Problem set 5: An introduction to time series

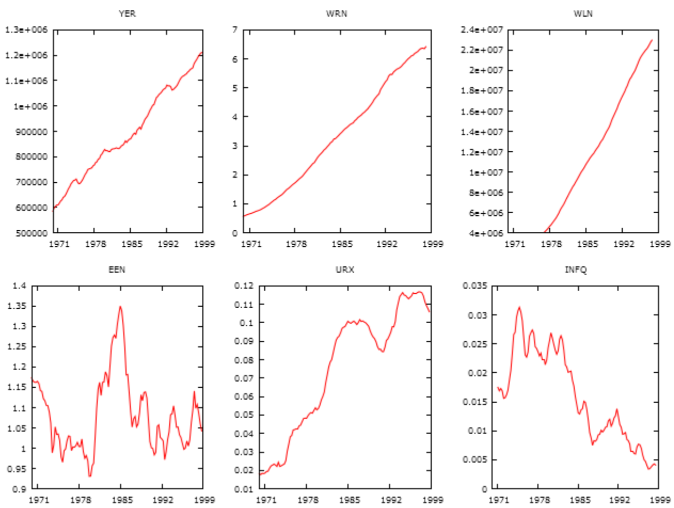
September 20, 2013

1 Introduction

This problem set accompanies the Youtube lecture series, and roughly corresponds to videos 156-180, covering an introduction to time series.

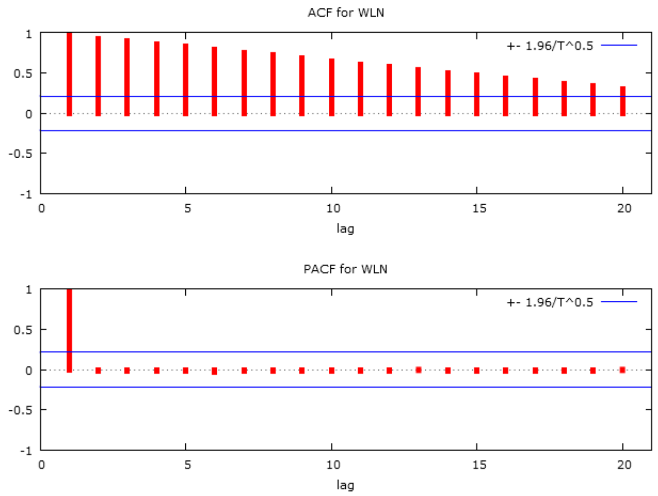
2 Practical - Eurozone & World economic data

1. (a) The time series scatters of the various series are shown below.



   (b) Looking at each of the time series, they all appear non-stationary. The real wage rate and wealth could potentially be trend-stationary on first appearances, although most likely not.

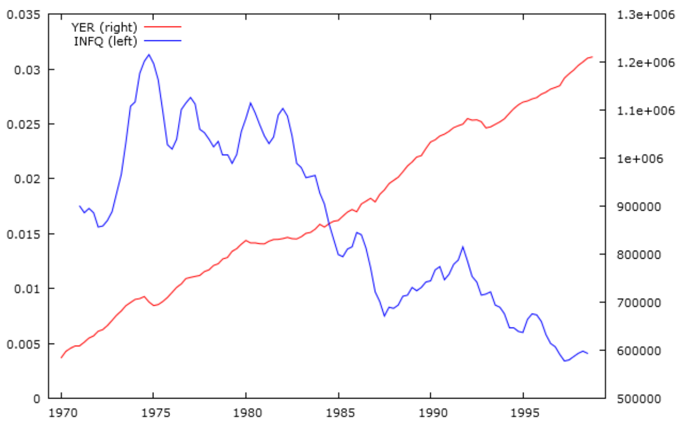
   (c) All the correlograms appear to show evidence of AR(1) serial correlation. This is evidenced by the strong (but decaying) autocorrelation as the number of lags increases. Also the partial autocorrelation peaked at the first lag suggests that it is a AR(1) process, not a higher order AR process. Possible exceptions are *EEN* and *INFQ* which show significant *negative* partial autocorrelations with the second lag. This could indicate that they could be modelled perhaps as having an MA(2) component, but this is quite a tentative conclusion.



(d) *GDP*, *Wealth*, *Wage* I would suggest would be the only series which look like they could be trend-stationary. The others look more random-walk-ish.

