November 29, 2022

Abstract

Exercise 1: Spurious Regression

1. Set-up and Definitions

- time series are generated as standardized random walk processes:

$$
\begin{aligned}
&x_t = \phi x_{t-1} + u_{x,t}, x_0 = 0, u_{x,t} \overset{iid}{\sim} \mathcal{N}(0,1) \\
&y_t = \phi y_{t-1} + u_{y,t}, y_0 = 0, u_{y,t} \overset{iid}{\sim} \mathcal{N}(0,1) \\
&\phi \in \{0.8, 1\}
\end{aligned}
$$

  A DGP with $\phi = 0.8$ is referred to as **AR(1)** and a DGP with $\phi = 1$ as **random walk** in the text.

- two regression models are defined:

$$
\begin{aligned}
y_t &= \beta_1 + \beta_2 x_t + v_t, && \text{(spurious regression)} \\
y_t &= \beta_1 + \beta_2 x_t + \beta_3 y_{t-1} + v_t, && \text{(valid regression)}
\end{aligned}
$$

  The **spurious regression** model should be expected to produce an insignificant estimate of β_2 and an R^2 near 0. The **valid regression** model reduces to the DGP for y_t given $\beta_1 = \beta_2 = 0$ and $\beta_3 = 1$.

- The combination of above DGPs and regression models yield 4 permutations for which the authors report how the rejection frequencies of $H0 : \beta_2 = 0$ change with increasing sample size n.

$$
\begin{aligned}
&LRM_1 : y_t = \beta_1 + \beta_2 x_t + v_t, && x_t, y_t : I(1), \phi = 1 \\
&LRM_2 : y_t = \beta_1 + \beta_2 x_t + v_t, && x_t, y_t : AR(1), \phi = 0.8 \\
&LRM_3 : y_t = \beta_1 + \beta_2 x_t + \beta_3 y_{t-1} + v_t, && x_t, y_t : I(1), \phi = 1 \\
&LRM_4 : y_t = \beta_1 + \beta_2 x_t + \beta_3 y_{t-1} + v_t, && x_t, y_t : AR(1), \phi = 0.8
\end{aligned}
$$

- In each simulation the DGP produce timeseries of length T = 6, 12, 60, 120, 240, 360, 480.

- A total of $N_{sim} = 100,000$ is run to produce the distributions of t-stat and R^2.

- Each regression model is estimated with OLS under asymptotic theory[1]

  A.1: independent observations y_i, x_i are i.i.d, which is equivalent to $v_t|x$ and $v_t|y$ being i.i.d.

  A.2: regressors are uncorrelated with errors $E[v_t x_t] = E[v_t y_t] = 0$. This assumption is weaker than strict exogeneity and is not restricting the utilization of lagged values of the dependent variable into the regressors matrix.

  A.3 Identification $Q = E[x_t x_t']$ exists, is positive definite and has rank corresponding to the total number of regressors (no regressor is a linear combination of the others).

  A.4: Finite Moments $S = E[v_t^2 x_t x_t']$ exists and is positive definite, with

$$\begin{aligned} S &= E[v_t^2 x_t x_t'] \\ &= E[v_t^2]E[x_t x_t'] + cov(v_t, x_t) \\ &= E[v_t^2]E[Q], by A.2 \end{aligned}$$

[1] ref: class notes, Ch 2, p. 60

1.a Replication of Figure 14.1 in Davidson MacKinnon (2005, book)

(i) Compute for each sample size T the distribution of the R^2 of the MC simulations with either 7 separate histograms, or one unique figure where you report on the y-axis the 5%, 10% 25%, 50%, 75%, 90% and 95% quantiles of the distributions of the simulated R^2, and on the x-axis you have T = 6, 12, 60, 120, 240, 360, 480.

LRM_1 spurious regression of random walks with $\phi = 1$ should produce R^2 near 0 given that β_2 is expected to be insignificant. The distribution of simulated R^2 however indicates that this is not the case for any length of timeseries.

