Οικονομετρία

Γρηγόρης Κόρδας

29 Οκτωβρίου 2020.

Περιεχόμενα

1 ΕΜΠΕΙΡΙΚΕΣ ΣΧΕΣΕΙΣ ΣΤΑ ΟΙΚΟΝΟΜΙΚΑ **5**

1.1 Εισαγωγή 5

2 ΜΕΤΡΑ ΠΙΘΑΝΟΤΗΤΑΣ **7**

2.1 Εισαγωγή 7

2.2 Μέτρα Πιθανότητας σε Πεπερασμένους Δειγματικούς Χώρους . 9

2.3 Μέτρα Πιθανότητας σε Αριθμήσιμα Άπειρους Δειγματικούς Χώρους 12

3 ΠΑΛΙΝΔΡΟΜΗΣΗ **13**

3.1 Εισαγωγή 13

3.2 Ελάχιστα Τετράγωνα 14

3.3 Απλή Παλίνδρομηση 16

3.4 Η Γεωμετρική Ερμηνεία των Ελαχίστων Τετραγώνων 18

3.5 Ο Συντελεστής Πολλαπλού Προσδιορισμού 20

3.6 Το Στατιστικό Υπόδειγμα της Πολλαπλής Παλινδρόμησης 22

3.7 Στατιστικές Ιδιότητες των Συντελεστών Ελαχίστων Τετραγώνων 25

4 ΠΑΡΑΒΙΑΣΕΙΣ ΤΩΝ ΥΠΟΘΕΣΕΩΝ ΤΟΥ ΚΛΑΣΙΚΟΥ ΥΠΟΔΕΙΓΜΑΤΟΣ **27**

4.1 Εισαγωγή 27

4.2 Η Υπόθεση της Γραμμικότητας 27

4.3 Λάθη στον Προσδιορισμό του Υποδείγματος 27

4.3.1 Παράλειψη σημαντικών ερμηνευτικών μεταβλητών . . . 27

4.3.2 Εισαγωγή άσχετων ερμηνευτικών μεταβλητών 28

1. ΕΜΠΕΙΡΙΚΕΣ ΣΧΕΣΕΙΣ ΣΤΑ ΟΙΚΟΝΟΜΙΚΑ

1.1. Εισαγωγή

Σκοπός της Οικονομικής Επιστήμης είναι η ανάλυση των σχέσεων που υπάρχουν μεταξύ διαφόρων οικονομικών μεταβλητών. Για παράδειγμα, ένας οικονομολόγος

2. ΜΕΤΡΑ ΠΙΘΑΝΟΤΗΤΑΣ

2.1. Εισαγωγή

Αν και η λέξη *πιθανότητα* χρησιμοποιείται συχνά στην καθημερινή μας ομιλία, είναι δύσκολο να προσδιορίσουμε τι ακριβώς εννοούμε με τον όρο αυτό. Για παράδειγμα, υπάρχει γενική συμφωνία ότι η πιθανότητα $\frac{1}{2}$ πρέπει να ανατεθεί στο ενδεχόμενο να λάβουμε γράμματα στην ρίψη ενός δίκαιου νομίσματος, αλλά είναι και φυσικό να αναζητήσουμε την εξήγηση αυτής της παροδοχής.

Μια πρώτη γραμμή σκέψης είναι ότι αν υπάρχουν μόνο δύο δυνατά ενδεχόμενα και αν είναι γενικά παραδεκτό ότι κανένα από τα δύο δεν φαίνεται να προτιμάται έναντι του άλλου, τότε θα πρέπει τα δύο ενδεχόμενα να είναι ισοπίθανα, δηλ. το καθένα θα πρέπει να λάβει πιθανότητα $\frac{1}{2}$.

Η Αρχή του Ανεπαρκούς Λόγου του Laplace: Αν υπάρχουν n αμοιβαίως αποκλειόμενα ενδεχόμενα και δεν υπάρχει κανένας λόγος (ή δεν υπάρχει επαρκής λόγος) να πιστεύουμε ότι κάποια απ′ αυτά είναι περισσότερο ή λιγότερο πιθανά, η πιθανότητα που θα πρέπει να ανατεθεί στο καθένα θα πρέπει να είναι $1/n$.

Ένας δεύτερος τρόπος να αναθέσουμε πιθανότητες στα δυνατά ενδεχόμενα η οποία υποστηρίχθηκε από τον Savage (1954) θα ήταν να επιχειρηματολογήσουμε ότι οι πιθανότητες είναι υποκειμενικές εκτιμήσεις της πιθανοφάνειας των επιμέρους ενδεχομένων. Δεν υπάρχει κανένας λόγος να συμφωνήσουμε όλοι στην πιθανότητα των γραμμάτων στην ρίψη ενός δίκαιου νομίσματος, αρκεί οι πιθανότητες που ο καθένας μας αναθέτει στα πιθανά ενδεχόμενα να ικανοποιούν κάποιες αυτονόητες συνθήκες, όπως ότι είναι θετικές και αθροίζουν στο 1.

Η Υποκειμενική Πιθανότητα του Savage: Αν όλοι συμφωνούμε στον δειγματικό χώρο S και όλοι υπακούμε στην συνθήκη ότι οι υποκειμενικές μας πιθανότητες είναι θετικές και αθροίζουν στο 1, μπορούμε να αναθέσουμε όποιο αριθμό μεταξύ 0 και 1 επιθυμούμε στα επιμέρους ενδεχόμενα.

Αυτός είναι ο ορισμός την λεγόμενης πιθανότητας κατά Bayes, με την οποία ο von Mises (1957) δεν θα μπορούσε να διαφωνεί περισσότερο. Επιχειρηματολόγησε ότι μια αμιγώς υποκειμενική αντίληψη της πιθανότητας δεν θα ήταν αρκετός για να εντάξει την Στατιστική στην οικογένεια των Θετικών Επιστημών. Αν δεν μπορούμε να συμφωνήσουμε σε έναν κοινό τρόπο να αναθέτουμε πιθανότητες στα επιμέρους ενδεχόμενα, τότε επιτρέπουμε το ενδεχόμενο διαφορετικοί επιστήμονες που δουλεύουν πάνω στο ίδιο πρόβλημα να φτάσουν σε διαφορετικά συμπεράσματα χωρίς να έχουμε τον τρόπο να αποφασίσουμε ποια απάντηση είναι η 'σωστή'.

Η Αντικειμενική Πιθανότητα του von Mises: Η πιθανότητα ενός ενδεχομένου είναι η μακροχρόνια συχνότητά του καθώς ο αριθμός των πειραμάτων τείνει στο άπειρο.

Αυτός είναι ο λεγόμενος 'κλασικός' ορισμός της πιθανότητας, ο οποίος έχει αρκετά μειονεκτήματα, το σημαντικότερο των οποίων ίσως είναι ότι βασίζεται στην ιδέα των τυχαίων πειραμάτων που επαναλαμβάνονται επ'άπειρον. Αλλά ούτε και η κριτική του von Mises στον Savage είναι και τόσο πειστική, ιδιαίτερα αν κάποιος φανταστεί την περίπτωση διαφόρων επενδυτών οι οποίοι βασιζόμενοι σε διαφορετικές εκτιμήσεις για την μελλοντική αξία ενός συνόλου χρεωγράφων φθάνουν σε διαφορετικές επενδυτικές αποφάσεις. Πράγματι, το ίδιο το γεγονός της αγοραπωλησίας μιας μετοχής καταδεικνύει ότι ο αγοραστής και ο πωλητής της έχουν διαφορετικές προσδοκίες για την μελλοντική της αξία.

Τα θέματα αυτά έχουν υπάρξει αντικείμενο μεγάλης διαμάχης ανάμεσα στους υποστηρικτές του κλασικού ορισμού της πιθανότητας και τους υποστηρικτές του ορισμού κατά Bayes. Εμείς όμως εδώ θα παρακάμψουμε αυτές τις φιλοσοφικές (μετα-μαθηματικές) αναζητήσεις και θα υιοθετήσουμε την αγνωστικιστική (μαθηματική) θέση του Kolmogorov.

Ο Αξιωματικός Ορισμός της Πιθανότητας του Kolmogorov: Όπως όλες οι μαθηματικές έννοιες, η πιθανότητα μπορεί (και πρέπει!) να ορισθεί σε αξιωματική βάση. Ακριβώς όπως οι έννοιες του 'σημείου' και της 'γραμμής' στην Ευκλείδιο Γεωμετρία ορίζονται μέσα από μια σειρά ιδιοτήτων που τους αναθέτουμε αξιωματικά και χωρίς δικαιολόγησή, η προέλευση και η σημασία των αξιωμάτων του λογισμού των πιθανοτήτων που θα παρουσιάσουμε παρακάτω *δεν* είναι μέρος των μαθηματικών.

2.2. Μέτρα Πιθανότητας σε Πεπερασμένους Δειγματικούς Χώρους

Μια διαδικασία όπως η ρίψη ενός νομίσματος, η ρίψη ενός ζαριού, ή η επιλογή ενός χαρτιού από μια τράπουλα ονομάζεται *πείραμα τύχης*. Αρχίζουμε με την ανάλυση πειραμάτων που έχουν πεπερασμένο αριθμό *στοιχειωδών ενδεχομένων* (τα λεγόμενα ω): στο πείραμα της ρίψης του νομίσματος έχουμε 2 δυνατά ενδεχόμενα (κορώνα H ή γράμματα T), στο πείραμα τη ρίψης ενός ζαριού υπάρχουν 6 $(1, .., 6)$, και στην επιλογή ενός χαρτιού από μια τράπουλα 52. Το *ενδεχόμενο A* είναι ένα σύνολο στοιχειωδών ενδεχομένων ω. Το σύνολο όλων των στοιχειωδών ενδεχομένων ονομάζεται *δειγματικός χώρος* και συμβολίζεται με το Ω.

