldsymbol{e}_2\boldsymbol{e}_\mathrm{T} \\ \vdots & \vdots & \vdots \\ \mathbf{E}\boldsymbol{e}_\mathrm{T}\boldsymbol{e}_1 & \dots & \mathbf{E}\boldsymbol{e}_\mathrm{T}^2 \end{bmatrix} = \boldsymbol{\sigma}_e^2 \begin{bmatrix} \mathbf{1} & \dots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots \\ \mathbf{0} & \dots & \mathbf{1} \end{bmatrix}$$

Οι βασικές υποθέσεις του πολλαπλού γραμμικού υποδείγματος

1. Η μεταβλητή e_t (διαταρακτικός όρος υποδείγματος πολλαπλής παλινδρόμησης του πληθυσμού) είναι τυχαία μεταβλητή με μέσο το μηδέν:
$\mathrm{E}(e_t) = 0$, **για** $t = 1, \ldots, \mathrm{T}$.

2. Η διακύμανση της τυχαίας μεταβλητής $\mathrm{e_t}$ είναι σταθερή (ομοσκεδαστικός όρος):
$\mathrm{Var}(\mathrm{e_t}) = \sigma_e^2$, **για** $t = 1, \ldots, \mathrm{T}$.

3. Δεν υπάρχει αυτοσυσχέτιση (autocorrelation) στους διαταρακτικούς όρους, δηλαδή οι τιμές των διαταρακτικών όρων είναι ανεξάρτητες
$Cov(e_t, e_s) = 0$, **για** $t \neq s$.

4. Ο διαταρακτικός όρος δεν συσχετίζεται με τις ανεξάρτητων μεταβλητές X:
$Cov(X_{1t}, X_{2t}, \ldots, X_{kt}, e_t) = 0$, **για** $t = 1, \ldots, \mathrm{T}$.

5. Η τυχαία μεταβλητή $\mathrm{e_t}$ ακολουθεί την κανονική κατανομή με μέσο μηδέν και σταθερή διακύμανση: $\mathrm{e_t} \sim N(0, \sigma_e^2)$.

6. Η μαθηματική (προσδιοριστική) σχέση μεταξύ των μεταβλητών Y_t και $X_{1t}, X_{2t}, \dots, X_{kt}$ είναι γραμμική.

7. Δεν υπάρχουν γραμμικές σχέσεις μεταξύ των ερμηνευτικών μεταβλητών. Η υπόθεση αυτή αποκλείει την ύπαρξη πολυσυγγραμικότητας μεταξύ των ερμηνευτικών μεταβλητών.

8. Ο αριθμός των παρατηρήσεων του δείγματος n είναι μεγαλύτερος από τον αριθμό των ανεξάρτητων μεταβλητών $k + 1$.

9. Το υπόδειγμα της πολλαπλής γραμμικής παλινδρόμησης είναι σωστά εξειδικευμένο.

10. Οι ανεξάρτητες μεταβλητές μετρούνται χωρίς σφάλμα.

Ολοκληρωμένη εξειδίκευση του πολλαπλού γραμμικού υποδείγματος

$$\mathbf{Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + e_t}$$

- $e_t \sim N(0, \sigma^2)$
  α) e_t: τυχαία μεταβλητή
  β) $E(e_t) = 0$
  γ) $Var(e_t) = E(e_t^2) = \sigma^2$
- $Cov(e_t, e_s) = E(e_t e_s) = 0$, για $t \neq s$ (ανεξαρτησία τυχαίων όρων)
- Οι ερμηνευτικές μεταβλητές δεν είναι στοχαστικές. Οι τιμές τους παραμένουν σταθερές και δεν είναι ίσες όλες μεταξύ τους.
- Δεν υπάρχουν ακριβείς γραμμικές σχέσεις ανάμεσα στις ερμηνευτικές μεταβλητές.
- Ο αριθμός των παρατηρήσεων του δείγματος είναι μεγαλύτερος από τον αριθμό των συντελεστών του υποδείγματος.

