1 Εκτιμητής Γενικευμένων Ελαχίστων Τετραγώνων (Generalized Least Squares (GLS) estimator)

Θεωρήστε το γραμμικό μοντέλο (γραμμική παλιδρόμηση) με πιθανή έλλειψη ομοσκεδαστικότητας ή/και ύπαρξη αυτοσυσχέτισης των σφαλμάτων (και άρα και των παρατηρήσεων):

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},\ \boldsymbol{\epsilon} \sim (\mathbf{0}, \boldsymbol{\Sigma}), \tag{1.1}$$

όπου $\mathbf{Y}$ είναι ένα $n \times 1$ διάνυσμα αποκρίσεων, $\mathbf{X}$ ο $n \times p$ πίνακας σχεδιασμού ή συμμεταβλητών (η πρώτη στήλη είναι ένα διάνυσμα μονάδων όταν έχουμε σταθερά (intercept) στην παλινδρόμηση), $\boldsymbol{\beta}$ είναι το $p \times 1$ διάνυσμα παραμέτρων, και $\boldsymbol{\epsilon}$ είναι ένα $n \times 1$ διάνυσμα σφαλμάτων. Επίσης υποθέτουμε πως $rank(\boldsymbol{X}) = p$, δηλαδή ο $\boldsymbol{X}$ έχει γραμμικά ανεξάρτητες στήλες.

Ο Πίνακας $\boldsymbol{\Sigma}$ είναι ο $n \times n$ πίνακας διακυμάνσεων-συνδυακυμάνσεων των σφαλμάτων, με i, j στοιχείο $\sigma_{i,j} = cov(Y_i, Y_j)$, όπου $i = 1, \ldots, n$ και $j = 1, \ldots, n$. Υποθέτουμε πως ο $\boldsymbol{\Sigma}$ είναι θετικά ορισμένος (p.d.). Σημείωση: όταν $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_n$ έχουμε την κλασσική παλινδρόμηση (δηλ. έχουμε ομοσκεδαστικά και ασυσχέτιστα σφάλματα).

Έχουμε δείξει πως ο εκτιμητής ελαχίστων τεραγώνων $\hat{\boldsymbol{\beta}}_{LS} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$ είναι Βέλτιστος Γραμμικός Αμεροληπτος Εκτιμητής (BLUE) του $\boldsymbol{\beta}$ όταν $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_n$. Το ερώτημα που θα απαντήσουμε τώρα είναι: ποιός είναι ο (BLUE) εκτιμητής του $\boldsymbol{\beta}$ στην γενική περίπτωση.

Θεώρημα 1. *Ο BLUE του* $\boldsymbol{\beta}$ *ισούται μϵ* $\hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{Y}$

Απόδειξη. Έχουμε δείξει πως η τεραγωνική ρίζα του p.d. πίνακα $\boldsymbol{\Sigma}$ (με φασματική ανάλυση $\boldsymbol{\Sigma} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T$) υπάρχει, είναι μοναδικός. και δίνεται από τον τύπο $\boldsymbol{\Sigma}^{1/2} = \boldsymbol{P}\boldsymbol{\Lambda}^{1/2}\boldsymbol{P}^T$. Ο αντίστροφος του ισούται με $\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{P}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{P}^T$. Πολλαπλασιάζουμε από δεξιά τα δύο μέλη της εξίσωσης του μοντέλου στο (1.1) με τον πίνακα $\boldsymbol{\Sigma}^{-1/2}$ κι έχουμε :

$$\mathbf{Y}^* = \mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\epsilon}^*,\ \boldsymbol{\epsilon}^* \sim (\mathbf{0}, I_n), \tag{1.2}$$

όπου

$$\begin{aligned} \boldsymbol{Y}^* &= \boldsymbol{\Sigma}^{-1/2}\boldsymbol{Y} \\ \boldsymbol{X}^* &= \boldsymbol{\Sigma}^{-1/2}\boldsymbol{X} \\ \boldsymbol{\epsilon}^* &= \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\epsilon}. \end{aligned}$$

Είναι προφανές πως $Var(\boldsymbol{\epsilon}^*) = Var(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\epsilon}) = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{I}_n$. Με αυτόν τον μετασχηματισμό καταλήγουμε από το μοντέλο στο (1.1) σε ένα ισοδύναμο μοντέλο κλασσικής παλινδρόμησης στο (1.2). Ο BLUE του $\boldsymbol{\beta}$ προφανώς είναι ο εκτιμητής

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{GLS} &= (\mathbf{X}^{*T}\mathbf{X}^*)^{-1}\mathbf{X}^{*T}\mathbf{Y}^* \\
&= (\mathbf{X}^T\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{-1/2}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{-1/2}\mathbf{Y} \\
&= (\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{Y}. \qquad (1.3)
\end{aligned}$$

