PROBLEM SET 6

Problem 1 (Pooled OLS, Random Effects and Fixed Effects)

A linear panel data model with $T = 2$ time periods, one regressor x_{it}, and scalar individual specific effect α_i is given by

$$\underbrace{\begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix}}_{=y_i} = \underbrace{\begin{pmatrix} x_{i1} \\ x_{i2} \end{pmatrix}}_{=x_i} \beta + \underbrace{\begin{pmatrix} u_{i1} \\ u_{i2} \end{pmatrix}}_{=u_i}, \qquad \begin{pmatrix} u_{i1} \\ u_{i2} \end{pmatrix} = \underbrace{\begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \end{pmatrix}}_{=\varepsilon_i} + \alpha_i, \qquad \alpha_i = (x_{i1} + x_{i2})\gamma + \tilde{\alpha}_i.$$

We assume that $\tilde{\alpha}_i$, x_{i1}, x_{i2}, ε_{i1} and ε_{i2} are mutually independent random variables, are independent across observations $i = 1, \ldots, n$, and are distributed as $\tilde{\alpha}_i \sim \mathcal{N}(0, \sigma_\alpha^2)$, $x_{it} \sim \mathcal{N}(0, \sigma_x^2)$, and $\varepsilon_{it} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$, $t = 1, 2$. Unknown parameters are $\beta \in \mathbb{R}$, $\gamma \in \mathbb{R}$, $\sigma_\alpha > 0$, $\sigma_x > 0$ and $\sigma_\varepsilon > 0$. Observed variables are y_i and x_i. We consider three estimators for β

$$\hat{\beta}_{\text{OLS}} = \frac{\sum_{i=1}^n x_i' y_i}{\sum_{i=1}^n x_i' x_i}, \qquad \hat{\beta}_{\text{GLS}} = \frac{\sum_{i=1}^n x_i' \Sigma^{-1} y_i}{\sum_{i=1}^n x_i' \Sigma^{-1} x_i}, \qquad \hat{\beta}_{\text{WG}} = \frac{\sum_{i=1}^n x_i' M y_i}{\sum_{i=1}^n x_i' M x_i},$$

where

$$\Sigma = \text{Var}(u_i) = \begin{pmatrix} \sigma_\varepsilon^2 + \sigma_\alpha^2 + 2\sigma_x^2\gamma^2 & \sigma_\alpha^2 + 2\sigma_x^2\gamma^2 \\ \sigma_\alpha^2 + 2\sigma_x^2\gamma^2 & \sigma_\varepsilon^2 + \sigma_\alpha^2 + 2\sigma_x^2\gamma^2 \end{pmatrix}, \qquad M = \begin{pmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{pmatrix}.$$

(a) Assume $\gamma = 0$. Are $\hat{\beta}_{\text{OLS}}$, $\hat{\beta}_{\text{GLS}}$, and $\hat{\beta}_{\text{WG}}$ consistent? Explain your answer.

(b) Assume $\gamma = 0$. Which of the consistent estimators in (a) has the smallest asymptotic variance (no proof required)?

(c) Assume $\gamma \neq 0$. Are $\hat{\beta}_{\text{OLS}}$, $\hat{\beta}_{\text{GLS}}$, and $\hat{\beta}_{\text{WG}}$ consistent? Explain your answer, and provide a consistency proof for one of the consistent estimators.

(d) Assume $\gamma \neq 0$. Somebody proposes to estimate β and γ by applying pooled OLS to the following regression model

$$y_{it} = x_{it}\beta + w_i\gamma + \tilde{u}_{it},$$

where $w_i = x_{i1} + x_{i2}$ and $\tilde{u}_{it} = \varepsilon_{it} + \tilde{\alpha}_i$. Would the resulting estimator for β be consistent? Explain your answer.

Problem 2 (Pooled OLS and Random Effects)

Consider the following panel data model with two time periods and one regressor

$$y_{it} = x_{it}\beta + u_{it}, \qquad\qquad u_{it} = \alpha_i + \varepsilon_{it},$$

where $i = 1, \ldots, n$ indexes individuals, and $t = 1, 2$ indexes time periods. We assume that x_{i1}, x_{i2}, α_i, ε_{i1} and ε_{i2} are all mutually independent, are all independently distributed across i, and that $x_{it} \sim \mathcal{N}(0, \sigma_x^2)$, $\alpha_i \sim \mathcal{N}(0, \sigma_\alpha^2)$, and $\varepsilon_{it} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. We assume that $\sigma_x^2 > 0$, $\sigma_\alpha^2 > 0$, $\sigma_\varepsilon^2 > 0$.

For each i the two-vectors $y_i = (y_{i1}, y_{i2})'$, $x_i = (x_{i1}, x_{i2})'$ and $u_i = (u_{i1}, u_{i2})'$, and the 2×2 matrix $\Sigma = \mathbb{E}(u_i u_i' | x_i)$. Consider the following two estimators for the scalar parameter β

$$\begin{aligned}
\hat{\beta}^{\text{OLS}} &= \frac{\sum_{i=1}^n \sum_{t=1}^2 x_{it} y_{it}}{\sum_{i=1}^n \sum_{t=1}^2 x_{it}^2} = \frac{\sum_{i=1}^n x_i' y_i}{\sum_{i=1}^n x_i' x_i}, \\
\hat{\beta}^{\text{GLS}} &= \frac{\sum_{i=1}^n x_i' \Sigma^{-1} y_i}{\sum_{i=1}^n x_i' \Sigma^{-1} x_i}.
\end{aligned}$$

(a) Show that $\hat{\beta}^{\text{GLS}}$ is a consistent as $n \to \infty$.

(b) Find an expression for Σ in terms of σ_α^2, and σ_ε^2. Calculate the inverse of Σ.

(c) Find expressions for the asymptotic variance of the two estimators in terms of only σ_x^2, σ_α^2 and σ_ε^2 (you can use the general result from the lecture). Which estimator has the smaller asymptotic variance?

Problem 3 (Applied Panel Data Methods)

We are once again investigate the determinants of wages. This time we have a panel data at our disposal (KEANE.csv). Use the 'plm' command (from the 'plm' package) to estimate the panel data linear models and solve the following problems.

1. Investigate the dataset KEANE.csv. What are the dimensions of the panel? Is it balanced?
2. We would like to estimate the influence of education on the wage of an individual. Develop a model describing the wage of an individual as a function of education and other variables from the dataset which you think should be included. Think carefully about which variables to include, and which sign you would expect the coefficient on each variable to have.
3. Estimate your proposed model by pooled OLS using the standard variance-covariance matrix. Try using a robust variance estimators as well - what difference does this make?
4. Does including year dummies affect the results of the pooled OLS?
5. Test whether experience in any form has a significant effect on the wage using both F (Wald) and LR-tests.
6. Estimate the original model (from question 3) using within transformation (fixed effects).
7. Estimate the model using random effects.
8. Estimate the model using first differences.
9. Compare the results from the four models. What are the assumptions of each model? Are these plausible? What are the pros and cons of each of the models.