Problem Set 3 Solutions

December 12, 2002

1 Problem 1

(a) Given that we do not know the aggregate variance-covariance matrix of the estimated coefficients, instead we will have to use a standard Normal test. It makes sense to use it for $\hat{\rho}$, and to forget about all the other samples. Ideally, we would like to use a two-dimensional Wald test for both ρ and δ, but that is unfeasible. Thus, the statistic for ρ is given by the formula,

$$\frac{\hat{\rho}-1}{std.err.\hat{\rho}} = \frac{0.9957-1}{0.01142} = -0.3765,$$

so the test fails to reject the Null as the statistic is less in absolute value than 1.96 which is the test value at the 5 percent two-sided level.

(b) The only difference between the two models is the presence of the Δy_{t-1} factor in the regression equation. Since we are doing an OLS estimation procedure, if the model with the added factor holds true and if we do not add that factor, then we will get a missing variable bias and our tests will have very low power. However, if the model holds true without the term Δy_{t-1} and if we add it, then we will get bias again, as under the Null: $y_t = \mu + y_{t-1} + \varepsilon_t$, the term y_{t-1} is potentially correlated with the term Δy_{t-1}. If we had a set of instruments to estimate that model by a method other than OLS, then we could add that factor in the regression equation and not get the above bias. However, the validity of the potential instruments would still remain under question, under the conditions that we had any such instruments. In general, whether or not to add that extra factor in there remains the option of the researcher doing the empirical work.

(c) This is a standard exercise in unit roots. To start with, note that,

$$n(\hat{\rho}-1) = n(\frac{\sum_t y_t y_{t-1}}{y_{t-1}^2} - 1) = n\frac{\sum_t y_{t-1}(\mu+\varepsilon_t)}{y_{t-1}^2} \sim^{D}_{n\to\infty} \frac{\int_0^1 BdB}{\int_0^1 B^2 dr},$$

where $B(r) = r\mu + W(r)$, where $W(r)$ is the standard Brownian motion with unit variance (the unit variance occurs because the ε's are having unit variance.) This completes the proof of this problem.

2 Problem 2

(a) To start with, note that the series looks non-stationary because of its auto-correlation function, and so because of that first differencing is required in order to make the series stationary.

Next, we test for whether the differenced series is white noise. To do so, apply the usual Box-Pierce Test for lag $K = 12$. The test is given by the formula,

$$Q = T \sum_{j=1}^{K} \hat{\rho}^2(j),$$

where $T = 210$, and then Q is distributed asymptotically as χ^2_K. The test statistic is 13.2, which is considerably larger than the χ^2 value at the 5 percent level of 5.2. Therefore we reject the Null. Given the limited information, it is hard to postulate another model. A choice would be an $ARIMA(2, 1, 0)$ model, but given our data any $I(1)$ model which is not white noise would be accepted given the limited available information.

(b) A test of the described Null would simply be a t-test for whether the coefficient of y_{t-1} in the appropriate regression setting would be equal to 1.

The advantage of adding the extra term Δy_{t-1} is that it allows for more general disturbance processes in the error term. One potential disadvantage is that in a general regression setting the above formula might lead to biases in the estimation of the other coefficients, and thus to inconsistent tests. However, the second disadvantage does not appear to play a role in our case, as the estimates are very little affected by the addition of the delta term in the regression setting.

The test by using the second regression has a value of 1.54, which is less than 1.96. Our test fails to reject the Null, and thus we accept the Null.

(c) As $y_t = y_{t-1} + \varepsilon_t$, we obtain that,

$$y_t = y_0 + \sum_{s=1}^{t} \varepsilon_s.$$

Simple algebra reveals to us that,(OLS First Order Conditions)

$$\hat{\rho} = \frac{(\sum_i y_i)(\sum_i y_{i-1}) - (n-1)(\sum_i y_i y_{i-1})}{(\sum_i y_{i-1})^2 - (n-1)(\sum_i y_{i-1}^2)}.$$

Next, standard unit root theory tells us that,

$$n(\hat{\rho} - 1) \sim^a_{n\to\infty} \frac{B(1)\int_0^1 B(r)dr - \frac{1}{2}(B(1)^2 - 1)}{\int_0^1 B^2 dr - (\int_0^1 B(r)dr)^2},$$

Note that adding the constant term in the regressioon equation changes the asymptotic distribution of the unit root estimator.