□

Συνήθως γράφουμε τον πίνακα διακυμάνσεων-συνδιακυμάνσεων ως $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{V}$, όπου $\sigma^2 > 0$. Τότε έχουμε

$$\hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{X}^T\boldsymbol{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{V}^{-1}\mathbf{Y} \qquad (1.4)$$

1.1 Ιδιότητες του Εκτιμητή Γενικευμένων Ελαχίστων Τετραγώνων (GLS) - Σύγκριση με τον Εκτιμητή Ελαχίστων Τετραγώνων (LS)

Για τον $\hat{\boldsymbol{\beta}}_{GLS}$ έχουμε

$$\begin{aligned}
E(\hat{\boldsymbol{\beta}}_{GLS}) &= \boldsymbol{\beta} \\
Var(\hat{\boldsymbol{\beta}}_{GLS}) &= Var((\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{Y}) = (\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}^T\boldsymbol{V}^{-1}\mathbf{X})^{-1} \qquad (1.5)
\end{aligned}$$

Από την άλλη, για τον $\hat{\boldsymbol{\beta}}_{LS}$ έχουμε

$$\begin{aligned}
E(\hat{\boldsymbol{\beta}}_{LS}) &= \boldsymbol{\beta} \\
Var(\hat{\boldsymbol{\beta}}_{LS}) &= Var((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}) = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\boldsymbol{\Sigma}\mathbf{X})(\mathbf{X}^T\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\boldsymbol{V}\mathbf{X})(\mathbf{X}^T\mathbf{X})^{-1} \qquad (1.6)
\end{aligned}$$

<u>Παράδειγμα 1:</u> Y_i είναι δειγματικοί μέσοι που προκύπτουν από k_i ανεξάρτητες παρατηρήσεις στο i-οστό εργαστήριο ($i = 1, \ldots, n$).

$$\begin{aligned}
Y_i &= \boldsymbol{x}_i^T\boldsymbol{\beta} + \epsilon_i \\
\epsilon_i &\sim^{ind} N(0, \sigma^2\frac{1}{k_i})
\end{aligned}$$

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1^T \\ \boldsymbol{x}_2^T \\ \vdots \\ \boldsymbol{x}_n^T \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1,p-1} \\ 1 & x_{21} & x_{22} & \ldots & x_{2,p-1} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{n,p-1} \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \sigma^2 \begin{bmatrix} \frac{1}{k_1} & 0 & \ldots & 0 \\ 0 & \frac{1}{k_2} & \ldots & 0 \\ \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & \ldots & \frac{1}{k_n} \end{bmatrix}$$

Εδώ ισχύει πως $\boldsymbol{V}$ είναι γνωστός (πράγμα σπάνιο).

Παράδειγμα 2: Παλινδρόμηση με AR(1) σφάλματα.

$$\begin{aligned} Y_t &= \boldsymbol{x}_t^T \boldsymbol{\beta} + \epsilon_t \\ \epsilon_t &= \phi \epsilon_{t-1} + \nu_t,\ \nu_t \sim^{iid} N(0, \sigma_\nu^2) \\ & \quad -1 < \phi < 1 \end{aligned}$$

Εύκολα μπορεί να δειχθεί πως

$$\begin{aligned} \epsilon_t &\sim N(0, \frac{\sigma_\nu^2}{1-\phi^2}) \\ cov(\epsilon_t, \epsilon_{t-k}) &= \phi^k \frac{\sigma_\nu^2}{1-\phi^2} \end{aligned}$$

$$\boldsymbol{\Sigma} = \sigma_\nu^2 \begin{bmatrix} \frac{1}{1-\phi^2} & \frac{\phi}{1-\phi^2} & \frac{\phi^2}{1-\phi^2} & \cdots & \frac{\phi^{n-1}}{1-\phi^2} \\ \frac{\phi}{1-\phi^2} & \frac{1}{1-\phi^2} & \frac{\phi}{1-\phi^2} & \cdots & \frac{\phi^{n-2}}{1-\phi^2} \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ \frac{\phi^{n-1}}{1-\phi^2} & \frac{\phi^{n-2}}{1-\phi^2} & \frac{\phi^{n-3}}{1-\phi^2} & \cdots & \frac{1}{1-\phi^2} \end{bmatrix}$$

Παράδειγμα 3: Απλό γραμμικό μοντέλο τυχαίων παραγόντων (one way random effects model).

$$\begin{aligned} Y_{ij} &= \mu_i + \nu_{ij},\ \nu_{ij} \sim^{iid} N(0, \sigma^2),\ j = 1, \ldots, n_i \\ \mu_i &\sim^{iid} N(\mu, \sigma_\pi^2),\ i = 1, \ldots, k \\ \nu_{ij} &\perp \nu_{i'j},\ i \neq i' \\ \nu_{ij} &\perp \mu_i,\ \forall (i,j) \end{aligned} \tag{1.7}$$

