Classwork 14: Random Walks

EC 421

2022-08-08

1) *(Analytic)* Let y_t be a random walk so that $y_t = y_{t-1} + u_t$, where $u_t \sim$ iid N(0, σ^2).

Suppose $y_1 = u_1$. Show that $y_t = \sum_t u_t$ so that $y_1 = u_1$, $y_2 = u_1 + u_2$, and $y_3 = u_1 + u_2 + u_3$. A rigorous mathematical proof is not necessary; just show it is true for those first few cases.

2) *(Analytic)* Use the previous result to show why random walks are nonstationary.

3) *(Analytic)* Suppose x_t and y_t are two unrelated random walks, so that

$x_t = x_{t-1} + w_t$

$y_t = y_{t-1} + v_t$

If we estimate $y_t = \beta_0 + \beta_1 x_t + u_t$, why do we often conclude that x has a strong association with y?

4) *(R)* Granger-Newbold (1974) Replication

Generate two time series x_t and y_t using `accumulate()` that are unrelated except that they are both random walks:

$y_t = y_{t-1} + w_t$

$x_t = x_{t-1} + v_t$

Where w_t and v_t are both iid N(0, 1) and t goes from 1 to 50.

Run the regression `y ~ x`. Can you reject the null hypothesis that $\beta_1 = 0$? That is, does x seem to effect y (p value less than .05)?

Then make this into a simulation using `map()`, where you'll generate 100 datasets, run `y ~ x` with each, and count the total number of times $\beta_1 = 0$ is rejected.

5) *(R)* Take the simulation above and adjust it to observe what happens when

we increase the number of observations from 50 to 100 to 300. Does a large sample size help to find $\beta_1 = 0$ more often?

6) *(R)* We can first difference a random walk to get a stationary time series

of u_t. Take your simulation above to estimate a new model: $y_t - y_{t-1} = \beta_0 + \beta_1(x_t - x_{t-1}) + u_t$. Does taking first differences help to find $\hat{\beta}_1 = 0$ more often?

7) *(R)* Use `accumulate` to generate an autocorrelated time series that follows this process:

$y_t = 0.9y_{t-1} + u_t$

where u_t is iid N(0, 1).

8) *(R)* Extra Credit (2 points): Continuing from the previous problem, we

learned that $y_t = \beta_0 + \beta_1 y_{t-1} + u_t$ generates biased but consistent coefficient estimates when there is no autocorrelation in u_t. Simulate this result with **one** (`geom_density`) plot that shows that as the sample size increases, running `lm(y ~ lag(y), data = .)` is biased but consistent. Use the same data generating process as the last problem:

$y_t = 0.9y_{t-1} + u_t$

where u_t is iid N(0, 1).