Economics 361
Method of Moments

Jun Ishii *
Department of Economics
Amherst College

Fall 2022

A Stroll through the Past: OLS, GLS, 2SLS

The estimation procedures we have examined thus far have been based on the ordinary least squares (OLS) approach. In the fortuitous case where the data generating process (DGP) underlying the given sample satisfied the classical OLS (Gauss Markov) assumptions, the OLS estimator is the best (minimum variance) linear unbiased estimator (BLUE/MVLUE) of BP$(Y|X)$. This result is courtesy of the Gauss-Markov Theorem and essentially states that one cannot find an unbiased, linear estimator that achieves lower variance (and, hence, lower mean-squared error (MSE) as we are only considering unbiased estimators).

We discussed scenarios where the Gauss-Markov assumptions were not satisfied. In particular, we spent time on the situation where $\text{Var}(Y|X) = \Sigma \neq \sigma^2 I$: the variance of Y conditional on X exhibited heteroskedasticity, autocorrelation, or both. Consequently, Gauss-Markov no longer held and applying OLS on the sample did not yield BLUE for $Y|X$. However, a modified OLS approach could possibly achieve BLUE status. The heuristic: we "transform" the sample into one where the classical OLS assumptions are satisfied. We introduced the matrix H defined such that $H'H = \Sigma^{-1}$ and showed that pre-multiplying H to the data yielded a transformed regression equation that satisfied the classical OLS assumptions

$$\underbrace{HY}_{\tilde{Y}} = \underbrace{HX}_{\tilde{X}}\beta + \underbrace{H\epsilon}_{\tilde{\epsilon}}$$

$$\begin{array}{rcl} E(\tilde{Y}|\tilde{X}) & = \tilde{X}\beta & \text{given } E(Y|X) = X\beta \\ V(\tilde{Y}|\tilde{X}) & = I & \text{given } V(Y|X) = \Sigma \\ (\tilde{X}'\tilde{X}) & \text{is non-singular} & \text{given } (X'X) \text{ is non-singular} \end{array}$$

Note that if we do not know Σ but know Ω (where $\Sigma = \alpha\Omega$ and α is an unknown scalar) then we can still apply GLS by choosing H such that $H'H = (\Omega)^{-1}$. But in this case, $\text{Var}(\tilde{Y}|\tilde{X}) = \alpha I$. α

*Office: Converse Hall 315 Phone: (413) 542-2901 E-mail: jishii@amherst.edu

would have to be estimated similar to the case of the unknown σ^2 in OLS (think $\alpha = \sigma^2$).

Lastly, we considered the scenario where the Linearity Condition, $E(Y|X) = X\beta$, is "violated" due to the inclusion of an "endogenous" variable in X.[1] We discussed three sources of endogeneity: omitted variables, measurement error, and simultaneity. We considered several alternative estimation procedures, including indirect least squares (ILS) and instrumental variables (IV), but focused primarily on a particular version of the least squares approach known as two-stage least squares (2SLS). The heuristic:

- We consider the regression equation implied by the expected value of Y conditional on some set of strictly **exogenous** variables – say Z
- We know that $E(Y|Z) = E(X|Z)\beta$. So, if we knew $E(X|Z)$, we could regress Y on $E(X|Z)$ to get an unbiased estimate of β. In general, we do not observe $E(X|Z)$. We use a proxy instead, the predicted value of X from OLS of X on Z.
- Regressing Y on this proxy yields an estimator of β with attractive asymptotic results: consistency (asymptotically converges in probability) and asymptotic normality

While all three estimation schemes (OLS, GLS, 2SLS) are based on the least squares approach, they also share more: each can be rationalized as an estimation procedure based on the moments/analogy principle: the normal equations (First Order Conditions) that define each estimation scheme are the sample analogy to population moments.

[1] e.g. Price and quantity in the Supply and Demand Model

OLS, GLS, and 2SLS as Moment Estimators

In OLS, we showed that the implicit moment condition imposed on the data was

$$
\begin{aligned}
\text{Estimator:} \quad & b = (X'X)^{-1}X'Y \\
\text{Regression Eqn:} \quad & Y = X\beta + \epsilon \\
\text{Population:} \quad & E(X'\epsilon) = 0 \\
\text{Sample (Normal Eqn):} \quad & X'e = X'(Y - Xb) = 0
\end{aligned}
$$