Ισοδύναμα, ορίζοντας $\pi_i = \mu_i - \mu$, γράφουμε το μοντέλο ως

$$\begin{array}{rcll} Y_{ij} & = & \mu + \pi_i + \nu_{ij},\ \nu_{ij} \sim^{iid} N(0,\sigma^2),\ j = 1,\ldots,n_i & \\ \pi_i & \sim^{iid} & N(0,\sigma_\pi^2),\ i = 1,\ldots,k & \\ \nu_{ij} & \perp & \nu_{i^{'}j},\ i \neq i^{'} & \\ \nu_{ij} & \perp & \pi_i,\ \forall (i,j) & (1.8) \end{array}$$

Το μοντέλο μπορεί να γραφεί ως Γενικό, Γραμμικό μοντέλο, δηλαδή στην μορφή (1.1). Ορίζω $\epsilon_{ij} = \pi_i + \nu_{ij}$. Έχουμε

$$\begin{array}{rcll} var(\epsilon_{ij}) & = & \sigma_\pi^2 + \sigma^2 & \\ cov(\epsilon_{ij}, \epsilon_{ij^{'}}) & = & \sigma_\pi^2,\ j \neq j^{'} & \\ cov(\epsilon_{ij}, \epsilon_{i^{'}j^{'}}) & = & 0,\ i \neq i^{'} & (1.9) \end{array}$$

Εδώ $\boldsymbol{Y} = (Y_{11},\ldots,Y_{1n_1},Y_{21},\ldots,Y_{2n_2},\ldots,\ldots,Y_{k1},\ldots,Y_{kn_k})^T$, $\boldsymbol{\beta} = \mu$, $\boldsymbol{X} = \mathbf{1}_N$, όπου $\mathbf{1}_N$ είναι ένα διάνυσμα με $N = n_1 + \ldots + n_k$ μονάδες, $\boldsymbol{\epsilon} = (\epsilon_{11},\ldots,\epsilon_{1n_1},\epsilon_{21},\ldots,\epsilon_{2n_2},\ldots,\ldots,\epsilon_{k1},\ldots,\epsilon_{kn_k})^T$ και $var(\boldsymbol{\epsilon}) = \boldsymbol{\Sigma}^*$ δίνεται από

$$\boldsymbol{\Sigma}^* = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2 & \ldots & \mathbf{0} \\ \vdots & \vdots & \ldots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \boldsymbol{\Sigma}_k \end{bmatrix} = \sigma^2 \begin{bmatrix} \boldsymbol{V}_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{V}_2 & \ldots & \mathbf{0} \\ \vdots & \vdots & \ldots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \boldsymbol{V}_k \end{bmatrix} = \sigma^2 \mathscr{V}$$

όπου $\boldsymbol{\Sigma}_i = var(\boldsymbol{\epsilon}_i)$ είναι ένας $n_i \times n_i$ πίνακας συνδιακυμάνσεων των παρατηρήσεων από το i-οστό εργαστήριο και δίνεται από

$$\boldsymbol{\Sigma}_i = \begin{bmatrix} \sigma_\pi^2 + \sigma^2 & \sigma_\pi^2 & \ldots & \sigma_\pi^2 \\ \sigma_\pi^2 & \sigma_\pi^2 + \sigma^2 & \ldots & \sigma_\pi^2 \\ \vdots & \vdots & \ldots & \vdots \\ \sigma_\pi^2 & \sigma_\pi^2 & \ldots & \sigma_\pi^2 + \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1+\phi & \phi & \ldots & \phi \\ \phi & 1+\phi & \ldots & \phi \\ \vdots & \vdots & \ldots & \vdots \\ \phi & \phi & \ldots & 1+\phi \end{bmatrix}$$

όπου $\phi = \frac{\sigma_\pi^2}{\sigma^2} > 0$

Παράδειγμα 4: Απλό γραμμικό μοντέλο με ένα σταθερό παράγοντα κι εναν τυχαίο παράγοντα (two way mixed effects model).

$$
\begin{aligned}
Y_{ij} &= \mu + \alpha_i + \beta_j + \nu_{ij},\ \nu_{ij} \sim^{iid} N(0,\sigma^2),\ j=1,\ldots,n \\
\alpha_i &\sim^{iid} N(0,\sigma_\alpha^2),\ i=1,\ldots,m \\
\nu_{ij} &\perp \nu_{i'j},\ i \neq i' \\
\nu_{ij} &\perp \alpha_i,\ \forall(i,j)
\end{aligned} \tag{1.10}
$$