Σε ένα δειγματικό χώρο με n στοιχειώδη ενδεχόμενα $\Omega = \{\omega_1, ..., \omega_n\}$ υπάρχουν δυνιτικά 2^n ενδεχόμενα που θα μπορούσαμε να κατασκευάσουμε. Για παράδειγμα, στο πείραμα της ρίψης ενός ζαριού υπάρχουν συνολικά

$$C_0^6 + C_1^6 + \cdots C_6^6 = 2^6 = 64,$$

δυνητικά ενδιαφέροντα ενδεχόμενα, όπου $C_x^n = n!/[(n-x)!\, x!]$ είναι ο αριθμός των συνδιασμών n αντικειμένων σε ομάδες x αντικειμένων.

Ορισμός 1 *Μια πεπερασμένη συλλογή $\mathscr{F}$ υποσυνόλων του Ω είναι ένα πεδίο αν*

i. $\Omega \in \mathscr{F}$;

ii. (κλείσιμο ως προς την συμπλήρωση) Αν $A \in \mathscr{F}$ τότε και $A^c \in \mathscr{F}$;

iii. *(κλείσιμο ως προς την ένωση) Αν* $A, B \in \mathscr{F}$ *τότε και* $A \cup B \in \mathscr{F}$.

Δεδομένης μιας συλλογής $\mathscr{F}$ είναι εύκολο να ελέγξουμε αν είναι πεδίο ή όχι. Για παράδειγμα, είναι εύκολο να δούμε ότι το $\mathscr{F} = \{\emptyset, \Omega\}$ είναι το μικρότερο δυνατό πεδίο. Στο πείραμα ρίψης ενός ζαριού, $\Omega = \{1, 2, 3, 4, 5, 6\}$, και η συλλογή $\mathscr{F} = \{\emptyset, 1 \cup 3 \cup 5, 2 \cup 4 \cup 6, \Omega\}$ υποσυνόλων του Ω είναι πράγματι ένα πεδίο. Παρατηρήστε ωστόσο ότι αυτό το πεδίο δεν περιέχει κανένα από τα στοιχειώδη ενδεχόμενα ω και παραλείπει διάφορα υποσύνολα του Ω (δηλ. ενδεχόμενα) όπως το $5 \cup 6$. Το πεδίο αυτό είναι στην πραγματικότητα το μικρότερο πεδίο που περιέχει τα μονά και ζυγά ενδεχόμενα και είναι γνωστό ως το πεδίο που *παράγεται* από αυτά τα ενδεχόμενα. Γενικά, το μικρότερο πεδίο που παράγεται από ένα ενδεχόμενο A είναι το $\mathscr{F} = \{\emptyset, A, A^c, \Omega\}$.

Είναι διδακτικό να δούμε πώς ακριβώς μια *συλλογή* ενδεχομένων παράγει ένα *πεδίο*. Ας θεωρήσουμε μια συλλογή $\mathscr{A} = \{A_1, A_2\}$ των ενδεχομένων $A_1, A_2 \subset \Omega$. Αρχικά θέτουμε το ίδιο το Ω στο πεδίο καθώς και το συμπληρωματικό του κενό ενδεχόμενο $\emptyset$. Μετά προσθέτουμε τα ενδοχόμενα και τα συμπληρωματικά τους και παίρνουμε $\Omega, \emptyset, A_1, A_1^c, A_2, A_2^c$. Σε αυτά μετά προσθέτουμε όλες τις δυνατές ενώσεις που δεν συμπεριλαμβάνονται ήδη: $A_1 \cup A_2, A_1 \cup A_2^c, A_1^c \cup A_2$ και $A_1^c \cup A_2^c$. Τέλος προσθέτουμε και όλα τα συμπληρωματικά αυτών των ενώσεων όπως το $(A_1 \cup A_2)^c$ για να παράξουμε το πεδίο.

Το επόμενο βήμα μας είναι να αναθέσουμε πιθανότητες στα ενδεχόμενα που περιέχονται στο πεδίο $\mathscr{F}$.

Ορισμός 2 *Το μέτρο πιθανότητας είναι μια συνάρτηση συνόλων (set function)* $P(\cdot)$, *που απεικονίζει στοιχεία του* $\mathscr{F}$ *στο* $[0, 1]$, *με τις ακόλουθες ιδιότητες:*

i. *Για κάθε* $A \in \mathscr{F}$, $0 \leq P(A) \leq 1$;

ii. $P(\emptyset) = 0$ *και* $P(\Omega) = 1$;

iii. *Για* $A_1, A_2 \in \mathscr{F}$, *αν* $A_1 \cap A_2 = \emptyset$ *τότε* $P(A_1 \cup A_2) = P(A_1) + P(A_2)$.

Παράδειγμα 1. Ας θεωρήσουμε το πείραμα της ρίψης ενός νομίσματος μία φορά. Τότε $\Omega = \{H, T\}$, το $\mathscr{F} = \{\emptyset, H, T, \Omega\}$ είναι ένα πεδίο, και φαίνεται λογικό να αναθέσουμε $P(H) = P(T) = \frac{1}{2}$, ώστε $P(\Omega) = 1$, $P(\emptyset) = 0$, και $P(\cdot)$ είναι ένα μέτρο πιθανότητας πάνω στο $\mathscr{F}$. ■

Παράδειγμα 2. Μία παραλλαγή του πειράματος στο προηγούμενο παράδειγμα είναι η ρίψη δύο πανομιότυπων νομισμάτων. Υπάρχουν τουλάχιστον δύο δειγματικοί χώροι που θα μπορούσαμε να σκεφτούμε σε αυτή την περίπτωση: $\Omega_1 = \{HH, HT, TT\}$, ή $\Omega_2 = \{HH, HT, TH, TT\}$. Ο πρώτος δειγματικός χώρος αντιστοιχεί σε ένα πείραμα στο οποίο η σειρά εμφάνισης των H και T δεν μας ενδιαφέρει, ενώ ο δεύτερος δειγματικός χώρος λαμβάνει υπόψιν την σειρά ενφάνισης των H και T. Αν τα νομίσματα είναι δίκαια και η σειρά δεν μας απασχολεί, μια λογική ανάθεση πιθανοτήτων θα ήταν, $P(HH) = P(TT) = \frac{1}{4}$ και $P(HT) = \frac{1}{2}$. Αν αντίθετα η σειρά εμφάνισης των H και T ληφθεί υπόψιν, θα έπρεπε αν αναθέσουμε τις πιθανότητες $P(HH) = P(HT) = P(TH) = P(TT) = \frac{1}{4}$. Το παράδειγμα αυτό καταδεικνύει την σημασία του εξάρχής προσδιορισμού των ενδεχομένων που μας ενδιαφέρουν, ώστε μετά να μπορέσουμε να ορίσουμε κατάλληλα τον δειγματικό χώρο Ω, το πεδίο $\mathscr{F}$, και το μέτρο πιθανότητας P. ■

Παράδειγμα 3. Στο πείραμα της ρίψης ενός ζαριού μία φορά, $\Omega = \{1, 2, 3, 4, 5, 6\}$, και το $\mathscr{F}$ θα μπορούσε να αποτελείται από όλα τα 2^6 δυνατούς συνδυασμούς ενδεχομένων στο Ω, δηλ. το κενό σύνολο $\emptyset$, τα C_1^6 στοιχειώδη ενδεχόμενα, τους C_2^6 συνδυασμούς δύο ενδεχομένων, τους C_3^6 συνδυασμούς 3 ενδεχομένων,..., ως και το $C_6^6 = 1$ ενδεχόμενο $1 \cup 2 \cup 3 \cup 4 \cup 5 \cup 6$, καθώς και το Ω. Όπως πριν τα στοιχειώδη ενδεχόμενα ανήκουν στο $\mathscr{F}$ και είναι λογικό να τους ανατεθεί $P(1) = P(2) = ... = P(6) = \frac{1}{6}$. Οι πιθανότητες των υπολοίπων μελών του $\mathscr{F}$ μπορούν τώρα να υπολογιστούν κάνοντας χρήση της ιδιότητας (iii) στον ορισμό του μέτρου πιθανότητας. ■

Παράδειγμα 4. Μία παραλλαγή του παραπάνω πειράματος θα ήταν το πείραμα στο οποίο $\Omega = \{1, 2, 3, 4, 5, 6\}$ όπως πριν, αλλά $\mathscr{F} = \{\emptyset, 1 \cup 3 \cup 5, 2 \cup 4 \cup 6, \Omega\}$, ώστε να ενδιαφερόμαστε μόνο για τα μονά και ζυγά ενδεχόμενα. Παρατηρήστε ότι τα στοιχειώδη ενδεχόμενα δεν είναι μέλη του συγκεκριμένου πεδίου $\mathscr{F}$, οπότε ο ορισμός του μέτρου πιθανότητας P που δόθηκε στο παράδειγμα 3 δεν είναι δυνατός. Μπορούμε παρόλα αυτά να ορίσουμε μια εξωτερική βοηθητική συνάρτηση $p(\cdot)$ και να αναθέσουμε $p(1) = p(2) = ... = p(6) = \frac{1}{6}$. Η συνάρτηση p μοιάζει με μέτρο πιθανότητας, αλλά αφού δεν είναι ορισμένη επάνω σε ένα πεδίο ούτε της έχουμε δώσει έχει τις προσθετικές ιδιότητες του μέτρου, δεν είναι μέτρο πιθανότητας. Παρόλα αυτά, η p μπορεί να χρησιμοποιηθεί για να ορισθεί ένα κανονικό μέτρο πιθανότητας P πάνω στο $\mathscr{F}$. Θέτουμε $P(1 \cup 3 \cup 5) = p(1) + p(3) + p(5)$, $P(2 \cup 4 \cup 6) = p(2) + p(4) + p(6)$, $P(\emptyset) = 0$, και $P(\Omega) = 1$. Το P αυτό ικανοποιεί όλες τις ιδιότητες στον ορισμό του μέτρου πιθανότητας, και είναι πράγματι ένα μέτρο πιθανότητας πάνω στο $\mathscr{F}$. ■

Στο τελευταίο παράδειγμα, τα στοιχειώδη ενδεχόμενα, όπως π.χ. το ‘6’, δεν ανήκουν στο $\mathscr{F}$, και λέμε ότι είναι *μη-μετρήσιμα* (non- measurable) ως προς το συγκεκριμένο $\mathscr{F}$.