Η μέθοδος των ελαχίστων τετραγώνων

Με τη μέθοδο ελαχίστων τετραγώνων εκτιμούμε τους συντελεστές της γραμμής παλινδρόμησης και ελαχιστοποιούμε το άθροισμα των τετραγώνων των καταλοίπων:

$$\min \sum_{t=1}^{n} e_t^2 = \min \sum_{t=1}^{n} (Y_t - \widehat{Y}_t)^2$$

$$= \min \sum_{t=1}^{n} (Y_t - \widehat{\beta}_0 - \widehat{\beta}_1 \mathrm{X}_{1t} - \widehat{\beta}_2 \mathrm{X}_{2t} - \cdots - \widehat{\beta}_k \mathrm{X}_{kt})^2$$

Για το γραμμικό υπόδειγμα παλινδρόμησης με τρεις μεταβλητές, οι εκτιμήσεις των συντελεστών $\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2$ γίνεται με τη μέθοδο ελαχίστων τετραγώνων λύνοντας το σύστημα εξισώσεων που προκύπτει αν μηδενίσουμε τις παρακάτω μερικές παραγώγους

$$\frac{\partial \sum_{t=1}^{n} e_t^2}{\partial \widehat{\beta}_0} = -2\sum_{t=1}^{n}(Y_t - \widehat{\beta}_0 - \widehat{\beta}_1 X_{1t} - \widehat{\beta}_2 X_{2t}) = 0$$

$$\frac{\partial \sum_{t=1}^{n} e_t^2}{\partial \widehat{\beta}_1} = -2\sum_{t=1}^{n} X_{1t}(Y_t - \widehat{\beta}_0 - \widehat{\beta}_1 X_{1t} - \widehat{\beta}_2 X_{2t}) = 0$$

$$\frac{\partial \sum_{t=1}^{n} e_t^2}{\partial \widehat{\beta}_2} = -2\sum_{t=1}^{n} X_{2t}(Y_t - \widehat{\beta}_0 - \widehat{\beta}_1 X_{1t} - \widehat{\beta}_2 X_{2t}) = 0$$

Λύνοντας αυτό το σύστημα εξισώσεων, προκύπτουν οι εκτιμητές των συντελεστών της γραμμής παλινδρόμησης του δείγματος

$$\widehat{\boldsymbol{\beta}}_1 = \frac{(\sum_{t=1}^{T} x_1 y)(\sum_{t=1}^{T} x_2^2) - (\sum_{t=1}^{T} x_2 y)(\sum_{t=1}^{T} x_1 x_2)}{(\sum_{t=1}^{T} x_1^2)(\sum_{t=1}^{T} x_2^2) - (\sum_{t=1}^{T} x_1 x_2)^2}$$

$$\widehat{\boldsymbol{\beta}}_2 = \frac{(\sum_{t=1}^{T} x_2 y)(\sum_{t=1}^{T} x_1^2) - (\sum_{t=1}^{T} x_1 y)(\sum_{t=1}^{T} x_1 x_2)}{(\sum_{t=1}^{T} x_1^2)(\sum_{t=1}^{T} x_2^2) - (\sum_{t=1}^{T} x_1 x_2)^2}$$

$$\widehat{\boldsymbol{\beta}}_0 = \bar{Y}_t - \widehat{\boldsymbol{\beta}}_1 \bar{X}_{1t} - \widehat{\boldsymbol{\beta}}_2 \bar{X}_{2t}$$

όπου $y = \Upsilon_t - \bar{Y}_t, x_1 = \mathrm{X}_{1t} - \bar{X}_{1t}, x_2 = \mathrm{X}_{2t} - \bar{X}_{2t}$

Οι εκτιμητές των συντελεστών της γραμμής παλινδρόμησης του δείγματος για k μεταβλητές δίνονται από την σχέση:

$$\widehat{\beta}_j = \frac{\Delta_j}{\Delta}$$

όπου $\Delta = \begin{vmatrix} \sum x_1^2 & \sum x_1 x_2 & \dots & \sum x_1 x_k \\ \sum x_1 x_2 & \sum x_2^2 & \dots & \sum x_2 x_k \\ \vdots & \vdots & \vdots & \vdots \\ \sum x_1 x_k & \sum x_2 x_k & \dots & \sum x_k^2 \end{vmatrix}$