Είναι ισοδύναμο με το μοντέλο στην άσκηση Ε6.6 στο βιβλίο. Μπορώ να γράψω το μοντέλο ώς εξής:

$$
\begin{aligned}
Y_{ij} &= (\mu + \beta_j) + \alpha_i + \nu_{ij},\ \nu_{ij} \sim^{iid} N(0,\sigma^2),\ j=1,\ldots,n \\
\alpha_i &\sim^{iid} N(0,\sigma_\alpha^2),\ i=1,\ldots,m \\
\nu_{ij} &\perp \nu_{i'j},\ i \neq i' \\
\nu_{ij} &\perp \alpha_i,\ \forall(i,j),
\end{aligned} \tag{1.11}
$$

όπου ο όρος $(\mu+\beta_j)$ είναι σταθερός. Δηλαδή έχω :

Το μοντέλο το γράφω σαν κανονικό Γενικό γραμμικό μοντέλο $(\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},\ \boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma}))$,

$$
\boldsymbol{Y} = \begin{bmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1n} \\ - \\ Y_{21} \\ Y_{22} \\ \vdots \\ Y_{2n} \\ - \\ \vdots \\ \vdots \\ - \\ Y_{m1} \\ Y_{m2} \\ \vdots \\ Y_{mn} \end{bmatrix} = \begin{bmatrix} \mu+\beta_1 \\ \mu+\beta_2 \\ \vdots \\ \mu+\beta_n \\ - \\ \mu+\beta_1 \\ \mu+\beta_2 \\ \vdots \\ \mu+\beta_n \\ - \\ \vdots \\ \vdots \\ - \\ \mu+\beta_1 \\ \mu+\beta_2 \\ \vdots \\ \mu+\beta_n \end{bmatrix} + \begin{bmatrix} \alpha_1 \\ \alpha_1 \\ \vdots \\ \alpha_1 \\ - \\ \alpha_2 \\ \alpha_2 \\ \vdots \\ \alpha_2 \\ - \\ \vdots \\ \vdots \\ - \\ \alpha_m \\ \alpha_m \\ \vdots \\ \alpha_m \end{bmatrix} + \begin{bmatrix} \nu_{11} \\ \nu_{12} \\ \vdots \\ \nu_{1n} \\ - \\ \nu_{21} \\ \nu_{22} \\ \vdots \\ \nu_{2n} \\ - \\ \vdots \\ \vdots \\ - \\ \nu_{m1} \\ \nu_{m2} \\ \vdots \\ \nu_{mn} \end{bmatrix}
$$

Εδώ $\boldsymbol{\beta} = (\mu, \beta_1, \ldots, \beta_n)^T$. Άρα ο $mn \times (n+1)$ Πίνακας σχεδιασμού είναι

$$\boldsymbol{X} = \begin{bmatrix} 1 & 1 & 0 & \ldots & 0 \\ 1 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \ldots & 1 \\ - & - & - & - & - \\ 1 & 1 & 0 & \ldots & 0 \\ 1 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \ldots & 1 \\ - & - & - & - & - \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ - & - & - & - & - \\ 1 & 1 & 0 & \ldots & 0 \\ 1 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \ldots & 1 \end{bmatrix}$$

και ο πίνακας διασποράς είναι ο $mn \times mn$ πίνακας

$$\boldsymbol{\Sigma}^* = \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ldots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \boldsymbol{\Sigma} \end{bmatrix} = \sigma^2 \begin{bmatrix} \boldsymbol{V} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{V} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ldots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \boldsymbol{V} \end{bmatrix} = \sigma^2 \mathscr{V}$$

όπου $\boldsymbol{\Sigma}$ είναι ένας $n \times n$ πίνακας συνδιακυμάνσεων των παρατηρήσεων από το i-οστό εργαστήριο και δίνεται από

$$\boldsymbol{\Sigma}_i = \begin{bmatrix} \sigma_\alpha^2 + \sigma^2 & \sigma_\alpha^2 & \ldots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma^2 & \ldots & \sigma_\alpha^2 \\ \vdots & \vdots & \ldots & \vdots \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \ldots & \sigma_\alpha^2 + \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1+\phi & \phi & \ldots & \phi \\ \phi & 1+\phi & \ldots & \phi \\ \vdots & \vdots & \ldots & \vdots \\ \phi & \phi & \ldots & 1+\phi \end{bmatrix}$$