(e) Conducting a test for unit roots for each of the series we find that in all cases (including a time trend for those series that are clearly trending), that the p value against the appropriate Dickey-Fuller distribution is above the 5% critical value. Hence we cannot reject the null hypothesis of each series having a unit root. Hence the series themselves are highly persistent, and non-stationary. I used four lags of the differences in the ADF tests, since we are dealing with quarterly data.

(f) There cannot be a long-run (cointegrated) relationship between these two variables. To see this look at the plot below with both series superimposed on top of one another. There is no single number which can multiply inflation to make the difference between them stationary.



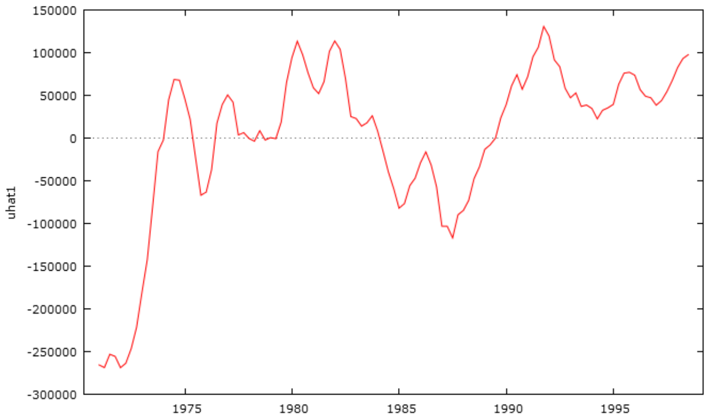
Another way to see this is to run a (spurious) regression of GDP on inflation. The results of this are shown below.

Model X: OLS, using observations 1971:1–1998:3 (T = 111)
Dependent variable: YER

	Coefficient	Std. Error	t-ratio	p-value
const	1.18241e+006	20320.5	58.1879	0.0000
INFQ	−1.75610e+007	1.13925e+006	−15.4145	0.0000

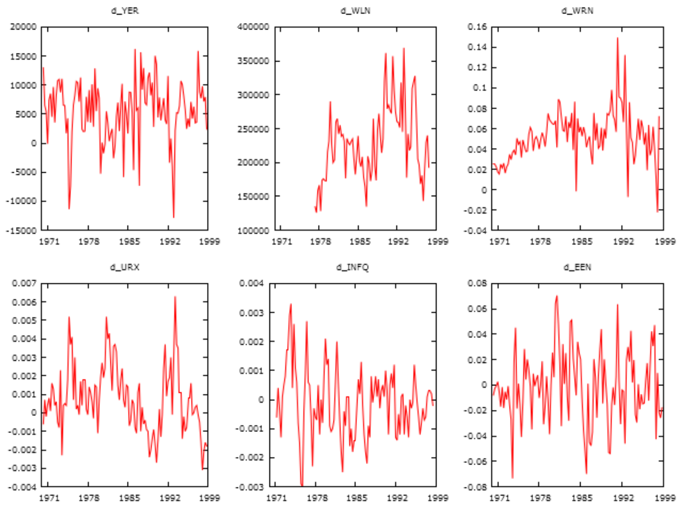
Mean dependent var	901020.4	S.D. dependent var	166945.7
Sum squared resid	9.64e+11	S.E. of regression	94048.96
R^2	0.685521	Adjusted R^2	0.682636
$F(1, 109)$	237.6054	P-value(F)	3.81e–29
Log-likelihood	−1427.617	Akaike criterion	2859.235
Schwarz criterion	2864.654	Hannan–Quinn	2861.433
$\hat{\rho}$	0.940485	Durbin–Watson	0.054640

Note that we have an R-squared which is far in excess of the value of the Durbin-Watson statistic - a classical sign of a spurious regression! Further, examining a time series plot of the residuals from this regression it is apparent that these are certainly non-stationary - another sign that they are not cointegrated.

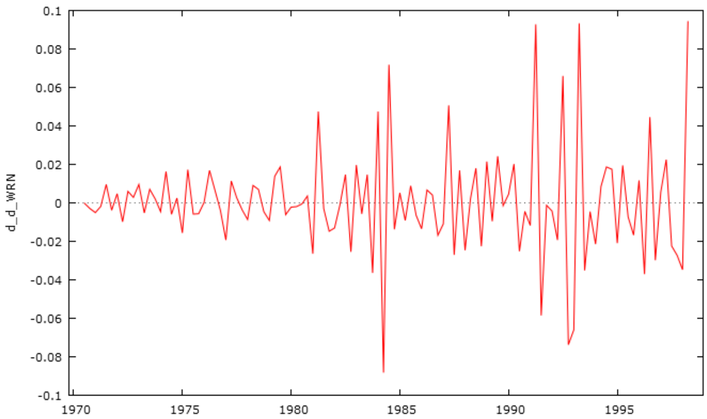


(g) See next part.