Ορισμός 3 *Τα ενδεχόμενα που ανήκουν στο $\mathscr{F}$ ονομάζονται μετρήσιμα (measurable) ως προς το $\mathscr{F}$, ενώ αυτά που δεν ανήκουν στο $\mathscr{F}$ ονομάζονται μη-μετρήσιμα (non-measurable) ως προς το $\mathscr{F}$. Η τριάδα $(\Omega, \mathscr{F}, P)$ ονομάζεται πιθανοχώρος (probability space).*

2.3. Μέτρα Πιθανότητας σε Αριθμήσιμα Άπειρους Δειγματικούς Χώρους

Ιτ ις νοω τιμε το εξτενδ τηε νοτιονς οφ α φιελδ ανδ α προβαβιλιτψ μεασυρε το ςασες ωηερε Ω ις ςομποσεδ οφ αν ινφινιτε νυμβερ οφ ουτςομες. Τηε σαμπλε σπαςε οφ μανψ εξπεριμεντς ςαν βε μοδελεδ βψ σεττινγ $\Omega = \{1, 2, ...\}$, τηε σετ οφ αλλ ποσιτιε ιντεγερς. Ιν τηεσε ςασες, τηε σαμπλε σπαςε ις σαιδ το βε *ςουνταβλψ ινφινιτε*. Εξαμπλες οφ εξπεριμεντς ωιτη τηις σαμπλε σπαςε μιγητ βε τηε νυμβερ οφ παρτιςλες εμιττεδ βψ α ραδιοαςτιε συβστανςε, οφ τηε νυμβερ οφ μεσσαγες αρριυνγ ατ α σωιτςηινγ βοαρδ ιν α γιεν περιοδ οφ τιμε. Τηε σαμπλε σπαςε ις περμιττεδ το βε ινφινιτε βεςαυσε τηερε ις νο ρεασον το φιξ αν υππερ βουνδ το τηε ςουντς τηατ ςουλδ βε οβσερεδ ιν αδανςε οφ τηε εξπεριμεντ.

3. ΠΑΛΙΝΔΡΟΜΗΣΗ

3.1. Εισαγωγή

Βασιζόμενος σε κάποια θεωρία ένας επιστήμονας πιστεύει ότι υπάρχει κάποια *συναρτησιακή σχέση* που συνδέει την βαθμωτή μεταβλητή Y με το διάνυσμα k μεταβλητών $\boldsymbol{x} = (X_1, ..., X_k)'$. Η συνάρτηση αυτή είναι γνωστό ότι είναι *γραμμική*

εκτός μόνο από ένα πεπερασμένο αριθμό παραμέτρων β τον οποίο και επιθυμεί να εκτιμήσει. Για το σκοπό αυτό λαμβάνει ένα *δείγμα*,

$$\{(y_1, x_1), (y_2, x_2), ..., (y_n, x_n)\} = \{(y_i, x_i) : i = 1, 2, ..., n\}$$

όπου το κάθε ζεύγος $(y_i, \boldsymbol{x}_i) \in \mathbb{R} \times \mathbb{R}^{(k+1)}$.

Το *γενικό γραμμικό υπόδειγμα της πολλαπλής παλινδρόμησης* είναι

$$y_i = \beta_0 x_{i0} + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + u_i, \qquad i = 1, 2, ..., n \tag{1.1}$$

όπου $x_{i0} \equiv 1$ για κάθε $i = 1, 2, ..., n$, είναι ο σταθερός όρος της παλινδρόμησης.

Συλλέγοντας όλα τα στοιχεία των ανεξαρτήτων μεταβλητών για την παρατήρηση i στο $1 \times (k + 1)$ διάνυσμα $\boldsymbol{x}_i' = (1, x_{i1}, ..., x_{ik})$ και όλους τις παραμέτρους στο $(k + 1) \times 1$ διάνυσμα $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_k)'$ η (1.1) μπορεί να γραφεί ως

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + u_i, \qquad i = 1, 2, ..., n. \tag{1.2}$$

Συλλέγοντας τα στοιχεία του δείγματος περεταίρω στα παρακάτω διανύσματα και πίνακες, γράφουμε την (1.1) ως

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}_{n\times 1} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix}_{n\times(k+1)} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}_{(k+1)\times 1} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}_{n\times 1} \tag{1.3}$$

$$\boldsymbol{y}_{n\times 1} = \boldsymbol{X}_{n\times(k+1)} \boldsymbol{\beta}_{(k+1)\times 1} + \boldsymbol{u}_{n\times 1}$$

όπου

- $\boldsymbol{y}$ είναι το $n\times 1$ διάνυσμα των παρατηρήσεων της εξαρτημένης μεταβλητής,
- $\boldsymbol{X}$ είναι ο $n \times (k+1)$ πίνακας των παρατηρήσεων των ανεξάρτητων μεταβλητών,
- $\boldsymbol{\beta}$ είναι το $(k+1) \times 1$ διάνυσμα των παραμέτρων προς εκτίμηση, και
- και $\boldsymbol{u}$ είναι το $n \times 1$ διάνυσμα των σφαλμάτων.

3.2. Ελάχιστα Τετράγωνα

Για κάθε $\boldsymbol{b} \in \mathbb{R}^{k+1}$ τα κατάλοιπα είναι

$$\boldsymbol{u}(\boldsymbol{b}) = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}. \tag{2.4}$$

Το άθροισμα των τετραγωνικών καταλοίπων (Residual Sum of Squared , RSS) που επιθυμούμε να ελαχιστοποιήσουμε είναι

$$\begin{aligned} RSS(\boldsymbol{b}) &= \boldsymbol{u}(\boldsymbol{b})'\boldsymbol{u}(\boldsymbol{b}) \\ &= (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}) \\ &= \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{X}\boldsymbol{b} - \boldsymbol{b}'\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{b}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} \\ &= \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{b}'\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{b}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b}, \end{aligned} \tag{2.5}$$

όπου χρησιμοποιήσαμε το γεγονός ότι οτί ο ανάστροφος ενός βαθμωτού είναι ο ευατός του, ώστε $\boldsymbol{y}'\boldsymbol{X}\boldsymbol{\beta} = (\boldsymbol{y}'\boldsymbol{X}\boldsymbol{b})' = \boldsymbol{b}'\boldsymbol{X}'\boldsymbol{y}$. Παραγωγίζοντας και θέτοντας την παράγωγο ίση με το μηδέν, λαμβάνουμε τις συνθήκες πρώτης τάξης για την ελαχιστοποίηση της $SSR(\boldsymbol{b})$ ως

$$\frac{\partial SSR(\hat{\boldsymbol{b}})}{\partial b} = -2\boldsymbol{X}'\boldsymbol{y} + 2\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{b}} = 0$$

ή

$$\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{b}} = \boldsymbol{X}'\boldsymbol{y}. \tag{2.6}$$

Οι εξισώσεις (2.6) ονομάζονται *κανονικές εξισώσεις*, και αποτελούν ένα σύστημα $(k+1)$ εξισώσεων με $(k+1)$ αγνώστους. Το σύστημα αυτό έχει μοναδική λύση εάν και μόνον εάν ο τετραγωνικός πίνακας $\boldsymbol{X'X}$, διαστάσεων $(k+1)\times(k+1)$, είναι αντιστρέψιμος, δηλαδή αν $|\boldsymbol{X'X}| \neq 0$. Η τελευταία συνθήκη ικανοποιείται εάν ο πίνακας $\boldsymbol{X}$ είναι πλήρους βαθμού, δηλαδή εάν

$$r(\boldsymbol{X}) = k + 1. \tag{2.7}$$

Στην περίπτωση αυτή ο πίνακας $\boldsymbol{X'X}$ είναι θετικά ορισμένος, ο αντίστροφός του $(\boldsymbol{X'X})^{-1}$ υπάρχει και είναι επίσης θετικά ορισμένος, και το σύστημα των κανονικών εξισώσεων (2.6) έχει μοναδική λύση.