(d) Exactly the same algebra as in part (c), except that now we also have the Δy_{t-1} term in our formulas to account for.

3 Problem 3

(a) To begin with, it is clear that x_t and y_t are non-stationary, as they are standard random walks.
(b)Define the two processes $X_n(r)$ and $Y_n(r)$, defined as $\frac{1}{\sqrt{n}}x_{[nr]}$ and $\frac{1}{\sqrt{n}}y_{[nr]}$, where n is the sample size. Then it will be case that,

$$\frac{1}{n^2}\sum_t x_t y_t = \sum_{t=1}^n \int_{\frac{t-1}{n}}^{\frac{t}{n}} X_n(r)dr.$$

By doing a similar calculation for the other series, namely $\frac{1}{n^{\frac{3}{2}}}\sum_t x_t^2$, and dividing both sides in the expression for $\hat{\beta}$ by $\frac{1}{n^2}$, we obtain $\hat{\beta}$ as the ratio of two integrals. Next, as the processes $X_n(r)$ and $Y_n(r)$ converge to standard Brownian motions, called W_2 and W_1, we obtain that,

$$\hat{\beta} \to^a_{n\to\infty} \frac{\int_0^1 W_1(r)W_2(r)dr}{\int_0^1 W_2(r)^2 dr}.$$

This result implies that when regressing one random walk on another independent random walk, in the limit this is the same as regression one Brownian Motion on another Brownian motion. Namely, we have shown that the regression coefficient is a continuous functional!
(c) The solution starts out by noting that,

$$\frac{1}{n^2}\sum_t (x_t - \bar{x})^2 = \frac{1}{n^2}(\sum_t x_t^2 - n(\bar{x})^2).$$

This sum is then approximated by,

$$\int_0^1 W_2^2(r)dr - (\int_0^1 W_2(r)dr)^2.$$

A similar calculation for the other sequence, reveal to us that the distribution for R^2 is asymptotically equal to,

$$(\frac{\int_0^1 W_1(r)W_2(r)dr}{\int_0^1 W_2(r)^2 dr})^2 \times \frac{\int_0^1 W_2^2(r)dr - (\int_0^1 W_2(r)dr)^2}{\int_0^1 W_1^2(r)dr - (\int_0^1 W_1(r)dr)^2}.$$

(d) The standard F-test will have the wrong size, as R^2 does not get the right distribution. The F-test does not do a good job here, because of the unit root problem, namely the standard asymptotically. In addition to that, the test also appears to be inconsistent.
(e) Note that as $y_t = x_t + v_t$, then it will be the case that,

$$\frac{1}{n^2}\sum_t x_t y_t = \frac{1}{n^2}\sum_t x_t(x_t + v_t) = \frac{1}{n^2}(\sum_t x_t^2 + \sum_t x_t v_t).$$

Then it will be the case that,

$$n(\hat{\beta} - 1) = \frac{\frac{1}{n}\sum_t x_t v_t}{\frac{1}{n^2}\sum_t x_t^2} \sim^a_{\to\infty} \frac{\int_0^1 W_2 dW_1}{\int_0^1 W_2^2 dr},$$

where W_1 and W_2 are two independent standard Brownian Motions.
(f) The test statistis is given by $s := n(\hat{\beta} - 1)$. Next pick a confidence level α. After that simulate the value $\frac{\int_0^1 W_2 dW_1}{\int_0^1 W_2^2 dr}$, and after estimate the value c so that,

$$Prob(s \in I(c)) = \alpha,$$

where the set $I(c)$ is defined by the formula,

$$x \in I(c) \text{ iff } \hat{f}(x) \geq c.$$

The interval I is estimated by way of simulation. If we can find tables for this distribution, then we can use those instead, and then there will be no need to use this simulation procedure.

This completes the solutions.