και Δ_j η ορίζουσα που προκύπτει από την Δ, όταν η στήλη που αντιστοιχεί στο συντελεστή β_j αντικατασταθεί με τα αθροίσματα $\sum x_j y$,

π.χ. για $j = 1$

$$
\Delta_1 = \begin{vmatrix} \sum x_1 y & \sum x_1 x_2 & \cdots & \sum x_1 x_k \\ \sum x_2 y & \sum x_2^2 & \cdots & \sum x_2 x_k \\ \vdots & \vdots & \vdots & \vdots \\ \sum x_k y & \sum x_2 x_k & \cdots & \sum x_k^2 \end{vmatrix}
$$

και

$$
\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \cdots - \hat{\beta}_k \bar{X}_k
$$

Θεώρημα των GAUSS-MARKOV

Δεδομένου ότι ισχύουν οι υποθέσεις του πολλαπλού γραμμικού υποδείγματος της παλινδρόμησης, οι εκτιμητές $\widehat{\beta}_0, \widehat{\beta}_1, \ldots, \widehat{\beta}_k$ που προκύπτουν από την μέθοδο ελαχίστων τετραγώνων είναι οι καλύτεροι γραμμικοί αμερόληπτοι εκτιμητές.

Το τυπικό σφάλμα του εκτιμητή $\widehat{\beta}$ (standard error of the estimate) ή το τυπικό σφάλμα της παλινδρόμησης ισούται με την εκτίμηση της τετραγωνικής ρίζας της διακύμανσης του διαταρακτικού όρου:

$$s_e^2 = \frac{u'u}{\mathrm{T} - k}$$

Διακυμάνσεις εκτιμητών

Οι γραμμικοί αμερόληπτοι εκτιμητές των διακυμάνσεων – συνδιακυμάνσεων των εκτιμώμενων συντελεστών του υποδείγματος είναι:

$$s_{\hat{\beta}}^2 = s_e^2 (\mathbf{X}'\mathbf{X})^{-1}$$

Π.χ. Στην περίπτωση Κ=2 ανεξάρτητων τυχαίων μεταβλητών:

$$s_{\hat{\beta}_1}^2 = s_e^2 \frac{\sum x_2^2}{\Delta}$$

$$s_{\hat{\beta}_2}^2 = s_e^2 \frac{\sum x_1^2}{\Delta}$$

$$s_e^2 = \frac{\sum \hat{u}_t^2}{\mathrm{T} - 3}$$

όπου $\Delta = \begin{vmatrix} \sum x_1^2 & \sum x_1 x_2 \\ \sum x_1 x_2 & \sum x_2^2 \end{vmatrix}$

Διάστημα εμπιστοσύνης των συντελεστών του πολλαπλού γραμμικού υποδείγματος

Όπως και στο απλό γραμμικό υπόδειγμα, έτσι και στο πολλαπλό υπόδειγμα μπορούμε να κατασκευάσουμε διαστήματα εμπιστοσύνης για τους συντελεστές (παραμέτρους) του πληθυσμού με βάση τις ιδιότητες των εκτιμητών ελαχίστων τετραγώνων:

$$\hat{\beta}_j - t_{T-(k+1),\frac{a}{2}} s_{\hat{\beta}_j} < \beta_j < \hat{\beta}_j + t_{T-(k+1),\frac{a}{2}} s_{\hat{\beta}_j}, j = 0, 2, .., k+1$$

- $s_{\hat{\beta}_j}$ **το τυπικό σφάλμα του εκτιμητή** $\hat{\beta}_j$
- $t_{T-(k+1),\frac{a}{2}}$ **η κριτική τιμή (τιμή των πινάκων) από κατανομή student με** $T-(k+1)$ **βαθμούς ελευθερίας**
- a **το επίπεδο σημαντικότητας**
- k **το πλήθος των ανεξάρτητων μεταβλητών**