Υποθέτοντας λοιπόν ότι ο πίνακας X είναι πλήρους βαθμού, οι εκτιμήτριες ελαχίστων τετραγώνων $\hat{\boldsymbol{b}} = (\hat{b}_0, \hat{b}_1, ..., \hat{b}_k)'$ της παλινδρόμησης (1.1) δίνονται από την σχέση

$$\begin{aligned} \hat{\boldsymbol{b}} &= (\boldsymbol{X'X})^{-1}\boldsymbol{X'y} \qquad (2.8) \\ &= (\Sigma \boldsymbol{x}_i \boldsymbol{x}_i')^{-1}\, \Sigma \boldsymbol{x}_i y_i. \end{aligned}$$

Η συνθήκη δεύτερης τάξης για ολικό ελάχιστο (global minimum) είναι ότι ο πίνακας των δευτέρων μερικών παραγώγων (Hessian matrix) εκτιμημένος στο $\hat{\boldsymbol{b}}$

$$\frac{\partial^2 SSR(\hat{\boldsymbol{b}})}{\partial \boldsymbol{b} \partial \boldsymbol{b}'} = 2\boldsymbol{X'X} \tag{2.9}$$

είναι θετικά ορισμένος, κάτι που ισχύει όπως έχουμε ήδη υποθέσει. Συμπεραίνουμε λοιπόν ότι το διάνυσμα $\hat{\boldsymbol{b}}$ είναι το ολικό ελάχιστο της συνάρτησης $SSR(\boldsymbol{b})$.

Το διάνυσμα των προσαρμοσμένων τιμών (fitted values) της y είναι

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{b}} = \boldsymbol{X}(\boldsymbol{XX})^{-1}\boldsymbol{X'y} \equiv \boldsymbol{P_X y}, \tag{2.10}$$

και τα κατάλοιπα (residuals) της παλινδρόμησης είναι

$$\hat{\boldsymbol{u}} = \boldsymbol{y} - \hat{\boldsymbol{y}} = \boldsymbol{y} - \boldsymbol{P_X y} = (\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y} \equiv \boldsymbol{M_X y}. \tag{2.11}$$

Μπορούμε, λοιπόν, να σπάσουμε το $\boldsymbol{y}$ σε ένα κομμάτι $\hat{\boldsymbol{y}}$ που εξηγήτει από τα $\boldsymbol{X}$, και σε ένα άλλο κομμάτι $\hat{\boldsymbol{u}}$ που μένει ανεξήγητο, δηλαδή να γράψουμε

$$\boldsymbol{y} = \hat{\boldsymbol{y}} + \hat{\boldsymbol{u}} = \boldsymbol{P_X y} + \boldsymbol{M_X y}. \tag{2.12}$$

Οι πίνακες $\boldsymbol{P_X}$ και $\boldsymbol{M_X}$ ονομάζονται (projection matrices) και έχουν ιδιέταιρες ιδιότητες. Ένας τρόπος να δούμε αυτούς τους πίνακες είναι να τους δούμε σαν φίλτρα που αποσυνθέτουν το $\boldsymbol{y}$ σε δύο μέρη: εφαρμόζοντας το $\boldsymbol{P_X}$ στο $\boldsymbol{y}$ αποσπούμε το $\hat{\boldsymbol{y}}$, ενώ εφαρμόζοντας το $\boldsymbol{M_X}$ στο $\boldsymbol{y}$ αποσπούμε το $\hat{\boldsymbol{u}}$.

3.3. Απλή Παλίνδρομηση

Σαν ένα παράδειγμα της εφαρμογής των τύπων του γενικού υποδείγματος της πολλαπλής παλινδρόμησης ας θεωρήσουμε την εφαρμογή τους στην περίπτωση της απλής παλινδρόμησης, δηλαδή όταν έχουμε μόνο μία ανεξάρτητη μεταβλητή στο υποδειγμά μας. Σε αυτή την περίπτωση το γενικό γραμμικό υπόδειγμα στην (1.1) γίνεται

$$y_i = \beta_0 + \beta_1 x_i + u_i, \qquad i = 1, 2, ..., n. \tag{3.13}$$

με $k = 2$ και ο πίνακας X των ανεξάρτητων μεταβλητών δίνεται από τον $n \times 2$ πίνακα

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}. \tag{3.14}$$

Έχουμε,

$$\boldsymbol{X'X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{bmatrix} \tag{3.15}$$

και

$$\boldsymbol{X'y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \Sigma y_i \\ \Sigma x_i y_i \end{bmatrix}, \tag{3.16}$$

άρα οι κανονικές εξισώσεις για $k = 2$ είναι

$$\begin{bmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} \Sigma y_i \\ \Sigma x_i y_i \end{bmatrix}, \tag{3.17}$$

ή

$$\begin{aligned} nb_0 + b_1 \Sigma x_i &= \Sigma y_i \\ b_0 \Sigma x_i + b_1 \Sigma x_i^2 &= \Sigma x_i y_i \end{aligned} \tag{3.18}$$

το οποίο είναι ένα σύστημα δύο εξισώσεων με δύο αγνώστους, b_0 και b_1. Υποθέτοντας ότι $r(\boldsymbol{X}) = 2$, ο αντίστροφος του $\boldsymbol{X}'\boldsymbol{X}$ υπάρχει και είναι

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \frac{1}{\Sigma x_i^2 - (\Sigma x_i)^2} \begin{bmatrix} \Sigma x_i^2 & -\Sigma x_i \\ -\Sigma x_i & n \end{bmatrix}. \tag{3.19}$$

Οι συντελεστές ελαχίστων τετραγώνων $\hat{b} = (\hat{b}_0, \hat{b}_1)'$ είναι

$$\begin{aligned} \hat{\boldsymbol{b}} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \\ &= \frac{1}{n\Sigma x_i^2 - (\Sigma x_i)^2} \begin{bmatrix} \Sigma x_i^2 & -\Sigma x_i \\ -\Sigma x_i & n \end{bmatrix} \begin{bmatrix} \Sigma y_i \\ \Sigma x_i y_i \end{bmatrix} \\ &= \frac{1}{n\Sigma x_i^2 - (\Sigma x_i)^2} \begin{bmatrix} \Sigma x_i^2 \Sigma y_i - \Sigma x_i \Sigma x_i y_i \\ n\Sigma x_i y_i - \Sigma x_i \Sigma y_i \end{bmatrix}. \end{aligned} \tag{3.20}$$

Η εκτιμήτρια της κλίσης της παλινδρόμησης $\hat{b}_1$ είναι

$$\hat{b}_1 = \frac{n\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2} = \frac{\Sigma x_i y_i - n\bar{x}\bar{y}}{\Sigma x_i^2 - n\bar{x}^2} \tag{3.21}$$

η οποία, χρησιμοποιώντας τους τύπους της δειγματικής διακύμανσης και της δειγματικής συνδιακύμανσης, μπορεί να γραφτεί ως

$$\hat{b}_1 = \frac{\Sigma(y_i - \bar{y})(x_i - \bar{x})}{\Sigma(x_i - \bar{x})^2} = \frac{Cov(y_i, x_i)}{Var(x_i)}, \tag{3.22}$$

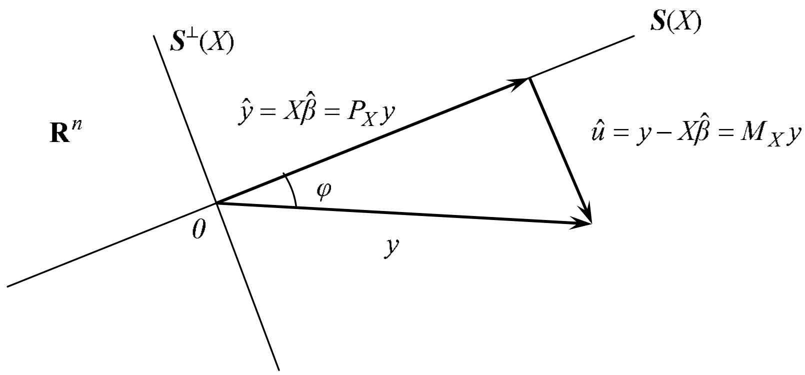
όπου $Cov(y_i, x_i)$ είναι η δειγματική συνδιακύμανση των y_i και x_i, και $Var(x_i)$ είναι η δειγματική διακύμανση των x_i.

Τέλος, η εκτιμήτρια της σταθεράς της παλινδρόμησης $\hat{b}_0$ είναι

$$\begin{aligned} \hat{b}_0 &= \frac{(n\bar{y})\Sigma x_i^2 - (n\bar{x})\Sigma x_i y_i}{n\Sigma x_i^2 - (n\bar{x})^2} \\ &= \frac{\bar{y}\left[\Sigma x_i^2 - (n\bar{x})\right] + \bar{y}(n\bar{x})^2 - (n\bar{x})\Sigma x_i y_i}{n\Sigma x_i^2 - (n\bar{x})^2} \\ &= \bar{y} - \frac{\Sigma x_i y_i - n\bar{x}\bar{y}}{\Sigma x_i^2 - n\bar{x}^2}\ \bar{x} \\ &= \bar{y} - \hat{b}_1\bar{x}. \end{aligned} \tag{3.23}$$

Παράδειγμα 5.

■



Σχήμα 3.1: Η γεωμετρία των ελαχίστων τετραγώνων.

