Session 7: OLS performance

MGT 581 | Introduction to econometrics

Michaël Aklin

PASU Lab | EPFL

Last time...

- Model evaluation

Today:

- Performance of OLS
- Bias/unbiasdness
- Consistency
- Efficiency

Readings:

- Stock and Watson (2011), ch 5, 7
- Verbeek (2018), ch 2.4-2.6, 4

Performance of OLS

What we want from an estimator...

- **Unbiasedness**: $E[\hat{\beta}] = \beta$
- **Consistency**: $\hat{\beta}$ converges in probability to β.

$$\lim_{n\to\infty} Pr[|\hat{\beta} - \beta| > \varepsilon] = 0 \forall \varepsilon > 0.$$

- **Efficiency**: $Var(\hat{\beta})$ should have the smallest feasible variance

Unbiasedness

Assumptions so far:

- $V(X) > 0$ (in all case)
- $\mathbf{X}'\mathbf{X}$ is invertible (in the multiple regression case).
- With this: OLS will yield a unique solution $\hat{\beta}$ to the SSR minimization problem
- Is this $\hat{\beta}$ unbiased? Let's derive it!

Starting point:

$$\begin{aligned}\hat{\beta} &= \frac{s_{yx}}{s_{xx}} \\ &= \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}\end{aligned}$$

Useful to simplify numerator:

$$
\begin{aligned}
\sum(x_i-\bar{x})(y_i-\bar{y}) &= \sum[y_i(x_i-\bar{x})-\bar{y}(x_i-\bar{x})] \\
&= \sum y_i(x_i-\bar{x})-\sum\bar{y}(x_i-\bar{x}) \\
&= \sum y_i(x_i-\bar{x})-\left(\sum\bar{y}(x_i)-\sum\bar{y}\bar{x}\right) \\
&= \sum y_i(x_i-\bar{x})-(\bar{y}\sum x_i-n\bar{y}\bar{x}) \\
&= \sum y_i(x_i-\bar{x})-(n\bar{y}\bar{x}-n\bar{y}\bar{x}) \\
&= \sum y_i(x_i-\bar{x})
\end{aligned}
$$

- We can rewrite $\hat{\beta}$:

$$\hat{\beta} = \frac{\sum y_i(x_i - \bar{x})}{\sum(x_i - \bar{x})^2}$$

- We need to bring in β to related $E[\hat{\beta}]$ to its true value.
- New Assumption: **correct specification of the model**. The population model is: $y = \mathbf{X}\beta + \varepsilon$.

Then:

$$\begin{aligned}
\sum y_i(x_i - \bar{x}) &= \sum[(\alpha + \beta x_i + \varepsilon_i)(x_i - \bar{x})] \\
&= \sum \alpha(x_i - \bar{x}) + \sum \beta x_i(x_i - \bar{x}) + \sum \varepsilon_i(x_i - \bar{x}) \\
&= \alpha \sum(x_i - \bar{x}) + \beta \sum x_i(x_i - \bar{x}) + \sum \varepsilon_i(x_i - \bar{x}) \\
&= 0\alpha + \beta \sum x_i(x_i - \bar{x}) + \sum \varepsilon_i(x_i - \bar{x}) \\
&= \beta \sum(x_i - \bar{x})^2 + \sum \varepsilon_i(x_i - \bar{x})
\end{aligned}$$

We now have:

$$
\begin{aligned}
\hat{\beta} &= \frac{\beta \sum (x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} + \frac{\sum \varepsilon_i (x_i - \bar{x})}{\sum (x_i - \bar{x})^2} \\
&= \beta + \frac{\sum \varepsilon_i (x_i - \bar{x})}{\sum (x_i - \bar{x})^2} \qquad (1)
\end{aligned}
$$

Is $E(\hat{\beta}) = \beta$?

$$
\begin{aligned}
E(\hat{\beta}) &= E\left(\beta + \frac{\sum \varepsilon_i (x_i - \bar{x})}{\sum (x_i - \bar{x})^2}\right) \\
&= E(\beta) + E\left(\frac{\sum \varepsilon_i (x_i - \bar{x})}{\sum (x_i - \bar{x})^2}\right) \\
&= \beta + E\left(\frac{\sum \varepsilon_i (x_i - \bar{x})}{\sum (x_i - \bar{x})^2}\right)
\end{aligned}
$$