Έλεγχος υποθέσεως για τους συντελεστές του πολλαπλού γραμμικού υποδείγματος

H_0: $\boldsymbol{\beta_j = 0}$

H_1: $\boldsymbol{\beta_j \neq 0}$

Στατιστικό ελέγχου $\boldsymbol{t = \dfrac{\hat{\beta}_j}{s_{\hat{\beta}_j}}}$

Αν $\boldsymbol{|t| > t_{T-(k+1),\frac{a}{2}}}$**, η H_0 απορρίπτεται.**

- $\boldsymbol{t_{T-(k+1),\frac{a}{2}}}$ **η κριτική τιμή (τιμή των πινάκων) από κατανομή student με** $\boldsymbol{T-(k+1)}$ **βαθμούς ελευθερίας**
- $\boldsymbol{a}$ **το επίπεδο σημαντικότητας**
- $\boldsymbol{k}$ **το πλήθος των ανεξάρτητων μεταβλητών**

Έλεγχος στατιστικής σημαντικότητας του υποδείγματος της πολλαπλής γραμμικής παλινδρόμησης

Ο έλεγχος της στατιστικής σημαντικότητας του υποδείγματος της πολλαπλής παλινδρόμησης του πληθυσμού αποτελεί ένα από τα κυριότερα στάδια στη μεθοδολογία της οικονομετρίας.

Έστω ότι έχουμε το παρακάτω υπόδειγμα της πολλαπλής παλινδρόμησης του πληθυσμού:

$$\mathbf{Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + e_t}$$

Η υπόθεση που θέλουμε να ελέγξουμε είναι αν οι συντελεστές του παραπάνω υποδείγματος είναι ίσοι με μηδέν, δηλαδή

H_0: $\beta_1 = \beta_2 = \cdots = \beta_k = 0$

H_1: ένας τουλάχιστον συντελεστής είναι διάφορος του μηδενός

Στην υπόθεση που εξετάζουμε δεν περιλαμβάνεται ο σταθερός όρος β_0, καθώς θέλουμε να ελέγξουμε είναι αν οι ανεξάρτητες μεταβλητές του υποδείγματος μπορούν να ερμηνεύσουν την εξαρτημένη μεταβλητή Υ.

Ο έλεγχος της H_0 γίνεται με την στατιστική $\mathbf{F}$ (F-statistics):

$$\mathrm{F} = \frac{\frac{\sum_{t=1}^{\mathrm{T}}(\hat{\Upsilon} - \bar{\Upsilon})^2}{k}}{\frac{\sum_{t=1}^{\mathrm{T}}(\Upsilon_t - \hat{\Upsilon})^2}{\mathrm{T} - (k+1)}} = \frac{\frac{SSR}{k}}{\frac{SSE}{\mathrm{T} - (k+1)}} \rightarrow \mathrm{F}_{k,\mathrm{T}-(k+1),a}$$

T: μέγεθος του δείγματος

k: πλήθος ανεξάρτητων μεταβλητών

SSR: ερμηνευτικό μέρος συνολικής μεταβλητότητας της Υ από την παλινδρόμηση

SSE: ανερμήνευτο μέρος συνολικής μεταβλητότητας της Υ

$\mathrm{F}_{k,\mathrm{T}-(k+1),a}$: κριτική τιμή κατανομής F (από πίνακα)

- **Αν F > $F_{k,T-(k+1),a}$, τότε η H_0 απορρίπτεται.**

  Για δεδομένο επίπεδο σημαντικότητας α (συνήθως $\alpha = 0.05$), αν η τιμή της F είναι μεγαλύτερη από την κριτική τιμή $F_{k,T-(k+1),a}$, τότε η τιμή της F είναι υψηλή, καθώς το $SSR = \sum_{t=1}^{T}(\hat{Y} - \bar{Y})^2$ είναι μεγάλο σε σχέση με το άθροισμα των τετραγώνων των καταλοίπων $SSE = \sum_{t=1}^{T}(Y_t - \hat{Y})^2$.

- **Αν F < $F_{k,T-(k+1),a}$, τότε η H_0 δεν απορρίπτεται.**

  Δηλαδή οι ανεξάρτητες μεταβλητές δεν ερμηνεύουν επαρκώς τη μεταβολή της εξαρτημένης μεταβλητής στο υπόδειγμα πολλαπλής γραμμικής παλινδρόμησης.