3.4. Η Γεωμετρική Ερμηνεία των Ελαχίστων Τετραγώνων

Το διάνυσμα $\boldsymbol{y}$ ανήκει στο $\mathbb{R}^n$, και αφού ο πίνακας $\boldsymbol{X}$ έχει βαθμό $(k+1) < n$, το $\boldsymbol{X}$ *παράγει* (generates or spans) ένα $(k+1)$-διάστατο υποχώρο του $\mathbb{R}^n$. Συμβολίζουμε με $\mathcal{S}(\boldsymbol{X})$ τον γραμμικό υποχώρο που παράγεται από τον $\boldsymbol{X}$

$$\mathcal{S}(\boldsymbol{X}) = \{\boldsymbol{X}\boldsymbol{b} : \boldsymbol{b} \in \mathbb{R}^{k+1}\},$$

και με $\mathcal{S}^{\perp}(\boldsymbol{X})$ τον υποχώρο του $\mathbb{R}^n$ που είναι ορθογώνιος προς τον $\mathcal{S}(\boldsymbol{X})$, έτσι ώστε $\mathcal{S}(\boldsymbol{X}) \bigcup \mathcal{S}^{\perp}(\boldsymbol{X}) = \mathbb{R}^n$. Άρα το $\boldsymbol{y}$ μπορεί να γραφτεί ως γραμμικός συνδυασμός στοιχείων του $\mathcal{S}(\boldsymbol{X})$ και του $\mathcal{S}^{\perp}(\boldsymbol{X})$.

Ενδιαφερόμαστε να βρούμε το στοιχείο $\boldsymbol{X}\hat{\boldsymbol{b}}$ του $\mathcal{S}(\boldsymbol{X})$ που ελαχιστοποιεί την απόσταση μεταξύ του $\boldsymbol{y}$ και του $\mathcal{S}(\boldsymbol{X})$. Γράφοντας $\boldsymbol{u} = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}$ για το διάνυσμα των καταλοίπων στο σημείο $\boldsymbol{b} \in \mathbb{R}^{k+1}$, θέλουμε να βρούμε το $\boldsymbol{u}$ με το *ελάχιστο μήκος*. Γεωμετρικά, το διάνυσμα καταλοίπων με το ελάχιστο μήκος $\hat{\boldsymbol{u}}$ είναι το διάνυσμα που είναι *κάθετο* στο $\mathcal{S}(\boldsymbol{X})$, δηλ.

$$\hat{\boldsymbol{u}} \perp \mathcal{S}(\boldsymbol{X}). \tag{4.24}$$

Η παραπάνω συνθήκη συνεπάγεται ότι $\hat{\boldsymbol{u}} \perp \boldsymbol{X}$ επίσης, κάτι που συνεπάγεται ότι $\boldsymbol{X}'\hat{\boldsymbol{u}} = 0$ (μηδενική συσχέτιση στο δείγμα ανάμεσα στα $\boldsymbol{X}$ και το $\hat{\boldsymbol{u}}$), από την οποία παίρνουμε

$$\begin{aligned} & \boldsymbol{X}'\hat{\boldsymbol{u}} = 0 \\ \Leftrightarrow\ & \boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}'\hat{\boldsymbol{b}}) = 0 \\ \Leftrightarrow\ & \boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{b}} = \boldsymbol{X}'\boldsymbol{y} \\ \Leftrightarrow\ & \hat{\boldsymbol{b}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}. \end{aligned} \tag{4.25}$$

Βλέπουμε λοιπόν ότι φτάσαμε στο διάνυσμα των ελαχίστων τετραγώνων χρησιμοποιώντας ένα καθαρά γεωμετρικό συλλογισμό. Έχει ενδιαφέρον να παρατηρήσουμε ότι φτάνουμε στο ίδιο αποτέλεσμα αν θέσουμε το $\hat{\boldsymbol{u}}$ κάθετο σε οποιονδήποτε γραμμικό συνδυασμό του $\boldsymbol{X}$, δηλαδή σε οποιοδήποτε στοιχείο του $\mathcal{S}(X)$: $u \perp \boldsymbol{Xc}, \boldsymbol{c} \in \mathbb{R}^{k+1}$ μας δίνει, $\boldsymbol{c}'\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{b}}) = 0 \Leftrightarrow \boldsymbol{c}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{b}} = \boldsymbol{c}'\boldsymbol{X}'\boldsymbol{y} \Leftrightarrow \hat{\boldsymbol{b}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$.

Το ότι ο εκτιμητής ελαχίστων τετραγώνων έχει αυτή την γεωμετρική ερμηνεία θα έπρεπε να είναι αναμενόμενο, αν συλλογιστούμε ότι το Ευκλείδειο μήκος ενός διανύσματος $\boldsymbol{z}$ είναι

$$||\boldsymbol{z}|| = \sqrt{\Sigma z_i^2} = \sqrt{\boldsymbol{z}'\boldsymbol{z}}$$

από το οποίο βλέπουμε ότι το πρόβλημα των ελαχίστων τετραγώνων

$$\hat{\boldsymbol{b}} = \underset{\boldsymbol{b}}{\operatorname{argmin}}\ ||\boldsymbol{u}(\boldsymbol{b})|| = \underset{\boldsymbol{b}}{\operatorname{argmin}}\ \sqrt{\boldsymbol{u}(\boldsymbol{b})'\boldsymbol{u}(\boldsymbol{b})} = \underset{\boldsymbol{b}}{\operatorname{argmin}}\ \boldsymbol{u}(\boldsymbol{b})'u(\boldsymbol{b})$$

ελαχιστοποιεί το Ευκλείδειο μήκος του διανύσματος των καταλοίπων. Από το Πυθαγόρειο Θεώρημα ξέρουμε ότι το μήκος αυτό ελαχιστοποιείται όταν οι προβολές του $\boldsymbol{y}$ στους υποχώρους $\mathcal{S}(\boldsymbol{X})$ και $\mathcal{S}^{\perp}(\boldsymbol{X})$ είναι κάθετες.

Πιο συγκεκριμένα ο $(n \times n)$ πίνακας $\boldsymbol{P_X}$ προβάλει κάθετα το $\boldsymbol{y}$ στον «υποχώρο των προσαρμοσμένων τιμών» $\mathcal{S}(\boldsymbol{X})$, και ο $(n \times n)$ πίνακας $\boldsymbol{M_X}$ προβάλει κάθετα το $\boldsymbol{y}$ στον «υποχώρο των καταλοίπων» $\mathcal{S}^{\perp}(\boldsymbol{X})$.

Ένας τετραγωνικός και συμμετρικός πίνακας $\boldsymbol{A}$ ονομάζεται *ταυτοδύναμος* (idempotent) αν $\boldsymbol{A} = \boldsymbol{AA}$, δηλαδή αν ισούται με το τετράγωνό του. Οι πίνακες $\boldsymbol{P_X}$ και $\boldsymbol{M_X}$ είναι ταυτοδύναμοι αφού είναι τετραγωνικοί $(n \times n)$, συμμετρικοί (δηλαδή $\boldsymbol{P_X} = \boldsymbol{P_X}'$ και $\boldsymbol{M_X} = \boldsymbol{M_X}'$), και

$$\begin{aligned} \boldsymbol{P_X P_X} &= [\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}X'][\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'] = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' = \boldsymbol{P_X} \\ \boldsymbol{M_X M_X} &= [\boldsymbol{I} - \boldsymbol{P_X}][\boldsymbol{I} - \boldsymbol{P_X}] = \boldsymbol{I} - 2\boldsymbol{P_X} + \boldsymbol{P_X P_X} = \boldsymbol{I} - \boldsymbol{P_X} = \boldsymbol{M_X}. \end{aligned}$$

Η γεωμετρική ερμηνεία αυτής της ιδιότητας των $\boldsymbol{P_X}$ και $\boldsymbol{M_X}$ είναι απλή. Εφαρμόζοντας τον πίνακα $\boldsymbol{P_X}$ στο διάνυσμα $\boldsymbol{y}$ λαμβάνουμε την προβολή του $\boldsymbol{y}$ στον υποχώρο $\mathcal{S}(\boldsymbol{X})$, δηλαδή παίρνουμε το διάνυσμα των προσαρμοσμένων τιμών $\hat{\boldsymbol{y}} = \boldsymbol{P_X y}$. Επαναεφαρμόζοντας τον πίνακα $\boldsymbol{P_X}$ στο διάνυσμα $\boldsymbol{P_X y}$ όμως, δεν έχει κανένα αποτέλεσμα αφού το $\boldsymbol{P_X y}$ ανήκει ήδη στον υποχώρο $\mathcal{S}(\boldsymbol{X})$, δηλαδή $\boldsymbol{P_X P_X y} = \boldsymbol{P_X y}$. Αντίστοιχα, εφαρμόζοντας τον πίνακα $\boldsymbol{M_X}$ στο διάνυσμα $\boldsymbol{y}$ λαμβάνουμε την προβολή του $\boldsymbol{y}$ στον υποχώρο $\mathcal{S}^{\perp}(\boldsymbol{X})$, δηλαδή παίρνουμε το διάνυσμα των καταλοίπων $\hat{\boldsymbol{u}} = \boldsymbol{M_X y}$, ενώ επανεφαρμόζοντας τον $\boldsymbol{M_X}$ στο $\boldsymbol{M_X y}$ δεν έχει κανένα αποτέλεσμα αφού το $\boldsymbol{M_X y}$ ανήκει ήδη στον υποχώρο $\mathcal{S}^{\perp}(\boldsymbol{X})$, δηλαδή $\boldsymbol{M_X M_X y} = \boldsymbol{M_X y}$.

