Boston University Questrom School of Business

MF840 – Spring 2021

Eric Jacquier

REVIEW OF OLS

GLS

SUR

SUR: Seemingly Unrelated Regressions

Similar to but more general than the ***multivariate regression model***

Linear Regression Model: $\underset{T\times 1}{Y} = \underset{T\times K}{X} \; \underset{K\times 1}{\beta} + \varepsilon$

1 OLS

- Min $\varepsilon'\varepsilon$ $\Leftrightarrow$ Min $(Y - X\beta)'(Y - X\beta)$

  Min $(-2\beta'X'Y + \beta'X'X\beta)$

$$0 = -2X'Y + 2\beta X'X$$

$$\hat{\beta} = (X'X)^{-1} X' Y$$

$$= (X'X)^{-1} X' (X\beta + \varepsilon)$$

$$\hat{\beta} = \beta + (X'X)^{-1} X'\varepsilon \qquad [1]$$

No statistics yet in this, we did not specify anything for the distribution of ε

- Properties of $\hat{\beta}$, is it a good estimator?

Bias? Do we get the correct number *on average, sample after sample*, if we use OLS?

Precision? How far do we deviate from the correct number *on average sample after sample*?

"On average, time after time": **repeated sampling**

=> we get different realizations of ε each time, therefore different β.

- **Unbiasedness**

$E[\hat{\beta}_{OLS}] \quad = \beta \quad + E[\ (X'X)^{-1}\ X'\ \varepsilon\ ] \qquad ?$

If $\qquad E(\varepsilon \mid X) = 0$ [2]

Then: $E(\varepsilon) = 0$, $Cov(g(X), \varepsilon) = 0$, $\mathbf{E(\ g(X)\ \varepsilon\ ) = 0}$. (LN9, MF793)

Then: $\quad \mathbf{E(\hat{\beta}_{OLS})} \quad = \beta + E[\ (X'X)^{-1}X'\ \varepsilon\ ] \quad = \boldsymbol{\beta}$ [3]

- Questions:

1. What is the only source of randomness ?

2. Did we need to know the *distribution* of ε ?

- Question: When would we **not** have $E(\varepsilon \mid X) = 0$?

- **(Co)Variance:** $V(x) = E[\ (x-E(x))\ (x-E(x))'\ ]$ **[4]**

Apply $\hat{\beta}$: $V(\hat{\beta}_{OLS}) = E[\ (\hat{\beta}-E(\hat{\beta}))\ (\hat{\beta}-E(\hat{\beta}))'\ ]$ k x k matrix

By condition [2], $E(\hat{\beta}_{OLS}) = \beta$, (4) becomes:

$$V(\hat{\beta}_{OLS}) = E[\ (\hat{\beta} - \beta)\ (\hat{\beta} - \beta)'\ ]$$
$$= E[\ (X'X)^{-1} X' \varepsilon\ \ ((X'X)^{-1} X' \varepsilon)'\ ]$$
$$= (X'X)^{-1} X'\ E(\varepsilon\varepsilon')\ X (X'X)^{-1}$$

New assumption: the noise is i.i.d. $E(\varepsilon\varepsilon') = \sigma^2 I_T$

$$\boldsymbol{V(\hat{\beta}_{OLS})} = (X'X)^{-1} X'\ E(\varepsilon\varepsilon')\ X (X'X)^{-1} = \boldsymbol{\sigma^2 (X'X)^{-1}}$$ **[5]**

Question: What if X is random?

A bit more complex, [5] is just the conditional variance of $\hat{\beta}$ (conditional on X), we would have to take the expectation of [5] around X.

- Optimality result: (non-proven): **Gauss-Markov Theorem:**

$\hat{\beta}_{OLS}$ is the Best Linear Unbiased Estimator (BLUE)

Linear: of the form $\hat{\beta} = K\,Y$

Unbiased: $E(\hat{\beta}) = \beta$

"Best": has the smallest possible variance

- **Distribution of $\hat{\beta}_{OLS}$**

Two possibilities

1. ε normally distributed

   => $\hat{\beta}_{OLS}$ **exactly normally** distributed, as a linear combination of normals

2. Don't know ε's distribution, but have a large sample:

   => $\hat{\beta}_{OLS}$ **approximately normally** distributed, as a linear combination of a large number of random variables.