Need a new assumption: (**exogeneity**)

$$E(\varepsilon|X) = 0$$

Note that this implies:

$$\begin{aligned} Cov(X, \varepsilon) &= 0, \\ E(X\varepsilon) &= 0 \end{aligned}$$

We can further simplify:

$$
\begin{aligned}
E(\hat{\beta}) &= \beta + E\left(\frac{\sum \varepsilon_i (x_i - \bar{x})}{\sum (x_i - \bar{x})^2}\right) \\
&= \beta + E(\varepsilon_i | X) \frac{\sum [(x_i - \bar{x})]}{\sum (x_i - \bar{x})^2} \\
&= \beta
\end{aligned}
$$

Similarly for multiple regression

$$\begin{aligned}
\hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \varepsilon) \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon \\
&= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon
\end{aligned}$$

$$
\begin{aligned}
E(\hat{\beta}) &= E(\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon) \\
&= E(\beta) + E((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon) \\
&= \beta + E((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon) \\
&= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\varepsilon|\mathbf{X}) \\
&= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{0} \\
&= \beta
\end{aligned}
$$

Omitted variable formula

- Question: what happens if we get the model wrong?
- Eg:
  - True model: $y = \lambda + \tau x + \gamma z + \mu$
  - Estimated model: $y = \alpha + \beta x + \varepsilon$
- We can use the result on unbiasedness!

$\frac{s_{xz}}{s_{xx}}$ can be understood as follows:

$$\begin{aligned} x &= \phi + \omega z + \rho \\ &\equiv \phi + z\frac{s_{xz}}{s_{xx}} + \rho \end{aligned}$$

Thus:

$$\begin{aligned} E(\hat{\beta}) &= \beta + \gamma\frac{s_{xz}}{s_{xx}} \\ &= \beta + \gamma\omega \end{aligned}$$

Therefore, the bias is:

$$E(\hat{\beta}) - \beta = \gamma\omega$$

Recall...

What we want from an estimator...

- **Unbiasedness**: $E[\hat{\beta}] = \beta$
- **Consistency**: $\hat{\beta}$ converges in probability to β.

$$\lim_{n \to \infty} Pr[|\hat{\beta} - \beta| > \varepsilon] = 0 \forall \varepsilon > 0.$$

- **Efficiency**: $Var(\hat{\beta})$ should have the smallest feasible variance

Inconsistency

With omitted var: not only is OLS biased, but it is also inconsistent:

$$\begin{aligned}
\hat{\beta} &= \beta + \frac{\sum \varepsilon_i (x_i - \bar{x})}{\sum (x_i - \bar{x})^2} \\
plim \hat{\beta} &= plim \left(\beta + \frac{\sum \varepsilon_i (x_i - \bar{x})}{\sum (x_i - \bar{x})^2} \right) \\
&= plim \beta + \frac{plim \left[\frac{1}{n} \sum \varepsilon_i (x_i - \bar{x}) \right]}{plim \left[\frac{1}{n} \sum (x_i - \bar{x})^2 \right]} \\
&= \beta + 0
\end{aligned}$$

The last step is due to the fact that $1/n \sum \varepsilon_i (x - \bar{x})$ converges to 0, iff exogeneity is met. If not: $\hat{\beta}$ does not converge to β.

(Note: $\hat{\beta}$ converges to a constant β as $n \to \infty$)

Therefore, a variable Z that is both causing X and Y directly will make $\hat{\beta}_{\textbf{ols}}$ biased and inconsistent if it is omitted.

- $\hat{\beta}$ will be **biased** $E(\hat{\beta}) \neq \beta$.
- $\hat{\beta}$ will be **inconsistent** (increasing the sample size won't help).
- Note that you can try to infer the sign of the bias caused by a missing variable W by formulating educated guesses about the effect it has on Y and D. Does its absnce push $\hat{\beta}$ to zero or does it inflate it? Example?

Efficiency

- *If* homoskedasticity and independence: opens the door of **Gauss-Markov theorem**
- Gauss-Markov: under assumptions of random sampling, correctly specified model, X is of full rank, exogeneity, homoskedasticity, independent error terms (no serial/spatial correlation), then OLS is the **b**est **l**inear **u**nbiased **e**stimator (BLUE)

Proof of Gauss-Markov (optional)

Let's sketch a proof of the Gauss-Markov theorem. What we want to show is that any non-OLS linear unbiased estimator of β has a higher variance than $\hat{\beta}_{ols}$. I'll refer to an arbitrary competing estimator $\tilde{\beta}$.