όπου $\phi = \frac{\sigma_\alpha^2}{\sigma^2} > 0$

Προσοχή εδώ $rank(\boldsymbol{X}) = n < n+1$. Κανένα πρόβλημα! (;). Παραμετροποιώ το μοντέλο ορίζοντας $\mu_j = \mu + \beta_j$, κι έχω ως $\boldsymbol{\beta} = (\mu_1, \ldots, \mu_n)^T$. Δηλαδή δουλεύω με το ισοδύναμο μοντέλο:

$$
\begin{array}{ccl}
Y_{ij} & = & \mu_j + \alpha_i + \nu_{ij},\ \nu_{ij} \sim^{iid} N(0, \sigma^2),\ j = 1, \ldots, n \\
\alpha_i & \sim^{iid} & N(0, \sigma_\alpha^2),\ i = 1, \ldots, m \\
\nu_{ij} & \perp & \nu_{i'j},\ i \neq i^{'} \\
\nu_{ij} & \perp & \alpha_i,\ \forall (i, j)
\end{array}
\tag{1.12}
$$

και

$$
\boldsymbol{X} = \begin{bmatrix}
1 & 0 & \ldots & 0 & \\
0 & 1 & \ldots & 0 & \\
\vdots & \vdots & \vdots & \vdots & \\
0 & 0 & \ldots & 1 & \\
- & - & - & - & \\
1 & 0 & \ldots & 0 & \\
0 & 1 & \ldots & 0 & \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \ldots & 1 & \\
- & - & - & - & \\
\vdots & \vdots & \vdots & \vdots & \\
\vdots & \vdots & \vdots & \vdots & \\
- & - & - & - & \\
1 & 0 & \ldots & 0 & \\
0 & 1 & \ldots & 0 & \\
\vdots & \vdots & \vdots & \vdots & \\
0 & 0 & \ldots & 1 &
\end{bmatrix}
$$

Παράδειγμα 5: Κανονικό γραμμικό μοντέλο με τυχαίους συντελεστές (random coefficients model).

Θεωρήστε επαναλαμβανόμενες μετρήσεις σε άτομα που συμμετέχουν σε μια κλινική δοκιμή. Υποθέτουμε πως μετά την έναρξη θεραπείας η απόκριση Y (ενδεχομένως) ακολουθεί μια γραμμική τάση (αυξάνεται ή μειώνεται) μέσα σε ένα δεδομένο χρονικό διάστημα. Η τάση αυτή ΄ποικίλει΄από άτομο σε άτομο.

$$\begin{aligned} Y_{ij} &= b_{0i} + b_{1i}t_{ij} + \nu_{ij},\ i=1,\ldots,k,\ j=1,\ldots,n_i \\ b_{0i} &\sim^{iid} N(\beta_0, \sigma_0^2) \\ b_{1i} &\sim^{iid} N(\beta_1, \sigma_1^2) \\ cov(b_{0i}, b_{1i}) &= \sigma_{01} \\ \nu_{ij} &\sim^{iid} N(0, \sigma^2) \\ \nu_{ij} &\perp (b_{0i}, b_{1i})^T \end{aligned} \tag{1.13}$$

Είναι λογικό πως υπάρχει ανεξαρτησία των τυχαίων συντελεστών, σφαλμάτων (κι άρα και των παρατηρήσεων) μεταξύ των ατόμων. Ισοδύναμα, παίρνοντας $b_{0i} = \beta_0 + (b_{0i} - \beta_0) = \beta_0 + \pi_{0i}$ και $b_{1i} = \beta_1 + (b_{1i} - \beta_0) = \beta_1 + \pi_{1i}$ γράφουμε:

$$\begin{aligned} Y_{ij} &= \beta_0 + \beta_1 t_{ij} + \pi_{0i} + \pi_{1i}t_{ij} + \nu_{ij}\ i=1,\ldots,k,\ j=1,\ldots,n_i \\ \pi_{0i} &\sim^{iid} N(0, \sigma_0^2) \\ \pi_{1i} &\sim^{iid} N(0, \sigma_1^2) \\ cov(\pi_{0i}, \pi_{1i}) &= \sigma_{01} \\ \nu_{ij} &\sim^{iid} N(0, \sigma^2) \\ \nu_{ij} &\perp (\pi_{0i}, \pi_{1i})^T \end{aligned} \tag{1.14}$$

Το μοντέλο γράφεται εύκολα ως μικτό μοντέλο σε μορφή πινάκων ως εξής. Γραφουμε το διάνυσμα αποκρίσεων του ι-οστού ατόμου, ως $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i})^T$. Έχουμε

$$\begin{aligned} \boldsymbol{Y}_i &= \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{\pi}_i + \boldsymbol{\nu}_i,\ i=1,\ldots,k, \\ \boldsymbol{\pi}_i &\sim^{iid} \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{B}),\ \boldsymbol{\nu}_i \sim^{iid} N(0, \sigma^2 \boldsymbol{I}_{n_i}) \\ \boldsymbol{\nu}_i &\perp \boldsymbol{\pi}_{i'}, \forall (i, i'), \end{aligned} \tag{1.15}$$