- **We now need to estimate σ**

OLS has nothing to say about the estimation of σ.

We have an estimate of ε, the residual **e**: $\mathbf{e} = Y - X\,\hat{\beta}_{\mathbf{OLS}}$

We can compute its variance**:** $s^2 = \dfrac{e'e}{T-K}$

Why T-K, why not T ?
Recall the proof using the trace and the M matrix, the $\chi^2(T\text{-}k)$ result

- Finally, our estimate of the variance of $\hat{\beta}_{\mathbf{OLS}}$**:** $\mathbf{s^2(X'X)^{-1}}$

- Distribution of one element, $\hat{\beta}^k_{OLS}$: Student-t with T-K degrees of freedom.

2 GLS – Generalized Least Squares Framework

2.1 Complications $E(\varepsilon\varepsilon') \neq \sigma^2 I$ $E(\varepsilon\varepsilon') = \sigma^2\, \Omega = \Sigma$

- We write $\sigma^2\, \Omega$ to have a notation where the general model nests the iid model.

  But of course in many situations, σ is not identified. $\sigma^2\, \Omega = (2\sigma^2)\ (\Omega/2)$

- Heteroskedasticity: $E(\varepsilon\varepsilon') = \sigma^2 D$, $E(\varepsilon_i^2) \neq E(\varepsilon_j^2)$

$$\sigma^2 \mathbf{\Omega} = \sigma^2 \begin{bmatrix} \omega_1 & 0 & \cdots & 0 \\ 0 & \omega_2 & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & \omega_n \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}.$$

  - Examples:
    - Error variance proportional to one of the regressors: $\sigma_t = \sigma\sqrt{x_t}$
    - Error variance varies with time or previous shock: $\sigma_t = \alpha + \delta\, \varepsilon_{t-1}^2$

      ... the ARCH model

- Correlated errors $E(\varepsilon_i\varepsilon_j)\neq 0$ for $i\neq j$, $E(\varepsilon\varepsilon') = \sigma^2\,\Omega$

  For homoskedastic, correlated errors, σ^2 is the constant error variance and Ω is a correlation matrix.

- Examples:
  - Cross-sectional regression

    Group effects: $i\neq j$ same industry pair, $E(\varepsilon_i\varepsilon_j) \neq 0$

    $i\neq j$ different industries, $E(\varepsilon_i\varepsilon_j) = 0$

  - Time Series regression

    Autocorrelation of order 1: $\varepsilon_t = \rho\,\varepsilon_{t-1} + \nu_t$ AR(1)

$$\sigma^2\boldsymbol{\Omega} = \sigma^2 \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \cdots & \rho_{n-2} \\ & & \vdots & \\ \rho_{n-1} & \rho_{n-2} & \cdots & 1 \end{bmatrix}$$

where $\rho_K = \mathrm{Corr}(\varepsilon_t, \varepsilon_{t-k})$

2.2 Theoretical Solution: GLS estimator

- Matrix Result: For any covariance matrix Ω, $\exists$ P and D matrices such that

$$\Omega = P' D P \qquad [1]$$

Where $P'P = P P' = I$, D is Diagonal.

P: matrix of Eigen vectors, D: matrix of Eigen Values

- $D^{-0.5} P \;\; \boldsymbol{\Omega} \;\; P' \; D^{-0.5} = D^{-0.5} P \;\; P'DP \;\; P' \; D^{-0.5} = D^{-0.5} \; D \; D^{-0.5} = I_T$

$$P^* \;\; \Omega \;\; P^{*\prime} = I$$

- Pre-multiply our regression: $P^* Y = P^* X \beta + P^* \varepsilon \qquad E(\varepsilon\varepsilon') = \sigma^2 \Omega$

$$Y^* = X^* \beta + \varepsilon^* \qquad [2]$$

- Look at the error covariance matrix of [2]

$$E(\varepsilon^* \varepsilon^{*\prime}) = E(P^* \varepsilon \, \varepsilon' P^{*\prime}) = P^* \sigma^2 \Omega \; P^{*\prime} = \sigma^2 I$$

