1 Regression with Time Series Variables

- With time series regression, Y might not only depend on X, but also lags of Y and lags of X

- *Autoregressive Distributed lag* (or ADL(p, q)) model has these features:

$$\begin{aligned} Y_t &= \alpha + \delta t + \rho_1 Y_{t-1} + .. + \rho_p Y_{t-p} \\ &\quad + \beta_0 X_t + \beta_1 X_{t-1} + .. + \beta_q X_{t-q} + \varepsilon_t. \end{aligned}$$

- Here we mostly focus on one X, but same ideas hold for case with several.

- Estimation and interpretation depends on whether, X and Y, are stationary or not.

- Note: X and Y must have the same stationarity properties (either must both be stationary or both have a unit root).

- Before running any time series regression, you should do unit root tests as described in Chapter 6 for every variable in your analysis.

- Note: in the textbook, the topics of forecasting, variance decompositions, vector autoregressions (VARs) and the Johansen test for cointegration covered. But we do not have time to cover them in this course (I hope you will read material anyway, but you are not responsible for this on the exam).

1.1 Time Series Regression When X and Y are Stationary

- When X and Y are stationary, standard OLS methods ADL(p, q) are fine

- E.g. hypothesis testing can be done using t-statistics or F-statistics. Sequential testing procedures can be used to select p and q, etc.

- Can rewrite ADL in more convenient form:

$$
\begin{aligned}
\Delta Y_t &= \alpha + \delta t + \phi Y_{t-1} + \gamma_1 \Delta Y_{t-1} + .. + \gamma_{p-1} \Delta Y_{t-p+1} \\
&\quad + \theta X_t + \omega_1 \Delta X_{t-1} + .. + \omega_{q-1} \Delta X_{t-q+1} + \varepsilon_t .
\end{aligned}
$$

- Note: this form of ADL model less likely to run into multicollinearity problems.

- One thing researchers often calculate is the *long run* or *total multiplier*

- To motivate: suppose that X and Y are in an equilibrium or steady state. Then X rises (permanently) by one unit, affecting Y, which starts to change, settling down in the long run to a new equilibrium value.

- Difference between old and new equilibrium values for Y is long run effect of X on Y and called long run multiplier.

- This multiplier is often of great interest for policy-makers who want to know the eventual effects of their policy changes in various areas.

- For ADL(p, q) model long run multiplier is:

$$-\frac{\theta}{\phi}.$$

1.2 Time Series Regression When Y and X Have Unit Roots

- Now assume Y and X have unit roots.

- In practice, you would do Dickey-Fuller test to confirm this.

1.2.1 Spurious Regression

- Consider the regression:

$$Y_t = \alpha + \beta X_t + \varepsilon_t.$$

- OLS estimation of this regression can yield results which are completely wrong.

- Even if the true value of β is 0, OLS can yield an estimate, $\widehat{\beta}$, which is very different from zero.

- Statistical tests (using the t-statistic or P-value) may indicate that β is not zero.

- Furthermore, if $\beta = 0$, then the R^2 should be zero. In fact, the R^2 will often be quite large.

- If Y and X have unit roots then *all* the usual regression results might be misleading and incorrect.

- This is called *spurious regression problem*.

- We will not prove, but stress the practical implication:

With the one exception of cointegration (see below), *you should never run a regression of Y on X if the variables have unit roots. Same thing holds for ADL.*

1.2.2 Cointegration

- If Y and X are cointegrated, do not need to worry about spurious regression problem.

- Cointegration has nice economic intuition.

- Intuition for cointegration: errors in the above regression model are:

$$\varepsilon_t = Y_t - \alpha - \beta X_t.$$

- Errors are just a linear combination of Y and X.

- Since X and Y both have unit roots you would expect the error to also have unit root.

- After all, if you add two things with a certain property together the result generally tends to have that property.

- Error does indeed usually have a unit root (this is what causes spurious regression problem).

- However, it is possible that the unit roots in Y and X "cancel each other out" and that the resulting error is stationary. This is *cointegration*,

- To summarize: if Y and X have unit roots, but some linear combination of them is stationary, then Y and X are cointegrated.

Intuition for cointegration:

- X and Y have stochastic trends. However, if they are cointegrated, the error does not have such a trend. Y and X will not diverge from one another; Y and X will trend together.

- In economic model involving an equilibrium concept, ε is the equilibrium error. If Y and X are cointegrated then the equilibrium error stays small.

- If Y and X are cointegrated then there is an equilibrium relationship between them. If they are not, then no equilibrium relationship exists.

- If Y and X are cointegrated then their trends will cancel each other out.