Similarly for GLS (Note: I use Σ here but can sub in Ω, instead, for case where α is unknown)

$$
\begin{aligned}
\text{Estimator:} \quad & b = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y \\
\text{Regression Eqn:} \quad & HY = HX\beta + H\epsilon \\
\text{Population:} \quad & E(\tilde{X}'\tilde{\epsilon}) = 0 \\
\text{Sample (Normal Eqn):} \quad & \tilde{X}'\tilde{e} = X'H'(H(Y - Xb)) = X'\Sigma^{-1}(Y - Xb) = 0
\end{aligned}
$$

And for 2SLS (Note: I define Z as set of *exogenous* variables here[2])

$$
\begin{aligned}
\text{Estimator:} \quad & b = (\hat{X}'\hat{X})^{-1}\hat{X}'Y \quad \text{where} \quad \hat{X} = Z(Z'Z)^{-1}Z'X \\
\text{Regression Eqn:} \quad & Y = X\beta + \epsilon \\
\text{Population:} \quad & \hat{X}'\epsilon \xrightarrow{p} 0 \quad \text{from} \quad E(Z'\epsilon) = 0 \\
\text{Sample (Normal Eqn):} \quad & \hat{X}'e = \hat{X}'(Y - \hat{X}b) = 0
\end{aligned}
$$

Note: If $E[\epsilon|X] = 0$ then $E[X'\epsilon] = 0$ by Law of Iterated Expectations. Consequently, the population moment condition for OLS is sometimes referred to as being $E[\epsilon|X] = 0$. Similarly, the population moment condition for GLS is sometimes referred to as being $E[\tilde{\epsilon}|\tilde{X}] = 0$.

For all three above, the first order conditions (Normal equations) that define the estimator is simply the sample analog to the underlying population moment condition.

- **OLS**: $E(X'\epsilon) = 0 \quad \longrightarrow \quad X'e = X'(Y - Xb) = 0$
- **GLS**: $E(\tilde{X}'\tilde{\epsilon}) = 0 \quad \longrightarrow \quad \tilde{X}'\tilde{e} = X'H'(H'(Y - Xb)) = X'\Sigma^{-1}(Y - Xb) = 0$
- **2SLS**: $\hat{X}'\epsilon \xrightarrow{p} 0$ from $E(Z'\epsilon) = 0 \quad \longrightarrow \quad \hat{X}'e = \hat{X}'(Y - \hat{X}b) = 0$

[2] I do this in order to make the definition of X consistent for all three cases: OLS, GLS, 2SLS. While Goldberger prefers to use X to refer always to the exogenous variables, others prefer Z with X referring always to the "right hand side" explanatory variables. Confusing ... I know

OLS as a Likelihood Estimator

Let us consider the classical OLS model but with an important addition: let us further assume that $Y|X$ is distributed Normally

$$\begin{aligned} Y|X &\sim N(X\beta, \sigma^2 I) \\ (X'X) &\text{ is } \text{non-singular} \end{aligned}$$

Note that for a generic sample of Normal variables (may not be random sample), $Y \sim N(\mu, \Sigma)$, the joint distribution can be written down as

$$\begin{aligned} f(Y) &= (2\Pi)^{-N/2} \, |\Sigma|^{-1/2} \, \exp\left\{(Y-\mu)'(2\Sigma)^{-1}(Y-\mu)\right\} \\ f(Y_i) &= \frac{1}{\sqrt{2\Pi\sigma_i^2}} \, \exp\left\{\frac{(Y_i-\mu_i)^2}{2\sigma_i^2}\right\} \end{aligned}$$

So for the regression model given above

$$\begin{aligned} f(Y|X) &= (2\Pi)^{-N/2} \, |\sigma^2 I|^{-1/2} \, \exp\left\{(Y-X\beta)'(2\sigma^2 I)^{-1}(Y-X\beta)\right\} \\ &= (2\Pi)^{-N/2} \, [\ (\sigma^2)^N\ ]^{-1/2} \, \exp\left\{\frac{1}{2\sigma^2}(Y-X\beta)'(Y-X\beta)\right\} \\ &= (2\Pi)^{-N/2} \, [\ (\sigma^2)^N\ ]^{-1/2} \, \exp\left\{\frac{1}{2\sigma^2}\epsilon'\epsilon\right\} \end{aligned}$$

Consider the natural log-transformation of the above joint distribution[3]

$$\underbrace{\ln(f(Y|X))}_{L(\beta,\sigma^2)} = -\frac{N}{2}\ln(2\Pi) - \frac{N}{2}\ln(\sigma^2) - \frac{1}{2}\left(\frac{\epsilon'\epsilon}{\sigma^2}\right)$$

