1 Properties of estimators

1.1 Unbiasedness

Explanatory variables are supposed to be deterministic in elementary Econometrics, to show unbiasedness of Least squares estimates of linear models.

In more general treatments, the alternative assumption is made: $E(u/X) = 0$, which means: $E(x_{it}.u_s) = 0 \forall t, s$, which we usually know as strict exogeneity.

It is usually hard to make a strong argument on the validity of that condition.

It is easy to figure out why can it fail to hold, but it is much harder to argue in its favor.

Since

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

The condition implies:

$$E(\hat{\beta}) = \beta + E\left[(X'X)^{-1}X'u\right] = \beta + E\left[(X'X)^{-1}X'E(u/X)\right] = \beta$$

But, should we care about unbiasedness in Economics, being a property that relates to the universe of possible samples?

1.2 Variance-covariance matrix of estimates

If the vector error term has covariance matrix,

$$Var(u) = \sigma_u^2 \Sigma$$

The variance-covariance matrix of least squares estimates is,

$$Var(\hat{\beta}) = \sigma_u^2 (X'X)^{-1}(X'\Sigma X)(X'X)^{-1} \quad (1)$$

If de not allow for a scalar factor σ_u^2, which is not necessary, then $Var(u) = \Sigma$ and $Var(\hat{\beta}) = (X'X)^{-1}(X'\Sigma X)(X'X)^{-1}$.

To estimate Σ we will need to use residuals from some initial estimation.

So, we can start by using OLS, and use the residuals to estimate the structure we assume in Σ.

If, for instance, we postulate that $E(u_i.u_j) = 0 \forall i \neq j$, while $E(u_i.u_j) = kz_i$ for $i = j$, we will then run a regression of the square OLS residuals on z, without intercept.

Whether we identify σ_u^2 with k and Σ with a diagonal matrix with z_i along the diagonal, or make those elements equal to kz_i and skip the σ_u^2 factor, is irrelevant.

There are special cases, those in which Σ is almost diagonal, when the variance-covariance matrix reduces to $\sigma_u^2(X'X)^{-1}$, but it is unfortunate that this matrix is widely presented in a first discussion of least squares methods in econometrics textbooks as being the variance covariance matrix of the least squares estimator.

The elements of $\sigma_u^2(X'X)^{-1}$ are biased estimates of the variances and covariances of the least squares estimator, not bearing any specific relationship with the unbiased $\sigma_u^2(X'X)^{-1}(X'\Sigma X)(X'X)^{-1}$ values. The biased, standard estimates may be either larger or smaller than the unbiased ones without any special reason.

Nothing is lost by computing (1) in all situations.

1.3 Efficiency

The standard, efficiency properties of least squares shown in introductory courses emerge from its coincidence with Maximum Likelihood under a Normal distribution for the error term, and provided we have a right specification for the variance-covariance matrix of the error term.

The first condition is unlikely in many situations in Economics.

In general, efficiency is shown only under deterministic or strictly exogenous explanatory variables.

Heteroscedasticity leads to lack of efficiency in least squares estimation.

It does not bias the estimates or produce inconsistency.

Autocorrelation in static models has similar implications

Dealing with Heteroscedasticity or autocorrelation as usual (Feasible GLS) is usually subject to important sample errors

$\Rightarrow$ Use OLS and compute robust variance-covariance matrix of estimates: White, Newey-West

In general, it is hard to figure out the properties of least squares estimates.

$\Rightarrow$ we need to worry about consistency and precision (related to efficiency).

1.4 Consistency

Consistency is a one-sample property, and all it requires is: $p\lim\left(\frac{1}{T}X'u\right) = 0_k$.

$$p\lim(\hat{\beta}) = \beta + p\lim\left[\left(\frac{1}{T}X'X\right)^{-1}\left(\frac{1}{T}X'u\right)\right] = \beta + \left[p\lim\left(\frac{1}{T}X'X\right)^{-1}\right]\left[p\lim\left(\frac{1}{T}X'u\right)\right] = \beta$$

Under light conditions (law of large numbers) this condition will hold if the error term is uncorrelated with the set of explanatory variables.

It is important that we now do not need exogeneity.

All we need is lack of correlation between regressors and error term, i.e., we do not need zero autocorrelation at all leads and lags of X and u.

Situations when correlation is not zero:

- Simultaneity
- Errors in variables
- Dynamic models with autocorrelated errors

1.5 Instrumental variables

We then need instrumental variables, Z, satisfying $E(Z/u) = 0,$.at the same time $E(Z.X) \neq 0$.

We lose consistency if the first condition fails to hold, and we lose precision because the correlation between Z and X is less than one (otherwise, we would still have the lack of consistency situation).

In most cases, it is usually hard to figure out what are valid instruments outside the model, and often, models are silent with respect to valid instruments.

Models with expectations, or dynamic panel data models suggest instruments that are already present in the model.

Precision means that standard errors are small relative to estimated parameters.

Precision depends, among others on: the quantity and quality of data, parameter stability.

1.6 Hypothesis testing

Most often, we compare nested models, and versions of likelihood ratio tests are appropriate

We should specifically worry about testing hypothesis in the face of low precision estimates.

Do not run hypothesis tests in the face of estimates obtained with low precision

Low precision in estimation leads to a bias in the results of any given test by too often not rejecting the null hypothesis (any null hypothesis)

So, when running significance tests, we would tend to conclude for non informative explanatory variables to often.

The t-statistic for significance ids the ratio between the estimated coefficient and its estimated standard error. The t-statistic can be too low, leadgin to not rejecting the null hypothesis of lack of significance if: *i*) the estimated coefficient is small to the point of being numerically irrelevant, *ii*) the standard deviation is large enough, *i.e.*, precision is very low, even if the estimated coefficient is numerically sizeable, *iii*) both, *i*) and *ii*).

Summarizing the sample information regarding the validity of a given null hypothesis in the value of a single test statistic value is too much information is an excessive reduction of the available information

Always examine residuals (or fit) from restricted and unrestricted models

Relative to significance tests:

- statistical significance of a given coefficient and economic relevance (or quantitative relevance) of the accompanying variable are very different concepts
- to evaluate the relevance of an estimated coefficient, multiply it by the standard deviation of the associated variable, and divide by the standard

deviation of the dependent variable. Or do a similar computation for the whole sample range or the interquartilic intervals of x and y.

- we can never test for the information content of a given variable in the context of a multiple regression model
- we can only test for whether a given variable *adds information* to that contained in the other explanatory variables already included in the model
- to test for information content in an absolute sense, we should estimate a simple regression model
- the estimated coefficient in a simple regression is a *biased* estimate of the partial effect of x. But is is an *unbiased* estimate of the global effect (direct effect plus indirect effects) on y of a change in x.
- each estimated coefficient in a multiple regression is an unbiased estimate of the partial effect (conditional on the other explanatory variables) on y of a change in x. It is a biased estimate of the effect on y of a change in x.