  Στην περίπτωση αυτή, το SSR είναι μικρό σε σχέση με το άθροισμα των τετραγώνων των καταλοίπων SSE, και η τιμή της F είναι χαμηλή.

Σε ορισμένα υποδείγματα πολλαπλής γραμμικής παλινδρόμησης, παρατηρούμε ότι ενώ οι συντελεστές είναι στατιστικά σημαντικοί (διάφοροι του μηδενός), η τιμή της στατιστικής F είναι πολύ χαμηλή, το οποίο σημαίνει ότι στο σύνολο τους οι συντελεστές του υποδείγματος είναι μη στατιστικά σημαντικοί.

Η ύπαρξη αυτού του φαινομένου οφείλεται στο πρόβλημα της πολυσυγγραμικότητας (multicollinearity).

Παράδειγμα

Θέλουμε να εξετάσουμε αν η εξαρτημένη μεταβλητή Υ σχετίζεται γραμμικά με τις ανεξάρτητες μεταβλητές

X_1: λόγος των τιμών των εισαγόμενων καταναλωτικών τιμών προς τις εγχώριες τιμές (σχετικές τιμές) και

X_2: διαθέσιμο εισόδημα, για την περίοδο 1958-1973.

Να εκτιμηθεί η γραμμή πολλαπλής παλινδρόμησης του δείγματος, οι συντελεστές του υποδείγματος, οι προβλεφθείσες τιμές της Υ ($\widehat{\Upsilon}$) και οι προβλεφθείσες τιμές του διαταρακτικού όρου ($\hat{u}_t$), η ελαστικότητα ως προς τις σχετικές τιμές και ως προς το εισόδημα.

Τα στοιχεία του παραδείγματος δίνονται στον παρακάτω πίνακα:

Έτος	Αξία εισαγωγών καταναλωτικών αγαθών (δισ. Δρχ.)	Σχετικές τιμές	Διαθέσιμο εισόδημα
1958	5,121	0,940	105,508
1959	4,134	0,967	107,497
1960	4,653	0,904	111,875
1961	5,622	0,896	124,676
1962	5,499	0,845	130,118
1963	6,453	0,897	142,140
1964	7,093	0,943	155,338
1965	8,907	0,937	171,456
1966	8,625	0,931	182,420
1967	9,204	0,919	192,895
1968	9,647	0,920	204,164
1969	10,167	0,892	221,908
1970	9,961	1,000	240,471
1971	10,580	1,045	267,849
1972	10,658	1,140	289,450
1973	13,139	1,205	318,550

Οι βασικοί υπολογισμοί που είναι απαραίτητοι για την εκτίμηση του υποδείγματος είναι οι ακόλουθοι:

$\Sigma \Upsilon = 129,463$	$\Sigma \Upsilon^2 = 1151,018$	$\Sigma y^2 = 103,476$
$\Sigma X_1 = 15,381$	$\Sigma X_1^2 = 14,921$	$\Sigma x_1^2 = 0,135$
$\Sigma X_2 = 2966,315$	$\Sigma X_2^2 = 617645,622$	$\Sigma x_2^2 = 67706,579$
$\bar{\Upsilon} = 8,091$	$\Sigma X_1 \Upsilon = 126,957$	$\Sigma x_1 y = 2,502$
$\bar{X}_1 = 0,961$	$\Sigma X_2 \Upsilon = 26541,949$	$\Sigma x_2 y = 2540,196$
$\bar{X}_2 = 185,394$	$\Sigma X_1 X_2 = 2927,634$	$\Sigma x_1 x_2 = 76,078$

όπου $y = \Upsilon_t - \bar{Y}_t, x_1 = X_{1t} - \bar{X}_{1t}, x_2 = X_{2t} - \bar{X}_{2t}$