We start with the model:

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

Since we are talking about linear estimators, we can write such an estimators $\tilde{\beta}$ as a function of y:

$$\tilde{\beta} = m + My$$

where m and M are a vector and a matrix of constants, respectively. In OLS, for instance, $M = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

Next, we assume that $\tilde{\beta}$ is unbiased. This implies that:

$$\begin{aligned} E(\tilde{\beta}|\mathbf{X}) &= E(m + My|\mathbf{X}) \\ &= m + ME(y|\mathbf{X}) \\ &= m + ME(\mathbf{X}\beta + \varepsilon|\mathbf{X}) \\ &= m + M\mathbf{X}\beta + ME(\varepsilon|\mathbf{X}) \\ &= m + M\mathbf{X}\beta \end{aligned}$$

The requirement for $\tilde{\beta}$ to be unbiased, thus, is that

$$\begin{aligned} m &= 0 \\ M\mathbf{X} &= \mathbf{I} \end{aligned}$$

Note that OLS satisfies this. The implication is that a linear unbiased estimator will look like:

$$\tilde{\beta} = My,$$

with M satisfying the above requirement. So, a typical non-OLS estimator will be (w/o loss of generalizability):

$$M = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + C$$

where C is any matrix.

Since $\tilde{\beta}$ is unbiased, we know that $M\mathbf{X} = \mathbf{I}$ must hold. At the same time:

$$\begin{aligned} M\mathbf{X} &= [(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + C]\mathbf{X} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} + C\mathbf{X} \\ &= \mathbf{I} + C\mathbf{X} \end{aligned}$$

The implication is that $C\mathbf{X} = 0$.

So, we can now compute the variance of $\tilde{\beta}$. Let's write:

$$
\begin{aligned}
\tilde{\beta} &= My \\
&= M(\mathbf{X}\beta + \varepsilon) \\
&= M\mathbf{X}\beta + M\varepsilon \\
&= \mathbf{I}\beta + M\varepsilon \\
\tilde{\beta} - \beta &= M\varepsilon
\end{aligned}
$$

The variance-covariance matrix is thus:

$$\begin{aligned} E(\tilde{\beta}-\beta)(\tilde{\beta}-\beta)' &= E(M\varepsilon(M\varepsilon)') \\ &= E(M\varepsilon\varepsilon' M') \\ &= M(E(\varepsilon\varepsilon'))M' \\ &= M\sigma^2\mathbf{I}M' \\ &= \sigma^2 MM' \end{aligned}$$

So what is MM' again? It is:

$$
\begin{aligned}
MM' &= [(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + C][(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + C]' \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&+ (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'C' + C\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} + CC' \\
&= (\mathbf{X}'\mathbf{X})^{-1} + CC'
\end{aligned}
$$

Thus, the variances of $\tilde{\beta}$ and $\hat{\beta}_{ols}$ are:

$$\text{Alternative} : \sigma^2 MM' = \sigma^2[(\mathbf{X}'\mathbf{X})^{-1} + CC']$$
$$\text{OLS} : \sigma^2 MM' = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

Since CC' is a quadratic term, it is positive. Thus, the variance of the rival estimator $\tilde{\beta}$ is larger than the variance of the OLS estimator, except if CC' happens to be zero. But if it is, then the estimator is actually OLS.

[end of optional section]

- Gauss-Markov: neat result to justify use of OLS
- If assumptions are met: not just unbiased and consistent, but lowest sampling variance among unbiased estimators
- But: hard to make the assumption of homoskedasticity. Eg: non-linear relations create heteroskedasticity.
- That's why we often used "**robust**" standard errors (aka Huber-Eicker-White)

Conclusion

- OLS is popular because it performs well
- When standard assumptions are met: unbiased and consistent
- Add homoskedasticity and no error correlation: efficiency (BLUE)
- Latter is unlikely to hold. With heteroskedasticity: still unbiased, but not efficient anymore
- Threat from endogeneity and omitted variable: creates bias and inconsistency

Questions?

References

Stock, James H., and Mark W. Watson. 2011. *Introduction to Econometrics, 3rd Edition*. Pearson.

Verbeek, Marno. 2018. *A Guide to Modern Econometrics 5th Edition*. Wiley.