Οι δύο υποχώροι του $\mathbb{R}^n$, $\mathcal{S}(\boldsymbol{X})$ και $\mathcal{S}^{\perp}(\boldsymbol{X})$ είναι κάθετοι μεταξύ τους, αφού οι πίνακες που προβάλουν διανύσματα του $\mathbb{R}^n$ σε αυτούς είναι κάθετοι, δηλαδή

$$\boldsymbol{P_X M_X} = \boldsymbol{P_X}[\boldsymbol{I} - \boldsymbol{P_X}] = \boldsymbol{P_X} - \boldsymbol{P_X P_X} = \boldsymbol{O}, \tag{4.26}$$

όπου $\boldsymbol{O}$ είναι ο $n \times n$ μηδενικός πίνακας. Γεωμετρικά αυτό σημαίνει ότι το μόνο κοινό σημείο των υποχώρων $\mathcal{S}(\boldsymbol{X})$ και $\mathcal{S}^{\perp}(\boldsymbol{X})$ είναι η αρχή των αξόνων $\boldsymbol{O}$.

3.5. Ο Συντελεστής Πολλαπλού Προσδιορισμού

Χρησιμοποιώντας την καθετότητα των προβολικών πινάκων P_X και M_X μπορούμε να αναλύσουμε το τετράγωνο του Ευκλείδειου μήκους του $\boldsymbol{y}$, $||\boldsymbol{y}||$, ως

$$\begin{aligned} ||\boldsymbol{y}||^2 = \boldsymbol{y}'\boldsymbol{y} &= (\hat{\boldsymbol{y}} + \hat{\boldsymbol{u}})'(\hat{\boldsymbol{y}} + \hat{\boldsymbol{u}}) \\ &= (\boldsymbol{P_X y} + \boldsymbol{M_X y})'(\boldsymbol{P_X y} + \boldsymbol{M_X y}) \\ &= \boldsymbol{y}'\boldsymbol{P_X y} + \boldsymbol{y}'\boldsymbol{M_X y} \\ &= ||\boldsymbol{P_X y}||^2 + ||\boldsymbol{M_X y}||^2 \\ &= ||\hat{\boldsymbol{y}}||^2 + ||\hat{\boldsymbol{u}}||^2, \end{aligned}$$

ή

$$\Sigma\, y_i^2 = \Sigma\, \hat{y}_i^2 + \Sigma\, \hat{u}_i^2. \tag{5.27}$$

Η σχέση αυτή δεν είναι τίποτε άλλο παρά το Πυθαγόρειο Θεώρημα.

Αυτό που μας ενδιαφέρει όμως στην πράξη είναι η ανάλυση της *διακύμανσης* της εξαρτημένης y. Γνωρίζουμε ότι

$$\Sigma\,(y_i - \bar{y})^2 = \Sigma\, y_i^2 - n\bar{y}^2 \tag{5.28}$$

και

$$\Sigma\,(\hat{y}_i - \bar{\hat{y}})^2 = \Sigma\,\hat{y}_i^2 - n\bar{\hat{y}}^2. \tag{5.29}$$

Αλλά αφού $\bar{\hat{y}} = \bar{y}$ και $\bar{\hat{u}} = 0$, η (5.27) μπορεί να γραφεί ως

$$\Sigma\,(y_i - \bar{y})^2 = \Sigma\,(\hat{y}_i - \bar{y})^2 + \Sigma\,\hat{u}_i^2. \tag{5.30}$$

ή

$$\frac{\Sigma\,(y_i - \bar{y})^2}{n-1} = \frac{\Sigma\,(\hat{y}_i - \bar{y})^2}{n-1} + \frac{\Sigma\,\hat{u}_i^2}{n-1}. \tag{5.31}$$

Βλέπουμε, λοιπόν, ότι η συνολική διακύμανση της y αναλύεται στο ερμηνευόμενο από την παλινδρόμηση μέρος που είναι η διακύμανση των $\hat{y}_i$, και το ανερμήνευτο μέρος που είναι η διακύμανση των καταλοίπων $\hat{u}_i$. Μπορούμε να γράψουμε την (5.30) ως

$$\text{TSS} = \text{ESS} + \text{RSS} \tag{5.32}$$

όπου TSS (Total Sum of Squares) είναι το συνολικό άθροισμα τετραγώνων, ESS (Explained Sum of Squares) είναι το ερμηνευόμενο άθροισμα τετραγώνων, και RSS (Residual Sum of Squares) είναι το ανερμήνευτο άθροισμα τετραγώνων.

Ο *συντελεστής πολλαπλού προσδιορισμού* R^2 ορίζεται ως ο λόγος του μέρους της διακύμανσης της Y που ερμηνεύεται από την παλινδρόμηση, προς τη συνολική διακύμανση της Y, δηλαδή

$$\begin{aligned} R^2 &= \frac{\Sigma\,(\hat{y}_i - \bar{y})^2}{\Sigma\,(y_i - \bar{y})^2} = \frac{ESS}{TSS} \\ &= 1 - \frac{\Sigma\,\hat{u}_i^2}{\Sigma\,(y_i - \bar{y})^2} = 1 - \frac{RSS}{TSS}. \end{aligned} \tag{5.33}$$

Είναι φανερό ότι

$$0 \leq R^2 \leq 1. \tag{5.34}$$

Ο συντελεστής πολλαπλού προσδιορισμού R^2 είναι ένα μέτρο της ερμηνευτικής ικανότητας της παλινδρόμησης και όσο πλησιάζει προς την μονάδα τόσο καλύτερα το γραμμικό μας υπόδειγμα προσαρμόζεται στα διαθέσιμα στατιστικά στοιχεία. Πρέπει όμως να είμαστε προσεκτικοί στην ερμηνεία του R^2 όταν το χρησιμοποιούμε για να αξιολογήσουμε μια παλινδρόμηση ή για να συγκρίνουμε δύο ή περισσότερες παλινδρομήσεις μεταξύ τους. Σχετικά πρέπει να επισημάνουμε τα εξής:

i. Η εισαγωγή πρόσθετων ερμηνευτικών μεταβλητών στην παλινδρόμηση αυξάνει το ερμηνευόμενο άθροισμα τετραγώνων ESS και μειώνει το ανερμήνευτο άθροισμα τετραγώνων RSS, δηλαδή αυξάνει το R^2. Η ιδιότητα αυτή είναι καθαρά μαθηματική και δεν εξαρτάται από το εάν οι επιπλέον ερμηνευτικές μεταβλητές που εισήχθησαν είναι σχετικές με την αιτιώδη σχέση που εκφράζεται από τη παλινδρόμηση ή όχι.

ii. Γενικά, αν αντί της Y χρησιμοποιήσουμε ως εξαρτημένη μεταβλητή κάποιο γραμμικό μετασχηματισμό

3.6. Το Στατιστικό Υπόδειγμα της Πολλαπλής Παλινδρόμησης

Για την προσαρμογή του πολυεπιπέδου των ελαχίστων τετραγώνων $\hat{y} = X\hat{b}$ δεν χρησιμοποιήσαμε καμία υπόθεση εκτός από το ότι ο πίνακας X είναι πλήρους βαθμού. Αν επιθυμούμε να προχωρήσουμε σε στατιστική επαγωγή για τις αληθείς τιμές των παραμέτρων $\beta_0, \beta_1, ..., \beta_k$ στον πληθυσμό, τότε απαιτούνται περαιτέρω υποθέσεις.

Υπόθεση 1: Γραμμικότητα / Σωστός Προσδιορισμός.

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}. \tag{υ.1}$$

Η υπόθεση αυτή είναι διτή. Πρώτον, η (υ.1) εκφράζει την πίστη μας ότι η πραγματική συναρτησιακή σχέση που συνδέει την εξαρτημένη μεταβλητή $\boldsymbol{y}$ με τις k ανεξάρτητες μεταβλητές $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_k)$ στον πληθυσμό είναι γραμμική ως προς τις παραμέτρους $\beta = (\beta_0, \beta_1, ..., \beta_k)'$. Αν αυτό ισχύει λέμε ότι η παλινδρόμηση του y πάνω στα X είναι γραμμική (liner regression). Δεύτερον, η (υ.1) λέει ότι οι $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_k)$ είναι ακριβώς οι μεταβλητές που χρειάζονται για να εξηγηθεί η $\boldsymbol{y}$, και ότι δεν έχει παραληφθεί καμία μεταβλητή που θα έπρεπε να συμπεριληφθεί, ούτε έχουν προστεθεί άσχετες με την $\boldsymbol{y}$ μεταβλητές. Αν αυτό ισχύει, λέμε ότι το υπόδειγμα είναι σωστά προσδιορισμένο (correctly specified model).

Υπόθεση 2: Εξωγένεια των $\boldsymbol{X}$ / Καθετότητα $\boldsymbol{u}$ και $\boldsymbol{X}$.

$$\boldsymbol{u} \perp \boldsymbol{X} \quad \text{ή} \quad E(\boldsymbol{u}|\boldsymbol{X}) = 0. \tag{υ.2}$$

Η υπόθεση αυτή λέει ότι ο διαταρακτικός όρος u είναι ανεξάρτητος από (ορθογώνιος προς) τις επεξηγητικές μεταβλητές $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_k)$ που περιλαμβάνονται στο υπόδειγμα. Αν αυτό είναι αλήθές λέμε ότι οι $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_k)$ είναι εξωγενείς μεταβλητές (exogenous variables). Η αναγκαιότητα και σημασία της υπόθεσης αυτής θα συζητηθεί αργότερα όταν θα δούμε πότε και γιατί δεν ικανοποιείται, δηλαδή πότε και γιατί τουλάχιστον ένα από τα X μπορεί να είναι ενδογενές (endogenuous). Προς το παρόν, θα την εκλάβουμε ότι σημαίνει κάτι πολύ πιο απλό, δηλαδή ότι ο μέσος των σφαλμάτων τις παλινδρόμησης είναι μηδέν, κάτι το οποίο είναι πάντα αληθές αν στο υπόδειγμά μας έχουμε συμπεριλάβει την σταθερά $\boldsymbol{x}_0 \equiv \boldsymbol{1}_n$, όπου το $\boldsymbol{1}_n$ είναι το $n \times 1$ διάνυσμα με όλα τα στοιχεία του ίσα με την μονάδα.

