Large Sample Property

Swiss Institute of Artificial Intelligence

Keith Lee

October 4, 2021

Large-Sample Analysis of OLS I

- Under the weaker assumption of **contemporaneous exogeneity** $\mathbb{E}(u_t|x_t) = 0$ we can obtain consistency of OLS in TS models
- Looks simular to unbiased result but for two key points:

Theorem

(Consistency of OLS for Time Series): Under the assumption of Stationarity and Weak Dependence, TS.1, TS.2, and TS.30, the OLS estimators are consistent

$$plim(\widehat{\beta}_j) = \beta_j, \;\; j = 0, ..., k.$$

- For this consistency result we will be assuming stationarity and weak-dependence which enables us to apply LLN and CLT.
- Unbiasedness requires TS.3 (strict exogeneity), consistency does not.
- Consistency requires weak dependence but unbiasedness does not.

Large-Sample Analysis of OLS II

- Consider the simple linear regression model:

$$
\begin{aligned}
y_t =& \beta_0 + \beta_1 x_t + u_t \ \ with \\
& \widehat{\beta}_1 = \beta_1 + \frac{\frac{1}{T}\sum(x_t - \bar{x})(u_t - \bar{u})}{\frac{1}{T}\sum(x_t - \bar{x})^2} = \beta_1 + \frac{SampleCov(x_t, u_t)}{SampleVar(x_t)}
\end{aligned}
$$

  - (Given stationarity and weak dependence) we can use LLN:

$$
\begin{aligned}
& plim \frac{1}{T}\sum(x_t - \bar{x})(u_t - \bar{u}) = Cov(x_t, u_t) \ \ and \\
& plim \frac{1}{T}\sum(x_t - \bar{x})^2 = Var(x_t), \ \ so \\
& plim \widehat{\beta}_1 = \beta_1 + \frac{Cov(x_t, u_t)}{Var(x_t)}
\end{aligned}
$$

  - Clearly for consistency we need $Cov(x_t, u_t) = 0$. Strict exogeneity, needed for unbiasedness, is not required.

Large-Sample Analysis of OLS for III

- The consistency result provides justification for estimating models where regressors are NOT strictly exogenous: while we cannot have unbiased estimators, they are consistent under fairly weak assumptions.
    - Consider the model:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 z_t + \beta_3 z_{t-1} + u_t, \ \textit{with} \ |\beta_1| < 1$$

        - y_{t-1} ensures that $\mathbb{E}(u_t|X) = 0$ **(TS.3) cannot be satisfied**
        - $\Rightarrow$ OLS be biased.
    - (Given stationariy and weak dependence (requires $|\beta_1| < 1$),) OLS will be consistent if we assume
        - $\mathbb{E}(u_t|x_t) = \mathbb{E}(u_t|y_{t-1}, z_t, z_{t-1}) = 0$ **(TS.3') is reasonable**
        - Requires that y_{t-1}, z_t and z_{t-1} are uncorrelated with u_t.

Large-Sample Analysis of OLS for IV

- For validity of the usual SEs, and test statistics, we will need to add the assumptions of homoskedastiticy and zero autocorrelation:
  - **TS.40 Homoskedasticity:** $Var(u_t|\mathbf{x_t}) = \sigma^2$
    - Weaker and more natural than TS.4 ($Var(u_t|\mathbf{x_1}, ..., \mathbf{x_t}, ..., \mathbf{x_T}) = \sigma^2$)
    - The error variance at time t does not depend on regressors x_t at time t.
  - **TS.50 No Serial Correlation:** For all $t \neq s$, $\mathbb{E}(u_t u_s|\mathbf{x_t}, \mathbf{x_s}) = 0$
    - If we include lagged dependent variable(s), Assumption TS.50 is more likely to be satistied
    - In fact, the goal is often to make it hold. The key point is this: if we include enough lagged y and possibly other variables there cannot be any serial correlation.

Large-Sample Analysis of OLS for V

Once we add these two assumptions (homoskedasticity and no serial correlation), we can rely on the standard inference procedures we used in cross section analysis.

Theorem

(Asymptotic normality of OLS for Time Series): Under the assumption of Stationarity and Weak Dependence, TS.1, TS.2, and TS.3'-TS.5', the OLS estimators is approximately normally distributed as $T \longrightarrow \infty$. Moreover, usual t-test, F-test and confidence interval are all asymptotically valid.