

ECON 704

Problem Set 2

All random variables in this problem set are scalar unless otherwise stated.

1. Suppose we are interested in the following model:

$$Y_i = X_i\beta + e_i \quad , \quad \mathbb{E}[X_i e_i] = \mathbb{E}[X_i^2 e_i] = 0$$

- a. Express the above as a moment equation model. That is, let $W_i = (Y_i, X_i)$ and write down the function $g(W_i, \beta)$ such that $\mathbb{E}[g(W_i, \beta)] = 0$.
 - b. Evaluate $G = \mathbb{E} \left[\frac{\partial g(W_i, \beta)}{\partial \beta} \Big|_{\beta} \right]$ and $\Omega = \mathbb{E}[g(W_i, \beta)g(W_i, \beta)']$.
 - c. Given $\{W_i\}_{i=1}^n \stackrel{i.i.d.}{\sim} F$, suppose we perform GMM using the $k \times k$ identity matrix as the weight matrix: $\mathbb{W}_n = \mathbb{I}_k$. Find the asymptotic distribution of $\hat{\beta}^{\text{GMM}}$. You may assume that all the conditions in Theorem 2 of the lecture slides hold.
 - d. Compute the optimal weight matrix.
 - e. Can you estimate β by OLS? How would you compare $\hat{\beta}^{\text{OLS}}$ with $\hat{\beta}^{\text{GMM}}$ using the optimal weight matrix?
2. Consider a panel data model with no covariates:

$$Y_{it} = \alpha + e_{it} \quad , \quad \mathbb{E}[e_{it}] = 0$$

Suppose also that $\mathbb{E}[e_{it}^2] = \sigma^2$, $\mathbb{E}[e_{it}e_{it-1}] = \rho$ for $t = 2, \dots, T$ and that $\mathbb{E}[e_{it}e_{it-k}] = 0$ for $k > 1$.

- a. Find the variance of $\hat{\alpha}^{\text{OLS}}$. Denote this variance \mathbb{V}^{C} .
 - b. Suppose now that $\rho = 0$. What is the variance of $\hat{\alpha}^{\text{OLS}}$? Denote this variance \mathbb{V}^{OLS} .
 - c. What is the Moulton factor in this example? That is, what is $\mathbb{V}^{\text{C}}/\mathbb{V}^{\text{OLS}}$? When do we expect it to be large?
 - d. A researcher wants to test the hypothesis that $\beta = 0$. What variance estimator should they use when $\rho \neq 0$?
3. Use the data in `mgy` for this exercise. [Meng, Qian and Yared \(2015\)](#) argue that the China's Great Famine, 1959 – 1961, occurred not because too little food was produced. Instead, an inflexible food procurement policy by the central government caused food to be poorly allocated across the different regions. In particular, too much food was removed from the rural, food-producing areas and sent into urban areas, resulting in high death rates in the rural areas.

To make this point, they show that food production and mortality is usually negatively correlated. However, this relationship became positively correlated during the famine. The main regression the authors are interested in is:

$$ldeath_{i,t} = lgrain_pred_famdum5860_{i,t}\beta_1 + lgrain_pred_{i,t}\beta_2 + lurbpop'_{j,t}\gamma_1 + ltotpop'_{i,t}\gamma_2 + \delta_t + \epsilon_{i,t}$$

Here, $ldeath_{i,t}$ is log number of deaths in province i during the year $t+1$. $lgrain_pred_{i,t}$ is log constructed grain production. $lgrain_pred_famdum5860_{i,t}$ is the interaction of $lgrain_pred_{i,t}$ with a dummy variable that takes value 1 in a famine year and 0 otherwise. δ_t is a full set of time fixed effects.

- a. Interpret β_1 and β_2 in the above regression.
- b. Estimate the above regression. Present (1) heteroskedasticity-consistent standard errors, (2) cluster-robust standard errors clustered at the province level and (3) cluster-robust standard errors clustered at the province and year level.
- c. Which of the standard errors in b. do you prefer? Explain.
- d. How many province and time periods are there? What concern might you have about the reliability of cluster-robust standard errors?