Econometrics

Chapter 16: Spurious Regression

In Choi

Sogang University

Spurious Correlation

- Yule (1926) tries to explain unreasonably high correlation between the mortality rate for the years 1866–1911 and the ratio of Church of England marriages to all marriages by using time series which, and the first difference of which, are positively correlated.
- See Aldrich (1995; "Correlations Genuine and Spurious in Pearson and Yule," Statistical Science, 10, 364–376.) for a historical account of spurious correlations.

Spurious Time Series Regression

- It is not uncommon to find, in applied time series econometric work, equation of an apparently high degree of goodness of fit (R^2), but with extremely low value of the D-W statistic.
- This phenomena are contradictory, because low value of the D-W statistic implies that the error terms are serially correlated, and that the model fitted is not adequate.
- Hence, low value of the D-W statistic should accompany low value of R^2, according to a standard econometric theory.

Spurious Time Series Regression

- This dubious regression results (high R^2, low D-W) are studied by simulations method in Granger and Newbold (1974; J. of Econometrics).
- Consider the regression equation

$$y_t = \alpha + \beta x_t + u_t \tag{1}$$

where

$$\begin{aligned} y_t &= y_{t-1} + v_t, \quad x_t = x_{t-1} + w_t \\ v_t &\sim N(0, \sigma_v^2), \quad w_t \sim N(0, \sigma_w^2) \end{aligned} \tag{2}$$

- Since v_t is independent of w_t, x_t and y_t have no statistical relation, and hence we expect that $H_0 : \beta = 0$ will be rejected by the usual t-test.

Spurious Time Series Regression

- Granger and Newbold found that the null hypothesis is rejected at the 5% level on about three-fourth of 100 simulations.
- This result shows an apparantely high degree of fit can very often be obtained simply by regressing independent random walks.
- Because the regression results are dubious in this case, we call the regression *"spurious"*.

Spurious Time Series Regression

- Phillips (1986; J. of Econometrics) develops asymptotic results for spurious regressions using the invariance principle.
- Phillips analyzes models (1) and (2) and reports that

$$\hat{\beta} \overset{p}{\nrightarrow} \beta$$

$$\begin{aligned} T^{-\frac{1}{2}}\hat{\alpha} &= O_p(1) \quad (\alpha \xrightarrow{p} 0) \\ T^{-\frac{1}{2}}t_\beta &= O_p(1) \quad (|t_\beta| \xrightarrow{p} \infty) \\ T^{-\frac{1}{2}}t_\alpha &= O_p(1) \quad (|t_\alpha| \xrightarrow{p} \infty) \\ R^2 &= O_p(1) \end{aligned}$$

$$DW \xrightarrow{p} 0$$

- These asymptotic results explain why we tend to reject the null hypotheses $\alpha = 0$ and $\beta = 0$ and find low D-W statistics in regressing two independent random walks.
- Furthermore, these results show that the OLS estimates of the coefficients α and β are not consistent.
- Phillips also analyzes the multiple regressions with integrated processes. All the results stated above still hold. This explains the simulations results of Granger and Newbold (1986, p.209).

Spurious Panel Regression

- See Choi (2013; Oxford Bulletin of Economics and Statistics).
- Consider the following two-period fixed effects model with a single regressor, x_{it}

$$y_{it} = \lambda_i + \alpha x_{it} + u_{it}, \ (i = 1, \ldots, n; \ t = 1, 2),$$

where

$$x_{it} = z_i + a_{it} \text{ and } u_{it} = v_i + b_{it}$$

and z_i and v_i are random variables.

Spurious Panel Regression

- As usual, λ_i is an individual effects variable correlated with x_{it}. Observed data are $\{y_{it}\}$ and $\{x_{it}\}$. But $\{z_i\}$ and $\{a_{it}\}$ are not separately observed. Random variables $\{a_{it}\}$ and $\{b_{it}\}$ bring time series variations to the observed data.

- Assume
  (i) $\begin{pmatrix} n^{\beta} a_{it} \\ n^{\gamma} b_{it} \end{pmatrix} \sim iid \left(0, \begin{bmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_b^2 \end{bmatrix}\right)$ for every n, i and t;
  (ii) $\beta > 0,\ \gamma < \beta - \frac{1}{2}$.

Spurious Panel Regression

- The first-differenced estimator is written as

$$\hat{\alpha}_d = \frac{\sum_{i=1}^{n} \Delta x_i \Delta y_i}{\sum_{i=1}^{n} (\Delta x_i)^2} = \alpha + \frac{\sum_{i=1}^{n} \Delta a_i \Delta b_i}{\sum_{i=1}^{n} (\Delta a_i)^2}.$$

Spurious Panel Regression

- Since $n^{\beta}\Delta a_i \sim iid\ (0, 2\sigma_a^2)$, $n^{\gamma}\Delta b_i \sim iid\ (0, 2\sigma_b^2)$ and $Cov(n^{\beta}\Delta a_i, n^{\gamma}\Delta b_i) = 0$ for every n, we obtain by the central limit theorem
$$\frac{n^{\beta+\gamma}}{\sqrt{n}} \sum_{i=1}^{n} \Delta a_i \Delta b_i \xrightarrow{d} N(0, \phi)$$
with $\phi = E\left[n^{2(\beta+\gamma)} (\Delta a_i)^2 (\Delta b_i)^2\right] = 4\sigma_a^2\sigma_b^2$.

Spurious Panel Regression

- The law of large numbers yields

$$\frac{n^{2\beta}}{n}\sum_{i=1}^{n}(\Delta a_i)^2 \xrightarrow{p} 2\sigma_a^2.$$

- Thus, $n^{\frac{1}{2}-\beta+\gamma}(\hat{\alpha}_d-\alpha) \xrightarrow{d} N\left(0, \frac{\sigma_b^2}{\sigma_a^2}\right)$. This shows that $\hat{\alpha}_d$ diverges in probability to $\pm\infty$ when $\frac{1}{2}-\beta+\gamma<0$.

Spurious Panel Regression

- Under the alternative hypothesis $H_1 : \alpha \neq 0$,

$$t_\alpha = \frac{\hat{\alpha}_d}{\sqrt{\hat{\sigma}^2 \left(\sum_{i=1}^{n} \left(\Delta x_i \right)^2 \right)^{-1}}} \xrightarrow{d} N(0,1).$$

  Thus, the t-test is not consistent.
- See Choi (2013) for further results.