όπου

$$\boldsymbol{X}_i = \begin{bmatrix} 1 & t_{i1} \\ 1 & t_{i2} \\ \vdots & \vdots \\ 1 & t_{in_i} \end{bmatrix}, \boldsymbol{B} = \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \boldsymbol{\pi}_i = \begin{bmatrix} \pi_{i0} \\ \pi_{i1} \end{bmatrix}, \boldsymbol{\nu}_i = \begin{bmatrix} \nu_{i1} \\ \nu_{i2} \\ \vdots \\ \nu_{in_i} \end{bmatrix}.$$

Εδώ έχουμε $\boldsymbol{X}_i = \boldsymbol{Z}_i$. Ισοδύναμα γράφουμε

$$\begin{aligned} \boldsymbol{Y} &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\pi} + \boldsymbol{\nu} && (1.16) \\ \boldsymbol{\pi} &\sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{I}_k \otimes \boldsymbol{B}),\ \boldsymbol{\nu} \sim \boldsymbol{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_N) && (1.17) \end{aligned}$$

όπου

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \\ \vdots \\ \boldsymbol{Y}_k \end{bmatrix}, \boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_k \end{bmatrix}, \boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Z}_1 & \boldsymbol{0} & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Z}_2 & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \dots & \boldsymbol{Z}_k \end{bmatrix}, \boldsymbol{\pi} = \begin{bmatrix} \boldsymbol{\pi}_1 \\ \boldsymbol{\pi}_2 \\ \vdots \\ \boldsymbol{\pi}_k \end{bmatrix}, \boldsymbol{\nu} = \begin{bmatrix} \boldsymbol{\nu}_1 \\ \boldsymbol{\nu}_2 \\ \vdots \\ \boldsymbol{\nu}_k \end{bmatrix}.$$

, από όπου προκύπτει πως

$$\boldsymbol{Y} \sim \boldsymbol{N}(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{Z}(\boldsymbol{I}_k \otimes \boldsymbol{B})\boldsymbol{Z}^T + \sigma^2 \boldsymbol{I}_N) \tag{1.18}$$

Στήν ειδική περίπτωση που $\boldsymbol{Z}_1 = \ldots = \boldsymbol{Z}_k = \mathscr{Z}$ έχουμε $\boldsymbol{Z} = \boldsymbol{I}_k \otimes \mathscr{Z}$.

Παράδειγμα 6: Κανονικό γραμμικό μοντέλο με τυχαίους συντελεστές και 2 ομάδες.

Το μοντέλο επεκτείνεται εύκολα να συμπεριλάβει την επίδραση μιας θεραπείας (φαρμάκου) σε σύγκριση με ένα (placebo). Έστω $d_{ij} = 1$ αν το άτομο παίρνει το φάρμακο (θεραπεία) και $d_{ij} = 0$ αν το άτομο παίρνει το placebo. Τότε μία πιθανή γενίκευση του μοντέλου είναι:

$$\begin{aligned} Y_{ij} &= \beta_0 + \beta_1 t_{ij} + \beta_2 d_{ij} + \pi_{0i} + \pi_{1i} t_{ij} + \nu_{ij}\ i = 1, \ldots, k,\ j = 1, \ldots, n_i \\ \pi_{0i} &\sim^{iid} N(0, \sigma_0^2) \\ \pi_{1i} &\sim^{iid} N(0, \sigma_1^2) \\ cov(\pi_{0i}, \pi_{1i}) &= \sigma_{01} \\ \nu_{ij} &\sim^{iid} N(0, \sigma^2) \\ \nu_{ij} &\perp (\pi_{0i}, \pi_{1i})^T \end{aligned} \tag{1.19}$$

Το παραπάνω μοντέλο λέει πως το φάρμακο ασκεί επίρροια μόνο στο (intercept) αλλά όχι στην κλίση, δηλαδή οι μέσες πληθυσμιακές ευθείες είναι παράλληλες!!! Άρα οι κλισεις είναι ίδιες. Για να εξετάσω αν υπάρχεια επίρροια του φαρμάκου στην κλίση εξετάζω το μοντέλο

$$\begin{array}{rcl} Y_{ij} & = & \beta_0 + \beta_1 t_{ij} + \beta_2 d_{ij} + \beta_3 d_{ij} t_{ij} + \pi_{0i} + \pi_{1i} t_{ij} + \nu_{ij} \; i = 1, \ldots, k, \; j = 1, \ldots, n_i \\ \pi_{0i} & \sim^{iid} & N(0, \sigma_0^2) \\ \pi_{1i} & \sim^{iid} & N(0, \sigma_1^2) \\ cov(\pi_{0i}, \pi_{1i}) & = & \sigma_{01} \\ \nu_{ij} & \sim^{iid} & N(0, \sigma^2) \\ \nu_{ij} & \perp & (\pi_{0i}, \pi_{1i})^T \end{array} \tag{1.20}$$

Εδώ βλέπουμε πως για την ομάδα placebo $E(Y_t) = \beta_0 + \beta_1 t$ ενώ για την ομάδα φαρμάκου $E(Y_t) = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)t$.