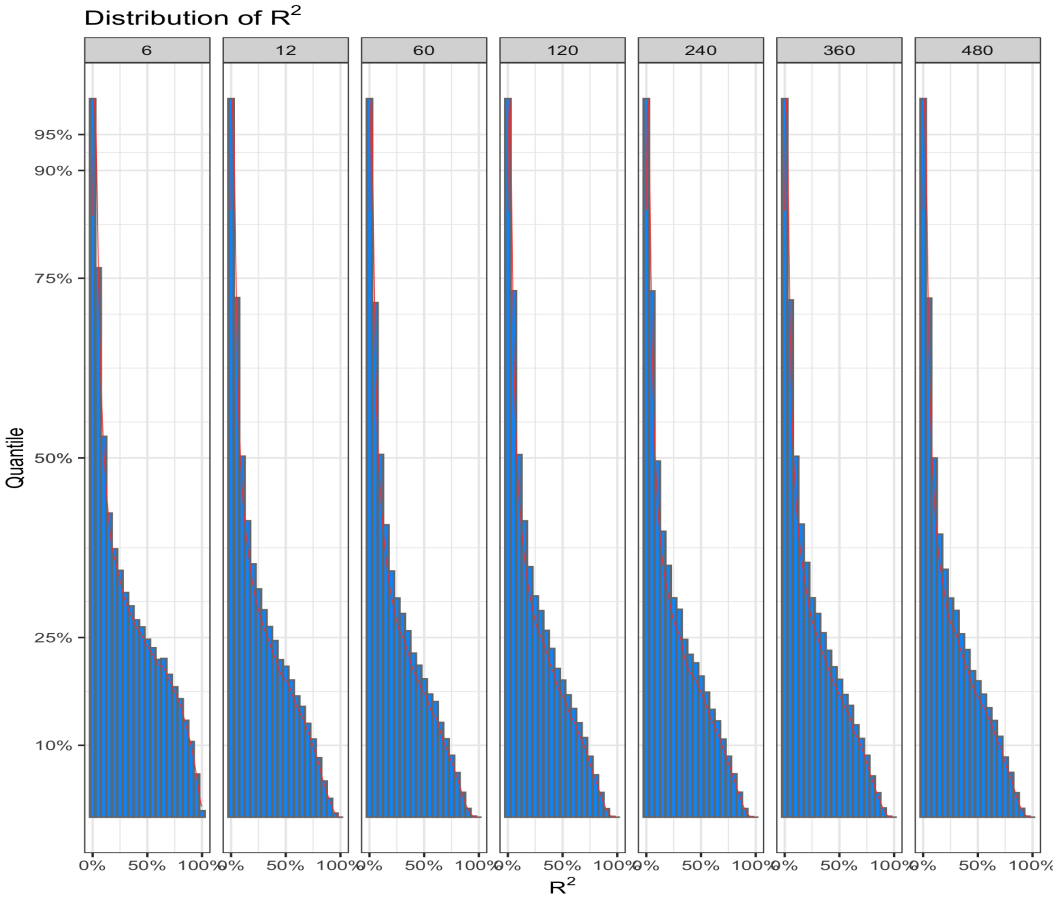

Figure 1: Distribution of R^2

(ii) Similarly (either with histograms, or with one plot of the quantiles) report the distributions of the estimates t-statistics for the test of the null H0 : $\beta_2 = 0$

The dispersion of the t-statistic $t_{\beta_2} = \frac{\beta_2 - 0}{\sigma_{\beta_2}}$ around 0 increases with increasing length of the timeseries.