Αν η (υ.2) ισχύει τότε

$$E(\boldsymbol{y}|\boldsymbol{X}) = E(\boldsymbol{X\beta} + \boldsymbol{u}|\boldsymbol{X}) = \boldsymbol{X\beta} + E(\boldsymbol{u}|\boldsymbol{X}) = \boldsymbol{X\beta},$$

δηλαδή ο υπό συνθήκη μέσος (conditional mean) της $\boldsymbol{y}$ δεδομένων των $\boldsymbol{X}$ ισούται με $\boldsymbol{X\beta}$. Σε αυτή την μορφή της, η (υ.2) σημαίνει ότι το $\boldsymbol{X\beta}$ είναι η προσδοκώμενη τιμή του y αν τα άτομα υπό εξέταση στο δείγμα έχουν χαρακτηριστικά X. Π.χ., αν y_i είναι ο μισθός που λαμβάνει το άτομο i και $\boldsymbol{x}_i = (1, x_{i1}, \cdots, x_{ik})'$ είναι το διάνυσμα των χαρακτηριστικών του (σταθερά, μόρφωση, εμπειρία, φύλο κ.τ.λ.), τότε $\boldsymbol{x}_i'\boldsymbol{\beta}$ είναι ο προσδοκώμενος μισθός του και $u_i = y_i - x_i'\beta$ είναι η απόκλιση του παρατηρούμενου από τον προσδοκώμενο μισθό.

<u>Υπόθεση 3:</u> Σφαιρικά Σφάλματα.

$$V(\boldsymbol{u}|\boldsymbol{X}) = \sigma^2 \boldsymbol{I}_n, \tag{υ.3}$$

όπου $V(\boldsymbol{u}|\boldsymbol{X})$ είναι ο υπό συνθήκη πίνακας διακυμάνσεων-συνδιακυμάνσεων (conditional variance-covariance matrix) των $\boldsymbol{u}$ δεδομένων των $\boldsymbol{X}$, και $\boldsymbol{I}_n$ είναι ο $n \times n$ μοναδιαίος πίνακας. Αναλυτικά η υπόθεση αυτή γράφεται ως

$$V(\boldsymbol{u}|\boldsymbol{X}) = E(\boldsymbol{uu}'|X) = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$$

και αποτελεί συντομογραφία των δύο επιμέρους υποθέσεων:

$$Var(u_i|\boldsymbol{x}_i) = E(u_i^2|\boldsymbol{x}_i) = \sigma^2, \qquad i = 1, 2, ..., n \qquad (υ.3α)$$

$$Cov(u_i, u_j|\boldsymbol{x}_i, \boldsymbol{x}_j) = E(u_i u_j|\boldsymbol{x}_i, \boldsymbol{x}_j) = 0, \quad i, j = 1, 2, ..., n, \ (i \neq j) \qquad (υ.3β)$$

Σύμφωνα με την (υ.3α) τα σφάλματα $u_1, u_2, ..., u_n$ έχουν την ίδια σταθερή διακύμανση σ^2, δηλαδή είναι ομοσκεδαστικά (homoskedastic errors)[1]. Επιπλέον, σύμφωνα με την (υ.3β) δύο διαφορετικά σφάλματα u_i και u_j, $i, j = 1, 2, ..., n, i \neq j$, είναι ασυσχέτιστα, οπότε λέμε ότι τα σφάλματα δεν είναι αυτοσυσχετισμένα (no autocorrelation). Όταν τα σφάλματα είναι ομοσκεδαστικά και ασυσχέτιστα λέμε ότι είναι σφαιρικά (spherical errors), ενώ αν είναι είτε ετεροσκεδαστικά (heteroskedastic) είτε αυτοσυχετισμένα (autocorrelated) λέμε ότι είναι μη-σφαιρικά (non-spherical errors).

Υπόθεση 4: Όχι Τέλεια Πολυσυγραμμικότητα.

$$r(\boldsymbol{X}) = k + 1. \qquad (υ.4)$$

Η υπόθεση αυτή είναι αναγκαία για να έχει λύση το σύστημα των κανονικών εξισώσεων και να ορίζονται οι εκτιμήτριες ελαχίστων τετραγώνων.

Υπόθεση 5: Όχι Σφάλματα Μέτρησης.

Οι μεταβλητές $\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_k$ έχουν μετρηθεί χωρίς σφάλμα. (υ.5)

Η υπόθεση αυτή λέει ότι οι ανεξάρτητες μεταβλητές στο υπόδειγμα έχουν μετρηθεί χωρίς σφάλματα (no errors in variables). Η σημασία της υπόθεσης αυτής θα συζητηθεί αργότερα όταν θα εξετάσουμε περιπτώσεις παραβίασής της.

Υπόθεση 6: Κανονικά Σφάλματα

$$\boldsymbol{u} \sim N(0, \sigma^2 \boldsymbol{I}_n). \qquad (υ.6)$$

[1] Η λέξη σκέδαση χρησιμοποιείται εδώ ως συνώνυμο της λέξης διακύμανση.

Η υπόθεση αυτή είναι αναγκαία για να κάνουμε επαγωγή όταν το δείγμα μας είναι μικρό. Σε μεγάλα δείγματα η επαγωγή βασίζεται στην χρήση του Κεντρικού Οριακού Θεωρήματος (ΚΟΘ). Η αντίστοιχη υπόθεση για τις εξαρτημένες μεταβλητές $y_1, y_2, ..., y_n$ είναι ότι

$$\boldsymbol{y} \sim N(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{I}_n).$$

3.7. Στατιστικές Ιδιότητες των Συντελεστών Ελαχίστων Τετραγώνων

Γράφουμε

$$\begin{aligned}
\hat{b} &= (X'X)^{-1}X'y \\
&= (X'X)^{-1}X'(X\beta + u) \\
&= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'u \\
&= \beta + (X'X)^{-1}X'u.
\end{aligned}$$

οπότε

$$\begin{aligned}
E(\hat{b}|X) &= \beta + (X'X)^{-1}X'E(u|X) \\
&= \beta
\end{aligned}$$

αφού από την (υ.2), $E(u|X) = 0$.

Ο πίνακας διακυμάνσεων-συνδιακυμάνσεων του $k + 1$ διανύσματος $\hat{b}$ είναι

$$\begin{aligned}
V(\hat{b}) &= Var(\hat{b}|X) \\
&= E[(\hat{b} - \beta)(\hat{b} - \beta)'|X] \\
&= E[(X'X)^{-1}X'uu'X(X'X)^{-1}|X] \\
&= (X'X)^{-1}X' \; E[uu'|X] \; X(X'X)^{-1} \\
&= (X'X)^{-1}X' \; \sigma^2 I_n \; X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}. \qquad (7.35)
\end{aligned}$$

Τα διαγώνια στοιχεία του $(k+1)\times(k+1)$ πίνακα $V(\hat{b})$ είναι οι διακυμάνσεις των επιμέρους εκτιμητριών $\hat{b}_j$, $j = 0, ..., k$, ενώ τα εκτός διαγωνίου στοιχεία είναι οι συνδιακυμάνσεις των εκτιμητριών $\hat{b}_j$ και $\hat{b}_{j'}$, $j \neq j'$, δηλαδή

$$V(\hat{b}) = \begin{bmatrix} Var(\hat{b}_0) & Cov(\hat{b}_0, \hat{b}_1) & \cdots & Cov(\hat{b}_0, \hat{b}_k) \\ Cov(\hat{b}_1, \hat{b}_0) & Var(\hat{b}_1) & \cdots & Cov(\hat{b}_1, \hat{b}_k) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(\hat{b}_k, \hat{b}_0) & Cov(\hat{b}_k, \hat{b}_1) & \cdots & Var(\hat{b}_k) \end{bmatrix} \tag{7.36}$$

Είναι φανερό ότι ο πίνακας $V(\hat{b})$ είναι συμμετρικός και θετικά ορισμένος.