Στο μοντέλο αυτό ισχύει ότι ισχύει και στο παράδειγμα 5, με τις εξής επισημάνσεις (διαφορές):

$$\boldsymbol{X}_i = \begin{bmatrix} 1 & t_{i1} & d_{i1} & d_{i1}t_{i1} \\ 1 & t_{i2} & d_{i2} & d_{i2}t_{i2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_{in_i} & d_{in_i} & d_{in_i}t_{in_i} \end{bmatrix}, \boldsymbol{Z}_i = \begin{bmatrix} 1 & t_{i1} \\ 1 & t_{i2} \\ \vdots & \vdots \\ 1 & t_{in_i} \end{bmatrix} \neq \boldsymbol{X}_i, \; \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$$

1.2 Εκτίμηση με την μέθοδο της μέγιστης Πιθανοφάνειας (Maximum Likelihood Estimation (MLE))

Με την υπόθεση κανονικότητας έχουμε

$$\boldsymbol{Y}|\boldsymbol{X} \sim N(\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}(\sigma^2, \boldsymbol{\phi}) = \sigma^2 \boldsymbol{V}(\boldsymbol{\phi})). \tag{1.21}$$

Η loglikelihood είναι:

$$l(\boldsymbol{Y}, \boldsymbol{X}; \boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi}) = -\frac{n}{2} ln 2\pi - \frac{n}{2} ln \sigma^2 - \frac{1}{2} ln |\boldsymbol{V}(\boldsymbol{\phi})| - \frac{1}{2\sigma^2} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T \boldsymbol{V}^{-1}(\boldsymbol{\phi}) (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) \tag{1.22}$$

Δοθέντος $\boldsymbol{\phi}$ (και σ^2), η loglikelihood μεγιστοποιείται ως προς $\boldsymbol{\beta}$ στην τιμή $\hat{\boldsymbol{\beta}}_{GLS}(\boldsymbol{\phi}) = (\mathbf{X}^T \boldsymbol{V}^{-1}(\boldsymbol{\phi}) \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{V}^{-1}(\boldsymbol{\phi}) \mathbf{Y}$. (γιατί;;)

Άρα αρκεί να μεγιστοποιήσουμε 'συγκεντρωμένη' πιθανοφάνεια

$$l_1(\sigma^2,\boldsymbol{\phi}) = -\frac{n}{2}ln2\pi - \frac{n}{2}ln\sigma^2 - \frac{1}{2}ln|\boldsymbol{V}(\boldsymbol{\phi})| - \frac{1}{2\sigma^2}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_{GLS}(\boldsymbol{\phi}))^T\boldsymbol{V}^{-1}(\boldsymbol{\phi})(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_{GLS}(\boldsymbol{\phi})) \quad (1.23)$$

Δοθέντος $\boldsymbol{\phi}$ η $l_1(\sigma^2,\boldsymbol{\phi})$ μεγιστοποιείται ως προς σ^2 στην τιμή

$$\hat{\sigma}^2(\boldsymbol{\phi}) = \frac{(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_{GLS}(\boldsymbol{\phi}))^T\boldsymbol{V}^{-1}(\boldsymbol{\phi})(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_{GLS}(\boldsymbol{\phi}))}{n}, \quad (1.24)$$

(γιατί;;).

Καθότι

$$\begin{aligned}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_{GLS})^T\boldsymbol{V}^{-1}(\boldsymbol{\phi})(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_{GLS}) =\\ (\boldsymbol{Y}-\boldsymbol{X}(\mathbf{X}^T\boldsymbol{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{V}^{-1}\mathbf{Y})^T\boldsymbol{V}^{-1}(\boldsymbol{Y}-\boldsymbol{X}(\mathbf{X}^T\boldsymbol{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{V}^{-1}\mathbf{Y})\\ = \boldsymbol{Y}^T(\boldsymbol{I}-(\mathbf{X}^T\boldsymbol{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{V}^{-1})^T\boldsymbol{V}^{-1}(\boldsymbol{I}-\boldsymbol{X}(\mathbf{X}^T\boldsymbol{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{V}^{-1})\mathbf{Y}\\ = \boldsymbol{Y}^T(\boldsymbol{V}^{-1}-\boldsymbol{V}^{-1}\boldsymbol{X}(\mathbf{X}^T\boldsymbol{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{V}^{-1})\boldsymbol{Y}\end{aligned} \quad (1.25)$$