One principle of estimation is to "choose" the values of (β, σ^2) such that they maximize the "likelihood" that the data generating process would yield the given, observed sample (Y, X). So replace (b, s^2) for (β, σ^2) and find the value of (b, s^2) that maximizes $L(b, s^2)$. Consider the relevant first order conditions (FOC)

$$\begin{aligned} \frac{\partial L}{\partial b} &= -\left(\frac{1}{2s^2}\right)\frac{\partial e'e}{\partial b} = 0 \quad \text{where } e = Y - Xb \\ \frac{\partial L}{\partial s^2} &= -\frac{N}{2s^2} + \frac{e'e}{2s^4} = 0 \end{aligned}$$

Note that the first FOC simply implies that b should be chosen to minimize $e'e$ which is simply ... the sum of squared residuals (SSR)!!! So the **maximum likelihood** estimator of β is the same as the OLS estimator of β. Moreover, the second FOC can be set to zero and solved such that $s^2 = \frac{e'e}{N}$ ... which is the consistent estimator of σ^2.

[3]Note that natural log is a monotonic transformation; furthermore, $f(Y|X)$ is strictly non-negative. Therefore, maximizing the log transformation is the same as maximizing $f(Y)$

Maximum Likelihood (Goldberger Ch. 12)

Asymptotic properties of the maximum likelihood estimator (MLE)

1. MLE is consistent: $\hat{\theta}_{ML} \xrightarrow{p} \theta$
2. MLE is asymptotically normal: $\sqrt{N}(\hat{\theta}_{ML} - \theta) \xrightarrow{d} N(0, \phi^2)$ therefore $\hat{\theta}_{ML}|X \overset{a}{\sim} N(\theta, \frac{\phi^2}{N})$

   $\phi^2 = \frac{1}{\mathrm{Var}(Z|X)} = \frac{1}{E(W|X)}$ where $Z = \frac{\partial L(\theta)}{\partial \theta}$ and $W = -\frac{\partial Z}{\partial \theta}$
3. MLE is the "best asymptotically normal" (BAN) estimator

   MLE asymptotically achieves the Cramer-Rao Lower Bound
4. Invariance property: the MLE of $h(\theta)$ where $h(\cdot)$ is a monotonic function of θ is simply $h(\hat{\theta}_{ML})$

Some important **caveats**

- In order for the properties above to hold, we must **know** the distribution from which the sample is generated. One pseudo-exception: **quasi-maximum likelihood**. Apply MLE assuming normality. This is because the quasi-ML estimator coincides with the OLS estimator and, therefore, have the same properties as the OLS estimator (but will not necessarily have the MLE properties listed above).[4]
- The properties are asymptotic. The finite sample properties of MLE differ depending on the assumed distribution $f(Y)$
- In general, MLE does not have an all-purpose analytical form, such as with OLS. The earlier estimator holds only for the case where the sample is distributed the given *i.i.d.* Normal process.
- Note that the variance of MLE depends on the true parameter values. Hence, we use the consistent estimate of ϕ^2. For the case of our earlier OLS example, the variance of (β, σ^2) is a function of (β, σ^2) (See below). Therefore, we replace (β, σ^2) with (b, s^2) to get a consistent estimate of ϕ. This requires (b, s^2) to be consistent estimates of (β, σ^2)

Example: Variance of the "OLS" ML Estimator

The variance of (b, s^2)

$$\underbrace{\begin{pmatrix} \frac{\partial^2 L}{\partial \beta \partial \beta'} & \frac{\partial^2 L}{\partial \beta \partial \sigma^{2\prime}} \\ \frac{\partial^2 L}{\partial \sigma^2 \partial \beta'} & \frac{\partial^2 L}{\partial \sigma^2 \partial \sigma^{2\prime}} \end{pmatrix}}_{-W} = \begin{pmatrix} -\frac{X'X}{\sigma^2} & -\frac{X'\epsilon}{\sigma^4} \\ -\frac{\epsilon' X}{\sigma^4} & \frac{N}{2\sigma^4} - \frac{\epsilon'\epsilon}{\sigma^6} \end{pmatrix}$$

$$\phi^2 = \frac{1}{E(W)} = \begin{pmatrix} \sigma^2 (X'X)^{-1} & 0 \\ 0' & \frac{2\sigma^4}{N} \end{pmatrix}$$

[4] For the case where the conditional mean of the distribution is a differentiable non-linear function of the parameters, quasi-ML estimator coincides with the Non-Linear Least Squares estimator