Regression to the mean in the normal case

מסגרת:

$$X_i = X + \varepsilon, \quad i=1,2$$
$$X \sim N(\mu, \sigma^2)$$
$$\varepsilon_i \sim N(0, \tau^2)$$
$$X, \varepsilon_1, \varepsilon_2 \text{ ב״ת}$$

מהנ״ל נובע כי

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu \\ \mu \end{bmatrix}, \eta^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right)$$

כאשר

$$\eta^2 = \sigma^2 + \tau^2$$
$$\rho = \sigma^2 / (\sigma^2 + \tau^2)$$

לפי התכונות של התפלגות הדו-נורמלית, יש לנו

$$E[X_2 \mid X_1 = x] = \mu + \rho(x-\mu) = (1-\rho)\mu + \rho x$$

לכן:

$$E[X_2 \mid X_1 \geq c] = \frac{1}{Pr(X_1 \geq c)} \int_c^\infty E[X_2 \mid X_1 = x] f_{X_1}(x)\, dx$$

צפיפות של X_1

$$= \frac{1}{Pr(X_1 \geq c)} \int_c^\infty [(1-\rho)\mu + \rho x] f_{X_1}(x)\, dx$$

$$= \left(\frac{1}{Pr(X_1 \geq c)} \int_c^\infty f_{X_1}(x)\, dx \right) ((1-\rho)\mu)$$
$$+ \rho \frac{1}{Pr(X_1 \geq c)} \int_c^\infty x f_{X_1}(x)\, dx$$

$$= (1-\rho)\mu + \rho E[X_1 \mid X_1 \geq c]$$

בס״ד

לכן:

$$E[X_2 - X_1 \mid X_1 \geq c]$$

$$= (1-\rho)\mu + \rho E[X_1 \mid X_1 \geq c] - E[X_1 \mid X_1 \geq c]$$

$$= -(1-\rho)\big(E[X_1 \mid X_1 \geq c] - \mu\big)$$

כאשר:

$$E[X_1 \mid X_1 \geq c] = \frac{1}{Pr(X_1 \geq c)} \int_c^\infty x f_{X_1}(x)\, dx$$

$$= \left(1 - \Phi\left(\frac{c-\mu}{\sigma}\right)\right)^{-1} \int_c^\infty x \left[\frac{1}{\eta} \varphi\left(\frac{x-\mu}{\eta}\right)\right] dx$$

$$\left[\begin{array}{ll} v = \frac{x-\mu}{\eta} & \text{ועכשיו נציב:} \\ x = \mu + \eta v & \\ dx = \eta\, dv & \end{array}\right]$$

$$= \left(1 - \Phi\left(\frac{c-\mu}{\eta}\right)\right)^{-1} \int_{c'}^\infty [\mu + \eta v\, \varphi(v)]\, dv$$

$$\left(c' = \frac{c-\mu}{\eta} \quad \text{כאשר} \right)$$

$$= \mu + \eta \left(1 - \Phi\left(\frac{c-\mu}{\eta}\right)\right)^{-1} \int_{c'}^\infty v\, \varphi(v)\, dv$$

$$= \mu + \eta \left(1 - \Phi\left(\frac{c-\mu}{\eta}\right)\right)^{-1} \int_{c'}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}v^2} v\, dv$$

$$= \mu + \eta \left(1 - \Phi\left(\frac{c-\mu}{\eta}\right)\right)^{-1} \left[-\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}v^2} \Big|_{c'}^{\infty} \right]$$

$$= \mu + \eta \left(1 - \Phi\left(\frac{c-\mu}{\eta}\right)\right)^{-1} \varphi\left(\frac{c-\mu}{\eta}\right)$$

כ ס"ן

ואם כן, יוצא כי

$$E[X_2 - X_1 \mid X_1 \geq c]$$
$$= -(1-\rho)\eta \frac{\varphi\left(\frac{c-\mu}{\eta}\right)}{1-\Phi\left(\frac{c-\mu}{\eta}\right)}$$