Economics 672: Econometric Analysis II
Winter 2018

Matias Cattaneo
Department of Economics
University of Michigan
Authored collaboratively by PhD entering class of 2017

January 23, 2018

Outline

Introduction

- Last lecture
  - Least squares finite sample properties
- This lecture
  - Generalized Least Squares (finite sample)
  - Partitioned Regression (actually presented in Lecture 8)
  - Least Squares Asymptotics

Outline

GLS Assumptions

- We maintain the assumptions that the CEF as linear in parameters, X is full rank, and the errors are exogenous, but relax the assumption of homoskedasticity.
- The model is:

$$y = X\beta + \varepsilon$$

- But instead of $\mathbb{V}[y|X] = \sigma_0^2 I_n$, we assume $\mathbb{V}[y|X] = \Omega_0$
- The simplest version of this assumes that Ω_0 is a diagonal matrix with elements σ_i^2, but in general it can be any (symmetric!) matrix.

GLS as WLS

- The GLS estimator is defined as the minimizer of the following objective function:

$$\|y - X\beta\|^2_{\Omega_0^{-1}} = (y - X\beta)'\Omega_0^{-1}(y - X\beta)$$

- If Ω_0 is diagonal, this can be written as:

$$\sum_{i=1}^{n} \sigma_i^{-2}(y_i - x_i'\beta)^2$$

- Here we can see an example of how GLS is a specific case of WLS, in which we are simply using the conditional VCV matrix of y as the weights (assuming it is known).

GLS as transformed OLS

- The objective function can also be written as:

$$(y - X\beta)'\Omega_0^{-1}(y - X\beta) = \varepsilon'\Omega_0^{-1}\varepsilon = (\Omega_0^{-\frac{1}{2}}\varepsilon)'(\Omega_0^{-\frac{1}{2}}\varepsilon)$$

- We can think of this as minimizing the residuals of a transformed model:

$$\begin{aligned} \Omega_0^{-\frac{1}{2}}y &= \Omega_0^{-\frac{1}{2}}X\beta + \Omega_0^{-\frac{1}{2}}\varepsilon \\ \tilde{y} &= \tilde{X}\beta + \tilde{\varepsilon} \end{aligned}$$

GLS and Gauss-Markov

- The GLS estimator is simply the OLS estimator of the transformed model:

$$\hat{\beta}_{GLS} = \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{y} = \left(X'\Omega_0^{-1}X\right)^{-1}X'\Omega_0^{-1}y$$

- Furthermore, in the transformed model we have:

$$\mathbb{V}[\tilde{y}|X] = \Omega_0^{-\frac{1}{2}}\mathbb{V}[y|X]\Omega_0^{-\frac{1}{2}} = \Omega_0^{-\frac{1}{2}}\Omega_0\Omega_0^{-\frac{1}{2}} = I_n$$

- Which means that the Gauss-Markov assumption of spherical errors is satisfied by the transformed model, and therefore that the GSL estimator is the BLU estimator of β

Feasible GLS

- If Ω_0 is not actually known, then the GSL estimator is not feasible (we can't actually compute it).
- This suggests a two-step procedure, where we first estimate Ω_0 with the data, and then plug in the this estimate for form a feasible GLS (or FGLS) estimator of β.
- We can, for example, use LS to construct $\hat{\Omega}$, and then the FGLS estimator is:

$$\hat{\beta}_{FGLS} = \left(X'\hat{\Omega}_0^{-1}X\right)^{-1} X'\hat{\Omega}_0^{-1}y$$

- This process can be iterated but updating our estimate of Ω_0 in each stage, and using this estimate to calculate the FGLS estimate or β in the next stage
- If repeated ad infinitum, this results in what is known as the Continuously Updated GMM Estimator.

Outline

Asymptotics

- Drop the assumption of Gaussianity (i.e. that $y|X \sim \mathcal{N}(X\beta_0, \sigma_0^2 I_n)$)
- Add the assumption that $\{(y_i, x_i) : 1 \leq i \leq n\}$ are i.i.d.
- Write the LS estimator as:

$$\hat{\beta} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} x_i y_i\right) = \beta_0 + \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} x_i \varepsilon_i\right)$$

- We can see that the expressions in parentheses are averages of i.i.d. random vectors, so under a suitable LLN they should converge to their expectations, which we can write as:

$$\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right) = \frac{X'X}{n} \to_p \mathbb{E}[x_i x_i'], \quad \left(\frac{1}{n}\sum_{i=1}^{n} x_i \varepsilon_i\right) = \frac{X'\varepsilon}{n} \to_p \mathbb{E}[x_i \varepsilon_i]$$

Note on Exogeneity Assumptions

- The strongest form of exogeneity is:

$$\varepsilon_i \perp\!\!\!\perp x_i \tag{1}$$

- The 2nd form of exogeneity is:

$$\mathbb{C}ov[\varepsilon_i | x_i] = 0 \tag{2}$$

- The weakest form of exogeneity is:

$$\mathbb{E}[x_i \varepsilon_i] = 0 \tag{3}$$

- In general (1) $\Rightarrow$ (2) $\Rightarrow$ (3) (not the other way around!)

Asymptotics

- So, as long as $\mathbb{E}[x_i\varepsilon_i]$ we will get consistency of the LS estimator for β_0
- Which exogeneity assumption we invoke will affect how we can interpret our estimates.
  - If we assume independence between errors and covariates (as in an RCT), then we can give a strong causal interpretation to the coefficients being estimated
  - If we use the conditional mean 0 assumption, then the coefficients are only causal inasmuch as we can parametrically model the CEF
  - If we use the 0 covariance assumption, then we can only interpret the coefficients as those of the best linear approximation to the CEF

Discussion

Other topics discussed in this class included:

- The dispute between those who are comfortable with the "best linear approximation" interpretation and use it to justify specification of a linear model, as opposed to those who think that any linear model is fundamentally misspecified and think it's better to specify a non-linear (e.g. logit, probit) model when you know that the truth is not linear
- The "principle of correspondence" and why we even bother with thinking about the finite sample properties and normalit of a LS model when we get almost all the same properties asymptotically
    - One answer was that large sample approximations very often lead to the same inferences as if the finite samples was assumed to follow a normal distribution

Outline

Next Class Topics and Readings

- Problem Set #
- Next topic: blah blah blah
- **Readings:**
    1. books
    2. more books