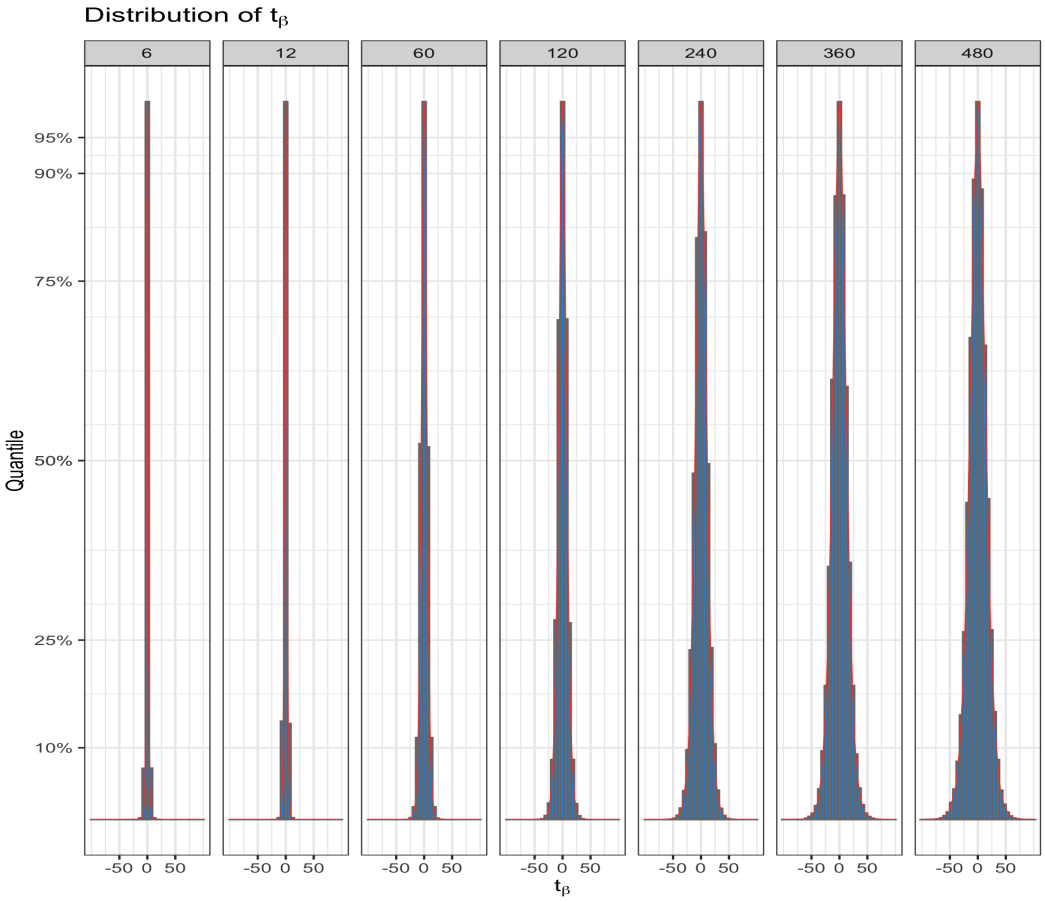

Figure 2: Distribution of t_{β_2}

(iii) Compute the empirical rejection frequencies (that is the empirical size of the tests), which is exactly the figure 14.1 in Davidson MacKinnon (2005, book).