OLS applies to [2], it is BLUE for [2]

- GLS is just OLS applied to the transformed system [2]

$$\hat{\beta}_{GLS} = (X^{*\prime} X^{*})^{-1} X^{*\prime} Y^{*} = \mathbf{(X'\Omega^{-1}X)^{-1} X'\Omega^{-1} Y} \qquad [3]$$

$$\mathbf{Var(\hat{\beta}_{GLS})} = \sigma^2 (X^{*\prime}X^{*})^{-1} = \boldsymbol{\sigma^2 (X'\Omega^{-1}X)^{-1}}. \qquad [4]$$

- We are not done!

We need to find an estimate of Ω to plug into [3][4]: $\hat{\Omega}$ **Feasible GLS**

- Ω is a TxT matrix.

  Under i.i.d. errors, we reduced it to 1 parameter, σ.

  Now we have T(T+1)/2 parameters and only T observations

- **Can't** estimate T(T+1)/2 parameters with T observations

Need to specify a model of correlated and / or heteroskedastic errors to reduce the number of parameters in Ω.

Examples: autocorrelated residuals with AR(1), Weighted Least Squares, GARCH errors

Problem: Wrong $\hat{\Omega}$ means wrong $\hat{\beta}_{GLS}$

2.3 "Feasible" GLS Approach

If we could find a "good" estimate of Ω, it would be that way:

- Feasible GLS: 1) Do OLS => OLS residual: e_{OLS}

  2) Compute a covariance matrix of the residual given our model.

  Likely that OLS errors yield a consistent estimate of Ω: $\widehat{\Omega}^{(1)}$

  3) Use $\widehat{\Omega}^{(1)}$ to get $\hat{\beta}_{GLS}$ by [3]

- **Iterated GLS**: 4) iterate: $e_{GLS} = Y - X\hat{\beta}_{GLS}$ => $\widehat{\Omega}^{(2)}$ => new e_{GLS}, etc…

  Applications: autocorrelated residuals, Cochrane-Orcutt
  simple forms for heteroskedasticity,
  weighted least squares

- Again remember: Wrong $\widehat{\Omega}$ means wrong $\hat{\beta}_{GLS}$

- GLS is sensitive to assumptions on the covariance matrix

  It can be biased and inefficient if assumptions on Ω are wrong

  Then its covariance matrix estimate is wrong too: $V(\hat{\beta}_{GLS}) \neq \sigma^2(X'\Omega^{-1}X)^{-1}$

2.4 Alternative to GLS: Remain with OLS, compute Robust standard Errors

OLS is possibly inefficient but we can adjust its standard errors

Use robust standard errors for:
Unspecified heteroskedasticity White (1983)
Unspecified autocorrelation Hansen-Hodrick, Newey-West.

You may not have the "best" estimator but you can specify the uncertainty properly.

- **HAC** standard errors (**H**eteroskedasticity – **A**utocorrelation – **C**onsistent)

  May not have the "best" estimator but you can specify the uncertainty properly

In R: command coeftest, needs packages sandwich and lmtest.

2.5 Regression with AR(1) errors, Cochrane-Orcutt, a Feasible GLS

$y_t = \alpha + \beta\, x_t + \varepsilon_t$ $\qquad\qquad$ $\varepsilon_t = \rho\, \varepsilon_{t-1} + \nu_t$, $\quad E(\nu_{t,} \nu_{t-k}) = 0$, $E(\nu_t^2) = \sigma_\nu^2$

$E(\varepsilon\varepsilon') =$

Feasible GLS:

1) Run OLS, Get $e_{OLS,}$ OLS is inefficient but unbiased and consistent

2) Compute $\widehat{\Omega}^{(1)}$: $\hat{\gamma}_1 = \frac{1}{T-1} \sum e_t\, e_{t-1}$ $\qquad$ $\hat{\gamma}_0 = \frac{1}{T-1} \sum e_t^2$

3) Use $\widehat{\Omega}^{(1)}$ to compute $\hat{\boldsymbol{\beta}}_{\mathbf{GLS}}$

Intuition:

Compute $Y_t^* = Y_t - \hat{\rho}\, Y_{t-1}$, X_t^*, Consider regression of Y_t^* on X_t^*. $\quad$ t = **2**, ..., T

What can you say of $\quad E(\varepsilon^*\, \epsilon^{*\prime})$?