Για να υπολογίσουμε τον πίνακα $V(\hat{b})$ στην πράξη πρέπει να έχουμε μία εκτιμήτρια της διακύμανσης των σφαλμάτων σ^2. Έχουμε

$$\begin{aligned} E(\hat{u}'\hat{u}|X) &= E(u'M_X u|X) \\ &= E[\mathrm{tr}\,(u'M_X u)|X] && [\text{διότι το } u'M_X u \text{ είναι βαθμωτό}] \\ &= E[\mathrm{tr}\,(M_X uu')|X] && [\text{διότι } \mathrm{tr}\,(AB) = \mathrm{tr}\,(BA)] \\ &= \mathrm{tr}\,[E(M_X uu'|X)] \\ &= \mathrm{tr}\,[M_X E(uu'|X)] \\ &= \sigma^2\,\mathrm{tr}\,(M_X) \\ &= \sigma^2\,\mathrm{tr}\,[I_n - X(X'X)^{-1}X'] \\ &= \sigma^2\{\mathrm{tr}\,(I_n) - \mathrm{tr}\,[X(X'X)^{-1}X']\} && [\text{διότι } \mathrm{tr}\,(A+B) = \mathrm{tr}\,(A) + \mathrm{tr}\,(B)] \\ &= \sigma^2\{n - \mathrm{tr}\,[X'X(X'X)^{-1}]\} \\ &= \sigma^2\{n - \mathrm{tr}\,(I_{k+1})\} \\ &= \sigma^2\{n - (k+1)\}. \end{aligned}$$

’ρα, μία αμερόληπτη εκτιμήτρια του σ^2 είναι

$$\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{n-(k+1)}, \tag{7.37}$$

και ο πίνακας διακυμάνσεων-συνδιακυμάνσεων του $\hat{b}$ μπορεί να εκτιμηθεί από

$$\hat{V}(\hat{b}) = \hat{\sigma}^2(X'X)^{-1}. \tag{7.38}$$

4. ΠΑΡΑΒΙΑΣΕΙΣ ΤΩΝ ΥΠΟΘΕΣΕΩΝ ΤΟΥ ΚΛΑΣΙΚΟΥ ΥΠΟΔΕΙΓΜΑΤΟΣ

4.1. Εισαγωγή

4.2. Η Υπόθεση της Γραμμικότητας

4.3. Λάθη στον Προσδιορισμό του Υποδείγματος

Οι συνέπειες από την παράλειψη σημαντικών ερμηνευτικών μεταβλητών ή από την εισαγωγή άσχετων ερμηνευτικών μεταβλητών σε μια παλινδρόμηση αποτελούν μέρος του γενικότερου προβλήματος του εσφαλμένου προσδιορισμού του υποδείγματος προς εκτίμηση.

4.3.1. Παράλειψη σημαντικών ερμηνευτικών μεταβλητών

Ας υποθέσουμε ότι η αληθής παλινδρόμηση που ερμηνεύει την Y είναι η

$$y = X_1 b_1 + X_2 b_2 + u, \tag{3.1}$$

και ότι είτε από λάθος είτε από αδυναμία συγκέντρωσης των απαραίτητων στατιστικών στοιχείων για τις μεταβλητές X_2, ο ερευνητής εκτιμά την ελλειπή παλινδρόμηση

$$y = X_1 a_1 + e. \tag{3.2}$$

Η εκτιμήτρια Ε.Τ. από την ελλειπή παλινδρόμηση είναι, κατά τα γνωστά, η

$$\begin{aligned}
\hat{a}_1 &= (X_1' X_1)^{-1} X_1' y \\
&= (X_1' X_1)^{-1} X_1' (X_1 b_1 + X_2 b_2 + u) \\
&= b_1 + (X_1' X_1)^{-1} X_1' X_2 b_2 + (X_1' X_1)^{-1} X_1' u \\
&= b_1 + \hat{D} b_2 + (X_1' X_1)^{-1} X_1' u
\end{aligned}$$

όπου $\hat{D} = (X_1'X_1)^{-1}X_1'X_2$. Εύκολα προκύπτει λοιπόν ότι

$$E(\hat{a}_1) = b_1 + \hat{D}b_2, \tag{3.3}$$

και άρα η $\hat{a}_1$ είναι γενικά μεροπληπτική εκτιμήτρια του υποδιανύσματος των παραμέτρων b_1.

Πιο αναλυτικά, ο πίνακας $\hat{D}$ είναι η εκτιμήτρια των συντελεστών της παλινδρόμησης

$$X_2 = DX_1 + U \tag{3.4}$$

των μεταβλητών X_2 πάνω στις μεταβλητές X_1. Άρα η μεροληψία της $\hat{a}_1$ είναι μηδέν εάν και μόνο εάν $\hat{D}b_2 = 0$, δηλαδή αν $\hat{D} = 0$, αφού σύμφωνα με την υπόθεσή μας $b_2 \neq 0$. Αυτό συμβαίνει αν οι μεταβλητές X_1 και X_2 είναι γραμμικά ασυσχέτιστες μεταξύ τους, δηλαδή είναι ορθογώνιες και έχουμε $X_1'X_2 = 0$.

Αν όμως, όπως συμβαίνει στις περισσότερες φορές, οι X_1 και X_2 είναι συσχετισμένες, η παράλειψη των μεταβλητών X_2 δημιουργεί μεροληψία στην $\hat{a}_1$ αφού η εκτιμήτρια αυτή μετρά τόσο την άμεση επίδραση των X_1 επάνω στο Y, (δηλαδή το b_1), όσο και την έμμεση επίδρασή τους μέσω της συσχέτισης τους με τις μεταβλητές X_2 (δηλαδή το $\hat{D}b_2$), δηλαδή οι X_1 λειτουργούν και ως proxies των απόντων μεταβλητών X_2.

Για παράδειγμα, αν το Y είναι ο μισθός ενός ατόμου, X_1 είναι η μόρφωσή του, και X_2 είναι το IQ του, τότε αν άτομα με υψηλότερο IQ τείνουν να είναι και πιο μορφωμένα, η παράλειψη του IQ θα δημιουργήσει μεροληψία στον συντελεστή της μόρφωσης, αφού κομμάτι αυτού που η παλινδρόμηση θα αποδίδει στην μόρφωση είναι στην πραγματικότητα αποτέλεσμα του IQ. Αυτό συμβαίνει συχνά σε μελέτες μισθών όπου τα στοιχεία για το IQ των συμμετασχόντων δεν είναι διαθέσιμα, με αποτέλεσμα *η επίδραση της μόρφωσης στους προσδοκώμενους μισθούς να υπερεκτιμάται*. Θα επιστρέψουμε σε αυτό το παράδειγμα όταν θα παρουσιάσουμε την μέθοδο των βοηθητικών μεταβλητών.

4.3.2. Εισαγωγή άσχετων ερμηνευτικών μεταβλητών

Έστω τώρα ότι η αληθής παλινδρόμηση είναι η

$$y = X_1b_1 + u \tag{3.5}$$

από την οποία λαμβάνουμε

$$\hat{b}_1 = (X_1'X_1)^{-1}X_1y, \quad E(\hat{b}_1) = b_1, \quad V(\hat{b}_1) = \sigma^2(X_1'X_1)^{-1}. \tag{3.6}$$

Υποθέτουμε τώρα ότι ο ερευνητής εισάγει ορισμένες άσχετες μεταβλητές X_2 και εκτιμά την παλινδρόμηση

$$y = X_1a_1 + X_2a_2 + e. \tag{3.7}$$

Το υπόδειγμα στην (3.7) μπορεί να γραφτεί ως:

$$y = [X_1 \;\; X_2]\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + e. \tag{3.8}$$

Οι κανονικές εξισώσεις για την εκτίμηση των a_1 και a_2 από την εξίσωση (3.8) είναι:

$$\begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}\begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \end{bmatrix} = \begin{bmatrix} X_1'y \\ X_2'y \end{bmatrix}. \tag{3.9}$$

από την επίλυση των οποίων προκύπτει ότι:

$$\hat{a}_1 = (X_1'Q_2X_1)^{-1}X_1'Q_2y \tag{3.10}$$

όπου

$$Q_2 = I - X_2(X_2'X_2)^{-1}X_2' \;\; \text{και} \;\; Q_2' = Q_2, \; Q_2^2 = Q_2, \; Q_2X_2 = 0. \tag{3.11}$$

Από την (3.10) εύκολα προκύπτει ότι:

$$E(\hat{a}_1) = b_1, \quad V(\hat{a}_1) = \sigma^2(X_1'Q_2X_1)^{-1}. \tag{3.12}$$

Από τις (3.6) και (3.12) έχουμε

$$V(\hat{a}_1) - V(\hat{b}_1) = \sigma^2[(X_1'Q_2X_1)^{-1} - (X_1'X_1)^{-1}]. \tag{3.13}$$

Αλλά

$$\begin{aligned} X_1'Q_2X_1 &= X_1'[I - X_2(X_2'X_2)^{-1}X_2']X_1 \\ &= X_1'X_1 - X_1'X_2(X_2'X_2)^{-1}X_2'X_1. \end{aligned} \tag{3.14}$$

Αν θέσουμε

$$A = X_1'X_1, \quad B = X_1'Q_2X_1 \quad \text{και} \quad C = X_1'X_2(X_2'X_2)^{-1}X_2'X_1$$

τότε

$$A = B + C$$

και επειδή οι πίνακες A και B είναι θετικά ορισμένοι και ο πίνακας C είναι μη αρνητικά ορισμένος, ο πίνακας

$$B^{-1}A^{-1} = (X_1'Q_2X_1)^{-1} - (X_1'X_1)^{-1} \tag{3.15}$$

που εμφανίζεται στην (3.13) είναι μη αρνητικά ορισμένος και άρα

$$V(\hat{a}_1) - V(\hat{b}_1) \geq 0. \tag{3.16}$$

Η ισότητα θα ισχύει αν $X_2 = 0$, ή αν οι X_1 και X_2 είναι ορθογώνιες, δηλαδή $X_1'X_2 = 0$. Σε κάθε άλλη περίπτωση η διακύμανση του $\hat{a}_1$ θα είναι μεγαλύτερη από αυτή του $\hat{b}_1$.

Συμπεραίνουμε λοιπόν ότι η εισαγωγή άσχετων ερμηνευτικών μεταβλητών δεν επηρεάζει την αμεροληψία της εκτιμήτριας $\hat{a}_1$, δηλαδή $E(\hat{a}_1) = b_1$, αλλά διογκώνει την διακύμανσή της, δηλαδή η $\hat{a}_1$ δεν είναι αποτελεσματική εκτιμήτρια του b_1 (εκτός εάν οι άσχετες μεταβλητές X_2 είναι ασυσχέτιστες με τις X_1).