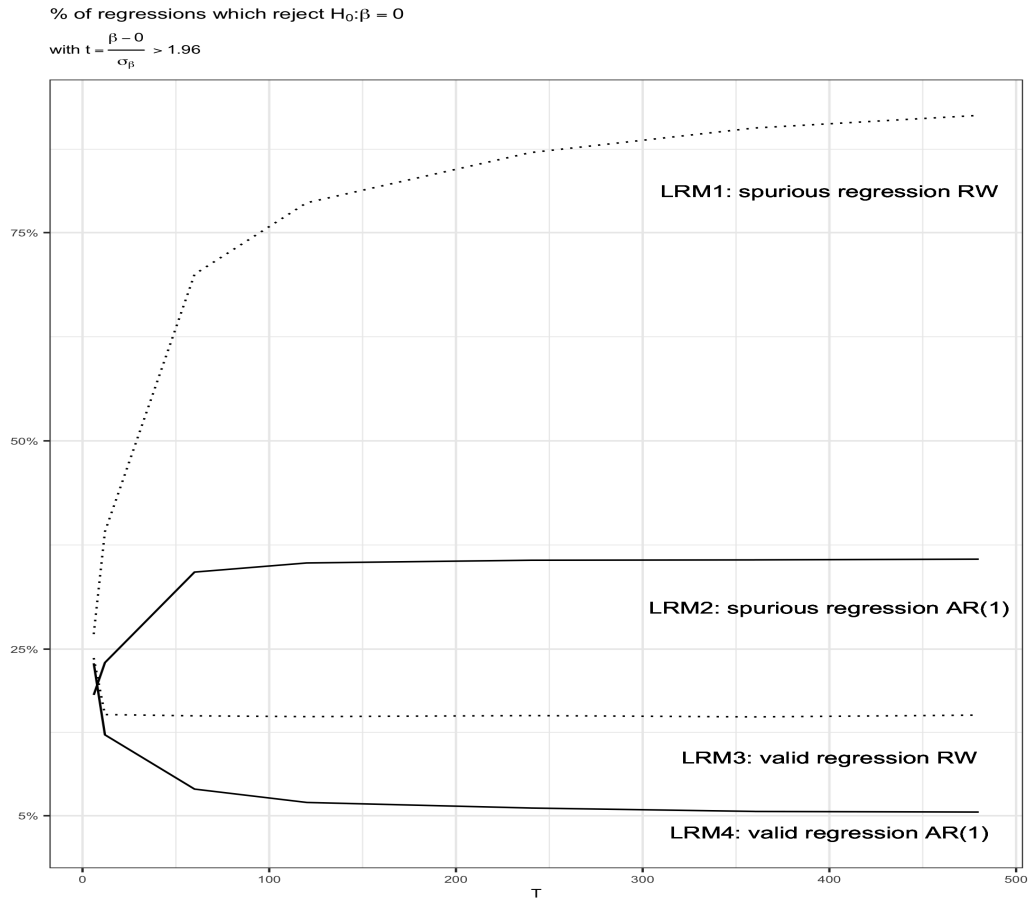

Figure 3: Rejection rate of $H0 : \beta_2 = 0$

1b: Summarize the problems of spurious regressions in econometrics

The results obtained are different from what we previously expected. Indeed, by looking at Figure 3, we can observe that:

- The simulated rate of rejection in LRM_4 converges asymptotically to 5% and significantly higher for the three other LRM_{1-3}.
- Additionally, we notice that the proportion of rejections keeps increasing in T for LRM_1 and LRM_2 implying a statistically significant relationship (β_2) between y_t and x_t even if this should not exist.
- Therefore, it emerges that the actual probability of a mistaken rejection of the true H0, i.e. committing a Type 1 error[2] is significantly higher than the assumed test size.

[tp 1] T?

Issues with these regressions:

- The bias of β_2 does not converge asymptotically in probability.
  - For $\hat{\beta}$ to converge to β_0 asymptotically, the bias $(\hat{\beta} - \beta_0)$ must be $O_p(1)$:

$$
\begin{aligned}
(\hat{\beta} - \beta_0) &= (X'X)^{-1}X'u, && \text{with} \\
(X'X)^{-1} &\in O_p(n^{-1}) && \text{and} \\
X'u &\in O_p(n^{\cdot 5}) && \text{consequently:} \\
n^{\cdot 5}(\hat{\beta} - \beta_0) &= n^{\cdot 5}(X'X)^{-1}X'u = n^{\cdot 5}O_p(n^{-1})O_p(n^{\cdot 5}) = O_p(1)
\end{aligned}
$$

  - The relevant assumption to be tested is therefore is $(X'X) \in O_p(n)$.
  - The random walks found in LRM_1 and LRM_3 are I(1), due to:

$$
\begin{aligned}
w_t &= w_{t-1} + \epsilon_t \\
w_t - w_{t-1} &= \epsilon_t \\
(1 - L)w_t &= \epsilon_t \\
(1 - \phi(z))w_t &= \epsilon_t \\
\phi(z) &= 1
\end{aligned}
$$

    Consequently, both x_t and y_t are I(1) given $\phi = 1$.
  - Further, LRM_1 and LRM_3 reduce recursively to $w_t = \sum_{s=1}^{t} \epsilon_s$,

[2] $P(R_{H_0}|H_0)$.

which enters as $X'X$ or:

$$\sum_{t=1}^{n}\left(\sum_{r=1}^{t}\sum_{s=1}^{t}\right)\epsilon_r\epsilon_s = \sum_{t=1}^{n}\sum_{r=1}^{t}E(\epsilon_r^2), \epsilon_r\epsilon_s = 0\ \forall r \neq s$$
$$\sum_{t=1}^{n}\sum_{r=1}^{t}\sigma^2 = \sum_{t=1}^{n}\sum_{r=1}^{t}1, \text{by assumption}$$
$$\sum_{t=1}^{n}t = \frac{1}{2}n(n+1)$$

  - Consequently, being I(1), $X'X \in O_p(n^2)$ and therefore cannot possibly converge to a finite probability limit. The bias $(\hat{\beta} - \beta_0)$ therefore does not converge asymptotically in probability.

- $X'X$ is no longer a positive definite matrix ... ...
- The distribution of the t-statistics does not converge to the Student's t even asymptotically causing an over-rejection of the null.
- Although the proportion of rejections converges and $E[X'X]$ is a finite positive definite matrix, we can see that the result for LRM_2 implies an over-rejection of the null while the empirical frequency for LRM_4 converges to the 5% significance level only after increasing significantly the number of simulations.

[tp 2] Need help with this statement

[tp 3] 0?

