Lecture 16

Recap...

Two-Stage Least Squares:

Consider the model:

$$y_1 = \alpha_0 + \alpha_1 y_2 + \alpha_2 x_1 + u$$
$$y_2 = \beta_0 + \beta_1 y_1 + \beta_2 x_2 + \beta_3 x_3 + v$$

- Endog.: y_1, y_2;
- Exog.: const., x_1, x_2, x_3.

First question: Is it identified?

- In the first equation, the number of omissions (x_2, x_3) is greater than the number of endogenous variables in the system (G) minus one $\rightarrow$ It is over identified.
- In the second equation, the number of omissions is equal to $G - 1 \rightarrow$ It is identified.

Step 1: Find instrumental variables (IV) for y_1 and y_2.

- Regress y_1 on constant, x_1, x_2, x_3: find $\hat{y}_1 = \hat{\lambda}_0 + \hat{\lambda}_1 x_1 + \hat{\lambda}_2 x_2 + \hat{\lambda}_3 x_3$.
- Regress y_2 on constant, x_1, x_2, x_3: find $\hat{y}_2 = \hat{\gamma}_0 + \hat{\gamma}_1 x_1 + \hat{\gamma}_2 x_2 + \hat{\gamma}_3 x_3$.
- $\hat{y}_1$ and $\hat{y}_2$ are IVs (they depend only on the x's').

Step 2: Estimate the system using the instrumental variables.

Multicollinearity

Say you have $x_1, \ldots, x_k$.

If $\exists\ \lambda_i\ (i = 1, \ldots, k) : \sum_{i=1}^{k} \lambda_i x_i = 0 \rightarrow$ **Perfect** Multicollinearity.

- E.g.: $x_1 = \frac{1}{\lambda_1}\left(\lambda_2 x_2 + \ldots + \lambda_k x_k\right)$.

If $\lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_k x_k = v_t$, where v_t is a random error $\rightarrow$ **Imperfect** Multicollinearity.

Example:

x_2	x_3	x_3^*
10	50	52
15	75	74
18	90	97
24	120	109
30	150	153

- $x_3 = 5x_2\ ;\ x_3^* = 5x_2 + v$.

- Using both x_2, x_3: perfect multicollinearity (cannot regress).
- Using both x_2, x_3^*: imperfect multicollinearity (might regress).

Summarizing:

- If there is perfect multicollinearity:
  - $\hat{\beta}_i$ undefined;
  - $\text{Var}(\hat{\beta}_i) \to \infty$.
- If there is imperfect multicollinearity
  - It is possible to estimate;
  - However, $\text{Var}(\hat{\beta}_i)$ will increase (and the t statistics decrease) as the x's are more and more correlated.

Estimation under imperfect multicollinearity:

1. Unbiased, consistent (only perfect multicollinearity violates Gauss-Markov);
   - OLS is still BLUE (Best Linear Unbiased Estimator).
2. The std. errors are big, but correctly estimated.
3. It is hard to get a small std. error (a small t) when the x's are highly correlated.
   - A small number of observations will produce the same problem.
   - The solution is the same: Get more data!
4. Estimators can be sensitive to small changes in data.
   - That, too, can apply to small samples.

Diagnostic:

- When individual t-statistics are insignificant, but the F-test of the joint significance is significant (R^2 usually high).
  - For a model $y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + u$:
    - t-tests: $H_0 : \beta_2 = 0$, $H_0 : \beta_3 = 0 \to$ low $|t|$'s.
    - F-test: $H_0 : \beta_2 = \beta_3 = 0 \to F$-statistic greater than critical value.
- High pair-wise correlation (say, $\rho > 0.8$).
  - If some variable x_2 has high correlation, not with other variables individually, but with a linear combination of them (say, $ax_3 + bx_4$), that can be found by regressing x_2 on x_3 and x_4.
    - If $\hat{a}$ and $\hat{b}$ are significant, there is most likely multicollinearity.

What to do?

1. Do nothing: it is a data (small sample) problem, not a model problem;
2. Collect more data;
3. Drop a variable;
4. Transform the data (log, square, differencing *etc*).

LM test for adding variables:

For a **restricted** model $y = \beta_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$.

- Test against **unrestricted** model: $y = \beta_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \boldsymbol{\beta_{k+1} x_{k+1}} + u$.

<u>Step 1</u>: Estimate restricted model: save $\hat{u}_i$'s.

<u>Step 2</u>: Regress $\hat{u}_i$ on *all* regressors: $\hat{u}_i = \alpha_1 + \alpha_2 x_2 + \ldots + \ldots \alpha_k x_k + \boldsymbol{\alpha_{k+1} x_{k+1}} + v_i$.

Step 3: $N \cdot R^2 \sim \chi^2(\mathrm{df})$, where $N \cdot R^2$ is the sample size times the coefficient of determination and df, the number of restrictions.