Задача (о последствиях гетероскедастичности для МНК-оценок). Пусть задана модель линейной регрессии $Y_t = \beta \cdot t + \varepsilon_t$, $t = 1, 2, 3$, где $\mathbb{E}[\varepsilon_t] = 0$, $\mathrm{D}(\varepsilon_t) = t\sigma^2$ и $\mathrm{cov}(\varepsilon_s, \varepsilon_t) = 0$ при $s \neq t$. Рассматриваются две оценки:

$$\hat{\beta}_{\text{МНК}} = \frac{\sum_{t=1}^{3} tY_t}{\sum_{t=1}^{3} t^2} \qquad \text{и} \qquad \tilde{\beta} = \frac{\sum_{t=1}^{3} Y_t}{\sum_{t=1}^{3} t}.$$

(a) Найдите $\mathbb{E}[\hat{\beta}_{\text{МНК}}]$ и $\mathbb{E}[\tilde{\beta}]$.

(b) Найдите $\mathrm{D}(\hat{\beta}_{\text{МНК}})$ и $\mathrm{D}(\tilde{\beta})$.

(c) Сравните $\mathrm{D}(\hat{\beta}_{\text{МНК}})$ и $\mathrm{D}(\tilde{\beta})$. Какой вывод можно сделать о последствии гетероскедастичности на МНК-оценки?

Решение. (a) Найдем математические ожидания оценок $\hat{\beta}_{\text{МНК}}$ и $\tilde{\beta}$:

$$\mathbb{E}[\hat{\beta}_{\text{МНК}}] = \mathbb{E}\left[\frac{\sum_{t=1}^{3} tY_t}{\sum_{t=1}^{3} t^2}\right] = \frac{\sum_{t=1}^{3} t\mathbb{E}[Y_t]}{\sum_{t=1}^{3} t^2} = \frac{\sum_{t=1}^{3} t \cdot \beta t}{\sum_{t=1}^{3} t^2} = \beta,$$

$$\mathbb{E}[\tilde{\beta}] = \mathbb{E}\left[\frac{\sum_{t=1}^{3} Y_t}{\sum_{t=1}^{3} t}\right] = \frac{\sum_{t=1}^{3} \mathbb{E}[Y_t]}{\sum_{t=1}^{3} t} = \frac{\sum_{t=1}^{3} \beta t}{\sum_{t=1}^{3} t} = \beta.$$

(b) Найдем дисперсии оценок $\hat{\beta}_{\text{МНК}}$ и $\tilde{\beta}$:

$$\mathrm{D}(\hat{\beta}_{\text{МНК}}) = \mathrm{D}\left(\frac{\sum_{t=1}^{3} tY_t}{\sum_{t=1}^{3} t^2}\right) = \frac{\mathrm{D}\left(\sum_{t=1}^{3} tY_t\right)}{\left(\sum_{t=1}^{3} t^2\right)^2} = \frac{\sum_{t=1}^{3} \mathrm{D}(tY_t)}{\left(\sum_{t=1}^{3} t^2\right)^2} = \frac{\sum_{t=1}^{3} t^2\,\mathrm{D}(Y_t)}{\left(\sum_{t=1}^{3} t^2\right)^2} =$$

$$= \frac{\sum_{t=1}^{3} t^2\,\mathrm{D}(\varepsilon_t)}{\left(\sum_{t=1}^{3} t^2\right)^2} = \frac{\sum_{t=1}^{3} t^2 t\sigma^2}{\left(\sum_{t=1}^{3} t^2\right)^2} = \sigma^2 \frac{\sum_{t=1}^{3} t^3}{\left(\sum_{t=1}^{3} t^2\right)^2} = \sigma^2 \frac{1^3 + 2^3 + 3^3}{(1^2 + 2^2 + 3^2)^2} \approx 0.18\sigma^2,$$

$$\mathrm{D}(\tilde{\beta}) = \mathrm{D}\left(\frac{\sum_{t=1}^{3} Y_t}{\sum_{t=1}^{3} t}\right) = \frac{\mathrm{D}\left(\sum_{t=1}^{3} Y_t\right)}{\left(\sum_{t=1}^{3} t\right)^2} = \frac{\sum_{t=1}^{3} \mathrm{D}(Y_t)}{\left(\sum_{t=1}^{3} t\right)^2} = \frac{\sum_{t=1}^{3} \mathrm{D}(\varepsilon_t)}{\left(\sum_{t=1}^{3} t\right)^2} =$$

$$= \frac{\sum_{t=1}^{3} t\sigma^2}{\left(\sum_{t=1}^{3} t\right)^2} = \sigma^2 \frac{\sum_{t=1}^{3} t}{\left(\sum_{t=1}^{3} t\right)^2} = \sigma^2 \frac{1+2+3}{(1+2+3)^2} \approx 0.17\sigma^2.$$

(c) Обе оценки $\hat{\beta}_{\text{МНК}}$ и $\tilde{\beta}$ являются несмещенными и линейными по вектору $Y = [Y_1, Y_2, Y_3]^T$ оценками неизвестного параметра β. При этом $\mathrm{D}(\hat{\beta}_{\text{МНК}}) > \mathrm{D}(\tilde{\beta})$. Стало быть, оценка $\tilde{\beta}$ является более эффективной по сравнению с МНК-оценкой $\tilde{\beta}$. Таким образом, в условиях гетероскедастичности МНК-оценки перестают быть наиболее эффективными среди всех несмещенных линейных по вектору Y оценками. ☐