- The $H_0 : \beta_2 = 0$ tested with LRM_1

$$y_t = \beta_1 + \beta_2 y_{t-1} + v_t$$

implies a DGP': $y_t = \beta_1 + v_t$, when y_t is actually generated using the DGP $y_t = y_{t-1} + v_t, y_0 = 0, v_t \sim iidN(0,1)$. The wrongly specified H_0 is rejected with increasing frequency in n. This merely confirms that y_t is not generated by the model implied DGP'. Correctly specifying the model as LRM_2

$$y_t = \beta_1 + \beta_2 x_t + \beta_3 y_{t-1} + v_t$$

and testing $H_0 : \beta_2 = 0$, implying $\beta_3 = 1$ reduces the model to the actual DGP. This treatment, however, does not completely eliminate the problem i.e. leaves the rejection rate still significantly above 0.

- The distortion for LRM_3 arises from the fact that neither the constant nor x_t have any explanatory power for y_t, therefore $y_t = v_t = 0.8y_{t-1} + u_{y,t}$. As y_t is modelled as an AR(1), the error term of LRM_3 becomes:

$$v_t = 0.8y_{t-1} + u_{y,t}$$
$$v_t = 0.8(u_{y,t-1} + 0.8y_{t-2}) + ... + u_{y,t}\ t = 1, ..., T$$

This intuition suggests that there is serial autocorrelation in the errors and the (asymptotic) solution to avoid too small standard errors in finite samples is to switch from the OLS VaR-Cov estimator to the Newey-West heteroskedasticity and autocorrelation consistent (HAC) one.

Original text from Giovanni [tp]

Usually, the bias of the β multipled by $n^{0.5}$ is $O_p(1)$, as per below demonstration and noting that $K'K = O_p(T)$ and $K'v = O_p(T^{0.5})$:

$$\begin{aligned}\hat{\beta} - \beta &= (K'K)^{-1}K'v \\ T^{0.5}(\hat{\beta} - \beta) &= T^{0.5}((K'K)^{-1}K'v) \\ &= T^{0.5}O_p(T^{-1}O_p(T^{0.5}) \\ &= O_p(1)\end{aligned}$$

The bias is eventually bounded for $T \to +\infty$, so it does not explode asymptotically.

However, we have here an issue of unit roots and spurious regressions. If, at least, one of the regressors is a unit root[3], we can derive that $E[K_t'K_t]$ is not a finite positive definite matrix anymore and there is a violation of the OLS assumptions under asymptotic theory (steps below are for DPG 1[4], same holds in case of multiple regressors):

$$\begin{aligned}E[x_T'x_T] &= \sum_{t=1}^{T} x_t^2 \\ &= T + (T-1) + ... + 2 + 1 \\ &= \frac{T(T+1)}{2}\end{aligned}$$

[tp 4] Both LRM_3 and LRM_1 have a unit root. But LRM_1 is also wrongly specified, ie does not reduce to the DGP

It follows that: i) $T^{-1}x_T'x_T$ is $O(T)$ and ii) this metric does not have a finite probability limit ($T \to +\infty$).

The distribution of the t-statistics does not converge to the Student's t even asymptotically causing an overrejection of the null.

Eventually, we observe that the issue of spurious regressions occurs even if all variables are stationary[5]. Although the proportion of rejections converges to a fixed number and $E[K_t'K_t]$ is a finite definite positive metric, we can see that the result for DPG 2 implies an overrejection of the null while the empirical frequency for DPG 4, the only correct regression model, converges to the 5% significance level only after increasing significantly the sample size (in the order of a few thousand observations).

The distortion for DPG 3 arises from the fact that neither the constant nor x_t

[3] This happens for LRM_1 and LRM_3, although for the latter the proportion of rejections does not converge to 1 as T increases. This is because we are reducing y_t by one lag turning the dependent variable into a white noise random walk

[4] $V(x_t) = E(x_t^2) = t$.

[5] AR(1) in this case.

have any explanatory power for y_t, therefore $y_t = v_t = 0.8y_{t-1} + u_{y,t}$.
As y_t is modelled as an AR(1), the error term of DPG 3 becomes:

$$v_t = 0.8y_{t-1} + u_{y,t}$$
$$v_t = 0.8(u_{y,t-1} + 0.8y_{t-2}) + +u_{y,t}\ t = 1, ..., T$$

This intuition suggests that there is serial autocorrelation in the errors and the (asymptotic) solution to avoid too small standard errors in finite samples is to switch from the OLS VaR-Cov estimator to the Newey-West heteroskedasticity and autocorrelation consistent (HAC) one.