Regression: Why?

February 6, 2020

Three justifications for estimating a linear regression

- Best least squares fit in population: $\min E[(Y - X\beta)^2]$.
- Mean of $Y \mid X$ is linear, we want to find it: $E[Y \mid X] = X\beta$.
- $Y \mid X$ has a known (or parametrically estimable) distribution with $X\beta$ as a location parameter.

Population LS fit

- Useful for predicting Y from new draws from same population of X's.
- Does not require any assumptions on the distribution, other than finite variances.
- In particular, not necessarily true that $E[Y \mid X] = X\beta$.
- OLS achieves the semi-parametric efficiency bound.

Semi-parametric efficiency bound?

- But don't we know that GLS is more efficient than OLS?

- SEB: Any other estimator that gives improved performance under some distributional assumption, must give worse performance under some other distribution that satisfies the same minimal identifying moment condition. ($E[X'Y] = X'X\beta$).

- For example, suppose we assume $E[Y \mid X] = X\beta$ and estimate a form for $\Omega = E[\varepsilon\varepsilon' \mid X]$. Then GLS will be more efficient (have smaller $\mathrm{Var}(\hat{\beta})$ in large samples) than OLS.

- But if instead, say, $E[Y \mid X]$ is nonlinear, GLS will generally be worse than OLS, indeed generally inconsistent. [EG]

Other arguments for focusing on OLS

- Since we don't try to estimate a nonlinear $E[Y \mid X]$ or a fancy distribution for the residuals $Y - X\beta$, we don't have to rely on a lot of complicated assumptions and we keep the number of parameters to consider small.

- This makes it plausible that asymptotic distribution theory applies well in a moderate sized sample. See Angrist and Pischke (2010).

- This leads to a recommendation to use OLS and rely on its asymptotic distribution as normal with mean β and asymptotic covariance matrix

$$\mathrm{Var}(\hat{\beta}) = (X'X)^{-1}E[X'\varepsilon\varepsilon'X](X'X)^{-1}$$

Another level of SEB

- If the assumption $E[\varepsilon \mid X] = X\beta$ holds and Ω is known, GLS achieves the SEB, regardless of the distribution of the residuals, so long as they are of finite variance. In other words, GLS, which is the MLE and small-sample efficient under the normality assumption, cannot be improved on, asymptotically, if we try to estimate the distribution of ε and use those estimates to improve our estimates of β.

- But we know that, e.g., correctly modeling a t distribution for the errors delivers efficiency gains.

- Since the normal distribution is a special case of the t, how can it not give us an advantage if we model the errors as t at the start, so long as we estimate ν, $1/$(degrees of freedom of the t), and thus allow for $\nu = 0$, the normal case?

Including $\nu = 0$ is not enough

- Just as in the case of GLS when the model is nonlinear, here the fact that we allow for $\nu = 0$, where OLS would be optimal, does not save us if in fact the residuals are neither normally or t distributed.

- Müeller example.

An empirical strategy that avoids these objections

- Check for deviations from assumptions. Expand the model as sample size grows and information about the needed complications accumulates.

*

References

Angrist, J. D. and J.-S. Pischke (2010): "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics," *Journal of Economic Perspectives*, 24, 3–30.