**Home Work 3**

**3.** In this exercise you will conduct a Monte Carlo experiment to study the phenomenon of spurious regression discussed in Class. In a Monte Carlo study, artificial data are generated using a computer, and then those artificial data are used to calculate the statistics being studied. This makes it possible to compute the distribution of statistics for known models when mathematical expressions for those distributions are complicated (as they are here) or even unknown. In this exercise, you will generate data so that two series, *Yt* and *Xt*, are independently distributed random walks. The specific steps are as follows:

i. Use your computer to generate a sequence of *T* = 100 i.i.d. standard normal random variables. Call these variables *e*1, *e*2, c, *e*100. Set *Y*1 = *e*1 and *Yt* = *Yt* - 1 + *et* for *t* = 2, 3, ┄, 100.

ii. Use your computer to generate a new sequence, *a*1, *a*2, ┄, *a*100, of *T* = 100 i.i.d. standard normal random variables. Set *X*1 = *a*1 and *Xt* = *Xt* - 1 + *at* for *t* = 2, 3, ┄, 100.

iii. Regress *Yt* onto a constant and *Xt*. Compute the OLS estimator, the regression *R*2, and the (homoskedastic-only) *t*-statistic testing the null hypothesis that b1 (the coefficient on *Xt*) is zero

Use this algorithm to answer the following questions:

1. Run the algorithm (i) through (iii) once. Use the *t*-statistic from (iii) to test the null hypothesis that b1 = 0, using the usual 5% critical value of 1.96. What is the *R* 2 of your regression?
2. Repeat (a) 1000 times, saving each value of *R* 2 and the *t*-statistic. Construct a histogram of the *R* 2 and *t*-statistic. What are the 5%, 50%, and 95% percentiles of the distributions of the *R* 2 and the *t*-statistic? In what fraction of your 1000 simulated data sets does the *t*-statistic exceed 1.96 in absolute value?
3. Repeat (b) for different numbers of observations, such as *T* = 50 and *T* = 200. As the sample size increases, does the fraction of times that you reject the null hypothesis approach 5%, as it should because you have generated *Y* and *X* to be independently distributed? Does this fraction seem to approach some other limit as *T* gets large? What is that limit?