έχουμε

$$\hat{\sigma}^2(\boldsymbol{\phi}) = \frac{1}{n}\boldsymbol{Y}^T(\boldsymbol{V}^{-1}(\boldsymbol{\phi})-\boldsymbol{V}^{-1}(\boldsymbol{\phi})\boldsymbol{X}(\mathbf{X}^T\boldsymbol{V}^{-1}(\boldsymbol{\phi})\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{V}^{-1}(\boldsymbol{\phi}))\boldsymbol{Y} \quad (1.26)$$

Άρα αρκεί να μεγιστοποιήσουμε 'συγκεντρωμένη'-'συγκεντρωμένη' πιθανοφάνεια.

$$\begin{aligned} l_1(\boldsymbol{\phi}) &= -\frac{n}{2}ln2\pi - \frac{n}{2}ln\hat{\sigma}^2(\boldsymbol{\phi}) - \frac{1}{2}ln|\boldsymbol{V}(\boldsymbol{\phi})| - \frac{1}{2\hat{\sigma}^2(\boldsymbol{\phi})}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_{GLS}(\boldsymbol{\phi}))^T\boldsymbol{V}^{-1}(\boldsymbol{\phi})(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_{GLS}(\boldsymbol{\phi}))\\ &= -\frac{n}{2}ln2\pi - \frac{n}{2}ln\hat{\sigma}^2(\boldsymbol{\phi}) - \frac{1}{2}ln|\boldsymbol{V}(\boldsymbol{\phi})| - \frac{n}{2} \end{aligned} \quad (1.27)$$

Δηλαδή <u>ελαχιστοποιούμε</u> την ποσότητα

$$nln\hat{\sigma}^2(\boldsymbol{\phi}) + ln|\boldsymbol{V}(\boldsymbol{\phi})| \quad (1.28)$$

ως προς $\boldsymbol{\phi}$ και παίρνουμε $\hat{\boldsymbol{\phi}}_{MLE}$. κι έχουμε

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{MLE} &= \hat{\boldsymbol{\beta}}_{GLS}(\hat{\boldsymbol{\phi}}_{MLE})\\ \hat{\sigma}^2_{MLE} &= \hat{\sigma}^2(\hat{\boldsymbol{\phi}}_{MLE}) = \frac{(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_{GLS}(\hat{\boldsymbol{\phi}}_{MLE}))^T\boldsymbol{V}^{-1}(\hat{\boldsymbol{\phi}}_{MLE})(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_{GLS}(\hat{\boldsymbol{\phi}}_{\boldsymbol{MLE}}))}{n}\end{aligned} \quad (1.29)$$

Εναλλάκτικά μπορώ να δουλέψω χωρίς να ξεχωρίζω το σ^2, δηλαδή με το $\mathbf{\Sigma}(\boldsymbol{\psi})$. Εδώ $\boldsymbol{\psi}$ είναι όλες οι παράμετροι στον πίνακα διασποράς συνδιασποράς.

Η loglikelihood είναι:

$$l(\boldsymbol{Y}, \boldsymbol{X}; \boldsymbol{\beta}, \boldsymbol{\psi}) = -\frac{n}{2} ln 2\pi - \frac{1}{2} ln|\mathbf{\Sigma}(\boldsymbol{\psi})| - \frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T \mathbf{\Sigma}^{-1}(\boldsymbol{\psi})(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) \tag{1.30}$$

και η συγκεντρωμένη πιθανοφάνεια ισούται με

$$l_1(\boldsymbol{\psi}) = -\frac{n}{2} ln 2\pi - \frac{1}{2} ln|\mathbf{\Sigma}(\boldsymbol{\psi})| - \frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{GLS}(\boldsymbol{\psi}))^T \mathbf{\Sigma}^{-1}(\boldsymbol{\psi})(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{GLS}(\boldsymbol{\psi})) \tag{1.31}$$

όπου $\hat{\boldsymbol{\beta}}_{GLS}(\boldsymbol{\psi}) = (\mathbf{X}^T \mathbf{\Sigma}^{-1}(\boldsymbol{\psi})\mathbf{X})^{-1}\mathbf{X}^T \mathbf{\Sigma}^{-1}(\boldsymbol{\psi})\mathbf{Y}$,

και ελαχιστοποιώ

$$l_1(\boldsymbol{\psi}) = ln|\mathbf{\Sigma}(\boldsymbol{\psi})| + (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{GLS}(\boldsymbol{\psi}))^T \mathbf{\Sigma}^{-1}(\boldsymbol{\psi})(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{GLS}(\boldsymbol{\psi})) \tag{1.32}$$