(h) In order to test the order of integration of each series, we first take the first differences of each of the series, and graph these. The graphs of the first differenced series are shown below. From the plots shown, the series which look particularly non-stationary are that of the wage rate, (the variance of the process looks to increase over time as well as the series showing other signs of persistence), and the unemployment rate (the first difference still appears to be 'random-walking').



We then conduct ADF tests on these series, bearing in mind the visual plots of the series. We find that for all series except the wage rate, we reject the null of the series having a unit root. Hence all series other than the wage rate appear to be I(1), since although the levels are non-stationary, the first differences are stationary. A time series plot of the second difference of the wage rate is shown below. Accordingly, we reject the null hypothesis of a unit root, and conclude that the wage rate appears I(2). This makes sense since, in the youtube lectures I spoke about prices as often being I(2), and wages are essentially just prices. The intuition is that the first difference represents inflation, which itself is non-stationary.



(i) Testing for a relationship in differences between variables is *not* equivalent to testing for a long run relationship in levels. See this youtube video for a full explanation: `http://tinyurl.com/osqvp7b`

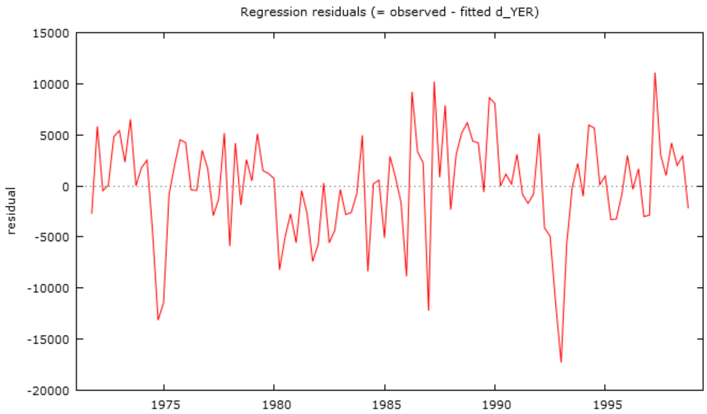
(j) The results of this regression are shown below.

Model 2: OLS, using observations 1971:4–1998:4 (T = 109)
Dependent variable: d_YER

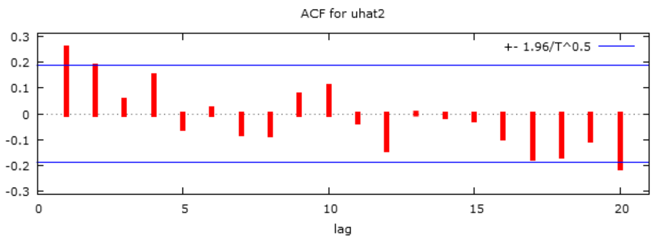
	Coefficient	Std. Error	t-ratio	p-value
const	5308.31	484.415	10.9582	0.0000
d_INFQ_1	1.61713e+006	497453.	3.2508	0.0015
d_INFQ_2	−2.05171e+006	497075.	−4.1276	0.0001

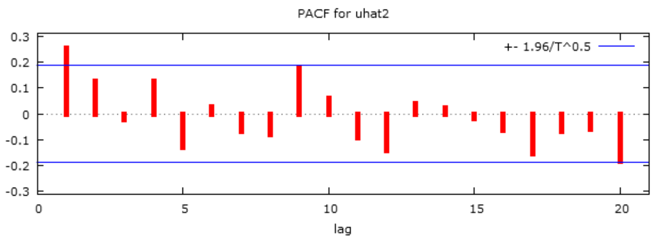
Mean dependent var	5366.872	S.D. dependent var	5390.701
Sum squared resid	2.68e+09	S.E. of regression	5023.645
R^2	0.147627	Adjusted R^2	0.131545
$F(2, 106)$	9.179361	P-value(F)	0.000211
Log-likelihood	−1082.032	Akaike criterion	2170.063
Schwarz criterion	2178.137	Hannan–Quinn	2173.337
$\hat{\rho}$	0.255408	Durbin–Watson	1.485667

We have an R-squared which is lower than the value of the Durbin-Watson statistic, suggestive that we may have avoided the issue of spurious regression. Also, since we are now using differences of the variables, these are stationary processes, meaning that we are less likely to run into the issue of spurious regression. A plot of the residuals from this regression are shown below, and appear to be stationary; again suggesting stationarity.