What is P* ?

3. SUR – SEEMINGLY UNRELATED REGRESSIONS (Arnold Zellner AZ)

- SUR is a fundamental model for the simultaneous analysis of several of regressions

- Say we regress portfolio 1 excess return r_1 on its benchmark return B_1, to find its systematic risk (β_1) and abnormal performance (α_1):

$$r_{1t} = \alpha_1 + \beta_1 B_{1t} + \varepsilon_{1t} \qquad t = 1, \ldots T, \; E(\varepsilon_1\varepsilon_1') = \sigma^2_1 I_T$$

  … and for portfolio 2:

$$r_{2t} = \alpha_2 + \beta_2 B_{2t} + \varepsilon_{2t} \qquad t = 1, \ldots T, \; E(\varepsilon_2\varepsilon_2') = \sigma^2_2 I_T$$

  … and 3, and 4, …. and N

- We assume that OLS assumptions are OK for each of these N … **seemingly unrelated** regressions.

- Still these regressions may be related via their errors: $\mathbf{E(\varepsilon_{it}\,\varepsilon_{jt}) = \sigma_{ij} \neq 0}$
  Does it matter? Yes.

  In this case, a more efficient GLS estimator follows by stacking these regressions.

- $$\underset{NT \times 1}{\begin{bmatrix} \underline{Y}_1 \\ \underline{Y}_2 \\ \vdots \\ \underline{Y}_N \end{bmatrix}} = \underset{NT \times NK}{\begin{array}{c} \overbrace{\quad}^{K} \qquad\qquad\qquad \\ \begin{bmatrix} X_1 & & & 0 \\ 0 & X_2 & 0 & \\ & 0 & \ddots & \\ & & & X_N \end{bmatrix} \end{array}} \underset{NK \times 1}{\begin{bmatrix} \underline{\beta}_1 \\ \underline{\beta}_2 \\ \vdots \\ \underline{\beta}_N \end{bmatrix}} + \begin{bmatrix} \underline{\varepsilon}_1 \\ \\ \\ \varepsilon_N \end{bmatrix}$$

- $$\underline{Y} = \underline{X}\,\underline{\beta} + \underline{\varepsilon}$$

- $$\underset{NT \times NT}{E(\varepsilon\varepsilon')} = E\left(\begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix}(\varepsilon_1' \cdots \varepsilon_N')\right) = E\begin{bmatrix} \varepsilon_1\varepsilon_1' & & & \\ & \varepsilon_2\varepsilon_2' & & \\ & \{\varepsilon_i\varepsilon_j'\} & \ddots & \\ & & & \varepsilon_N\varepsilon_N' \end{bmatrix}$$

- $$E(\varepsilon_j\varepsilon_j') = \underset{T \times T}{\begin{pmatrix} \sigma_j^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_j^2 \end{pmatrix}} \qquad E(\varepsilon_1\varepsilon_2') = \begin{pmatrix} \sigma_{12} & & 0 \\ & \ddots & \\ 0 & & \sigma_{12} \end{pmatrix}$$

Zero non contemporaneous correlations

$$\bullet\; E(\varepsilon\varepsilon') = \begin{bmatrix} \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_1^2 \end{pmatrix} & \begin{pmatrix} \sigma_{12} & & 0 \\ & \ddots & \\ 0 & & \sigma_{12} \end{pmatrix} & & \\ \begin{pmatrix} \sigma_{12} & & \\ & \ddots & \\ & & \sigma_{12} \end{pmatrix} & \begin{pmatrix} \sigma_2^2 & & \\ & \ddots & \\ & & \sigma_2^2 \end{pmatrix} & & \\ & & \ddots & \\ & & & \begin{pmatrix} \sigma_N^2 & & \\ & \ddots & \\ & & \sigma_N^2 \end{pmatrix} \end{bmatrix} = \Omega$$