Οπότε:

$$\widehat{\beta}_1 = \frac{(\sum_{t=1}^{T} x_1 y)(\sum_{t=1}^{T} x_2^2) - (\sum_{t=1}^{T} x_2 y)(\sum_{t=1}^{T} x_1 x_2)}{(\sum_{t=1}^{T} x_1^2)(\sum_{t=1}^{T} x_2^2) - (\sum_{t=1}^{T} x_1 x_2)^2} = -7,002$$

$$\widehat{\beta}_2 = \frac{(\sum_{t=1}^{T} x_2 y)(\sum_{t=1}^{T} x_1^2) - (\sum_{t=1}^{T} x_1 y)(\sum_{t=1}^{T} x_1 x_2)}{(\sum_{t=1}^{T} x_1^2)(\sum_{t=1}^{T} x_2^2) - (\sum_{t=1}^{T} x_1 x_2)^2} = 0,0453$$

$$\widehat{\beta}_0 = \overline{Y}_t - \widehat{\beta}_1 \overline{X}_{1t} - \widehat{\beta}_2 \overline{X}_{2t} = 6,409$$

Η γραμμή παλινδρόμησης του δείγματος είναι:

$$\widehat{Y}_t = 6,409 - 7,002 X_{1t} + 0,0453 X_{2t}$$

Οι προβλεφθείσες τιμές $\widehat{\Upsilon}_t$ προκύπτουν από την γραμμή παλινδρόμησης του δείγματος

$\widehat{\Upsilon}_t = 6,409 - 7,002\mathrm{X}_{1t} + 0,0453\mathrm{X}_{2t}$, για $t = 1, \dots, 16$

Για $t = 1$: $\widehat{\Upsilon}_1 = 6,409 - 7,002\mathrm{X}_{11} + 0,0453\mathrm{X}_{21} = 4,615$

κτλ

Οι προβλεφθείσες τιμές του διαταρακτικού όρου $\hat{\mathrm{u}}_t$ προκύπτουν από την σχέση $\hat{\mathrm{u}}_t = \Upsilon_t - \widehat{\Upsilon}_t$, για $t = 1, \dots, 16$

Για $t = 1$: $\hat{\mathrm{u}}_1 = \Upsilon_1 - \widehat{\Upsilon}_1 = 0,506$

κτλ

Ερμηνεία των μερικών συντελεστών παλινδρόμησης

Ο συντελεστής β_j στο πολυμεταβλητό γραμμικό υπόδειγμα παριστάνει τη μεταβολή στην προσδοκώμενη τιμή της εξαρτημένης μεταβλητής όταν η ερμηνευτική μεταβλητή X_j μεταβάλλεται κατά μια μονάδα και οι υπόλοιπες ερμηνευτικές μεταβλητές παραμένουν σταθερές.

Δηλαδή, όταν όλες οι υπόλοιπες ερμηνευτικές μεταβλητές είναι σταθερές, εκτιμάμε την επίδραση της μεταβλητής X_j επί της Y.

Ο μερικός συντελεστής παλινδρομήσεως $\widehat{\beta}_j$ του πολυμεταβλητού γραμμικού υποδείγματος είναι ο συντελεστής της απλής παλινδρόμησης ανάμεσα στην Y και στην X_j όταν έχουν αφαιρεθεί οι γραμμικές επιδράσεις όλων των υπολοίπων ερμηνευτικών μεταβλητών επί της X_j και της Y.

Βιβλιογραφία

Χρήστου Κ. Γεώργιος (2007) Εισαγωγή στην Οικονομετρία, Τόμος 1, Εκδότης: Γ. ΔΑΡΔΑΝΟΣ - Κ. ΔΑΡΔΑΝΟΣ Ο.Ε.

Stock H. James, Watson W. Mark, επιμέλεια Πραγγίδης Ιωάννης - Χρυσόστομος (2017) Εισαγωγή στην Οικονομετρία, Εκδότης: Γ. ΔΑΡΔΑΝΟΣ - Κ. ΔΑΡΔΑΝΟΣ Ο.Ε.

Χρήστου Κ. Γεώργιος (2006) Εισαγωγή στην Οικονομετρία Ασκήσεις, Εκδόσεις Gutenberg.

Δριτσάκη Ν. Χάιδω, Δριτσάκη Ν. Μελίνα (2013) Εισαγωγή στην Οικονομετρία με τη Χρήση του Λογισμικού EViews, Κλειδάριθμος ΕΠΕ Εκδόσεις.