(k) Looking at a correlogram for the above residual series, it is possible that the series may be well modelled as an MA(1) process. This is because of the fact that there is significant autocorrelation with the first lag, but no other further lags.





(l) We can use either the Durbin-Watson test or an Breusch-Godfrey test. Both tests conclude that there is significant serial correlation present. The Breusch-Godfrey suggests that this is due to correlation between the residuals and their first lag; suggesting either an AR(1) or MA(1) process. Combining the evidence of both this test and the correlogram results, we conclude that the process could be well-modelled as an MA(1) process.

3 Theory

2. (a) Yes. We know this because of the fact that the residuals from this regression are non-stationary (they are trending upwards over time), and we can only get non-stationary residuals if we have at least one non-stationary variable in the regression.

   (b) No. Since the residuals are non-stationary we cannot conclude that the relationship is cointegrated.

   (c) Nothing! We have likely run a spurious regression.

   (d) Since the p value is below the 5% critical value, we reject the null hypothesis that the series has a unit root, and could perhaps suggest that we have found a cointegrated relationship.

   (e) We cannot infer anything using the standard errors from this regression unfortunately, this is because under the null hypothesis the relationship is spurious! We could use a leads and lags estimator however to clean up the issue of the violation of strict exogeneity. Assuming that we have no serial correlation or heteroscedasticity, and that the errors appear normal, we could then proceed on to do inference using the typical standard errors. Alternatively, we could use some sort of an Error Correction Model.

3. (a) (*GDP*, *weather*) - weather (if not too severe) could cause temporary (one period) effects on GDP. Civil war would likely cause structural changes which persist for longer.

   (b) The mean of this process is zero. To see this back-substitute in for X_t to obtain
   $$X_t = \rho(\rho X_{t-2} + \varepsilon_{t-1} + \theta\varepsilon_{t-2}) + \varepsilon_t + \theta\varepsilon_{t-1} = \rho^t X_0 + \sum_{i=0}^{t-1} \rho^i(\varepsilon_{t-i} + \theta\varepsilon_{t-i-1})$$
   From this, it is apparent when we apply the expectations operator we will obtain $\mathbb{E}(X_t) = 0$.

   (c) The variance of the process can be obtained in similar fashion to that of a typical AR(1) process, equalling $\frac{\sigma^2(1+\theta^2)}{1-\rho^2}$. See video: `http://tinyurl.com/onmwwul`

   (d) It is no different to that of the derivation of the conditions for a typical AR(1) process to be stationary. We just require that $|\rho| < 1$. Intuitively, it doesn't matter how big the effect of a shock is in the next period (that given by the θ term), so long as the effect propagates through time with an ever-decreasing magnitude this means that the time series will be stationary and not persistent.

   (e) Yes. The correlation between X_t and $X_{t-\tau}$ is given (for $\tau > 1$) by ρ^τ. Hence if $|\rho| < 1$ the series is weakly dependent since the correlation tends to zero (quickly enough technically, but don't worry about that) as τ tends to infinity.

   (f) For an AR(2) process to be stationary, we require that the lags polynomial have roots which lie outside the unit circle. Or put another way, are greater than 1. For an AR(2) process we can get the lags polynomial by first writing the series as

$$X_t - \rho_1 X_{t-1} - \rho_2 X_{t-2} = \varepsilon_t$$

Or using the lag operator

$$(1 - \rho_1 \mathbb{L} - \rho_2 \mathbb{L}^2) X_t = \varepsilon_t$$

The lags polynomial is the term in parentheses on X_t, and it must be the case that solutions to

$$1 - \rho_1 \mathbb{L} - \rho_2 \mathbb{L}^2 = 0$$

Are outside the unit circle. In other words we must have that $\mathbb{L} > 1$.

This must hold of an ARMA process of any order, all we need to do is work out the AR polynomial term, and then test whether the solutions are all greater than 1. Checking this method works for an AR(1) process. We must have solutions to

$$1 - \rho \mathbb{L} = 0$$

Which are greater than 1, which means that $|\mathbb{L}| = |\frac{1}{\rho}| > 1$, or alternatively, $|\rho| < 1$, which is what we have proved by other methods! This lags methodology works!