$$= \begin{bmatrix} \sigma_1^2 I_T & \sigma_{12} I_T & \cdots & \sigma_{1N} I_T \\ \sigma_{21} I & \sigma_2^2 I & & \sigma_{2N} I \\ \vdots & & & \\ \sigma_{N1} I & \sigma_{N2} I & & \sigma_N^2 I \end{bmatrix} \equiv \underbrace{\begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ \{\sigma_{ij}\} & & \sigma_N^2 \end{pmatrix}}_{N \times N} \otimes I_{T \times T}$$

$$\equiv \underbrace{\Sigma_N}_{\substack{N \times N \text{ contemporaneous variance - covariance} \\ \text{matrix of the errors}}} \otimes I_T$$

- The correct LS estimator is the GLS on the "big" system:

$$\hat{\beta}_{SUR} = (\mathbf{X'}\Omega^{-1}\mathbf{X})^{-1}\,\mathbf{X'}\Omega^{-1}\mathbf{Y}$$
$$\text{Var}(\hat{\beta}_{SUR}) = (\mathbf{X'}\Omega^{-1}\mathbf{X})^{-1}$$

- Is there a feasible GLS for this? Can we get a consistent estimate of Ω?

  Yes we can: Ω is a huge matrix, NT by NT, but it is very sparse.

  Getting it right only requires a consistent estimator of Σ which is *only* NxN

  Condition: $T > N$ in fact $T \gg N$ is needed for reasonably precise results.

- Feasible GLS: Run N OLS regressions => $e_{i,OLS}$ => consistent $\hat{\Sigma}^{(1)}=\{e^{i'}e^{j}\}/(T-N)$

  $\hat{\Omega}^{(1)}_{GLS} = \hat{\Sigma}^{(1)}\otimes I$ => $\hat{\beta}_{GLS}$

  $e_{GLS} = \mathbf{Y} - \mathbf{X}\,\hat{\beta}_{GLS}$ => $\hat{\Omega}^{(2)}_{GLS}$

3.1 SUR when the independent variables X_i are the same

- All small X matrices are equal: $X_1 = X_2 = \ldots. = X_n = X$

$$\begin{pmatrix} Y_1 \\ Y_2 \\ . \\ Y_N \end{pmatrix} = \begin{pmatrix} X & 0 & . & 0 \\ 0 & X & . & 0 \\ . & . & . & 0 \\ 0 & 0 & . & X \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ . \\ \beta_N \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ \varepsilon_N \end{pmatrix} = (I \otimes X) \begin{pmatrix} \beta_1 \\ \beta_2 \\ . \\ \beta_N \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ \varepsilon_N \end{pmatrix}$$

What is the big **X** now: $\mathbf{X}_{NT,NK} = I_N \otimes X_{T,K}$, also remember: $\Omega_{NT} = \Sigma_N \otimes I_T$

$$\hat{\beta}_{SUR} = (\mathbf{X'} \; \Omega^{-1} \; \mathbf{X})^{-1} \; \mathbf{X'} \; \Omega^{-1} \; \mathbf{Y}$$

$$= [(I\otimes X)' (\Sigma\otimes I)^{-1} (I\otimes X)]^{-1} (I\otimes X)' (\Sigma\otimes I)^{-1} \mathbf{Y}$$

$$= [(I \, \Sigma^{-1} I) \otimes (X' \, I \, X)]^{-1} (\Sigma^{-1} \otimes X') \mathbf{Y}$$

$$= [\Sigma^{-1} \otimes (X'X)]^{-1} (\Sigma^{-1} \otimes X') \mathbf{Y} = [\Sigma \otimes (X'X)^{-1}] (\Sigma^{-1} \otimes X') \mathbf{Y}$$

$$= [I \otimes (X'X)^{-1} X'] \mathbf{Y}$$

$$= \hat{\beta}_{OLS}$$ ….. stacked OLS betas !

$V(\hat{\beta}_{SUR}) = (\mathbf{X'}\Omega^{-1}\mathbf{X})^{-1} = \Sigma \otimes (\mathbf{X'X})^{-1}$ … gives us **cross-equation covariances**