Example: Cointegration Between the Prices of Two Goods

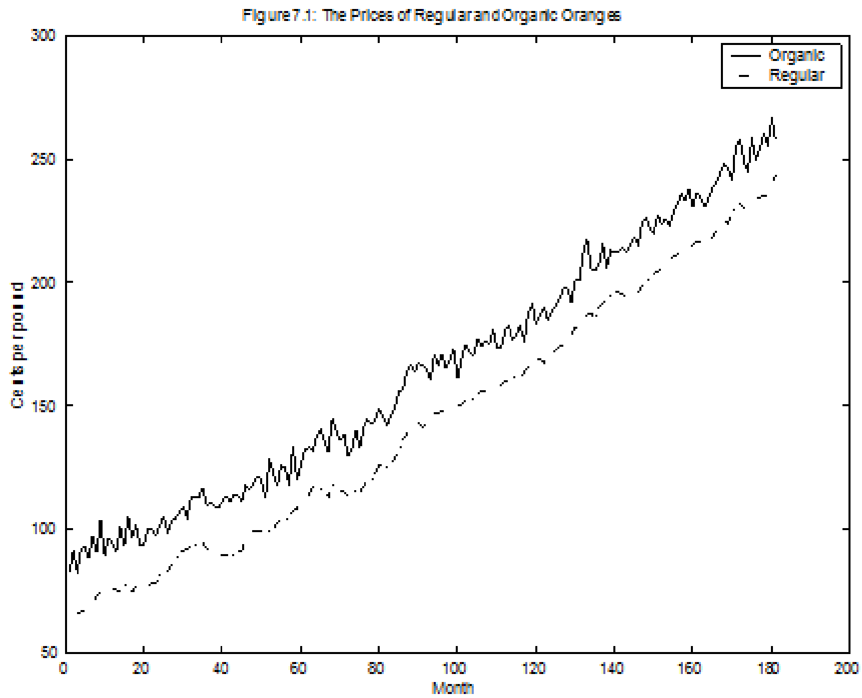
- Similar goods should be close substitutes for each other and therefore their prices should be cointegrated.

- Data for 181 months on the prices of regular oranges and organic oranges in a certain market.

- Although the prices of these two products will fluctuate due to the vagaries of supply and demand, market forces will always keep the price difference between the two goods roughly constant.

- Figure provides visual evidence for cointegration

- Many examples of cointegration, especially in macroeconomics and finance.



- Short term and long term interest rates

- Purchasing power parity and the permanent income hypothesis, theories of money demand, etc. etc.

1.2.3 Estimation and Testing with Cointegrated Variables

- If Y and X are cointegrated, then the spurious regression problem does not apply and OLS methods are fine.

- Coefficient from this regression is the long run multiplier.

- Regression of Y on X is called *cointegrating regression*.

- But it is important to verify that Y and X are cointegrated.

- Many tests for cointegration exist and some computer software packages will do several tests

- The textbook discussed a popular test called the *Johansen test* (but not covered in this course).

- But first introduce *Engle-Granger* test

- Basic idea: test for unit root in residual

- Remember: cointegration occurs if errors do not have unit root (and residuals are estimates of errors)

- Engle-Granger test has following steps:

1. Run the regression of Y on an intercept and X and save the residuals.

2. Carry out a Dickey-Fuller test on the residuals (without including a deterministic trend).

3. If the unit root hypothesis is rejected then conclude that Y and X are cointegrated. However, if the unit root is accepted then conclude cointegration does not occur.

- Note: Critical values (see Table 7.2 in textbook) are slightly different from the critical values for the Dickey-Fuller test.

- Note: Usually you not include a deterministic trend when doing this test (i.e. if it were included it could mean the errors could be growing steadily over time. This would violate the idea of cointegration.)

- Regression in Step 2 above is usually:

$$\Delta\widehat{\varepsilon}_t = \phi\widehat{\varepsilon}_{t-1} + \gamma_1\Delta\widehat{\varepsilon}_{t-1} + .. + \gamma_{p-1}\Delta\widehat{\varepsilon}_{t-p+1} + u_t.$$

- Remember: cointegration is found if we reject hypothesis of unit root in residuals (i.e. null hypothesis "no cointegration" and we conclude "cointegration is present" only if we reject unit root in errors hypothesis)

Example: Cointegration Between the Prices of Two Goods (continued)

- Regression of $Y =$ the price of organic oranges on $X =$ the price of regular yields:

$$\widehat{Y}_i = 20.686 + 0.996 X_i.$$

- Engle-Granger test: carry out a unit root test on the residuals, $\widehat{\varepsilon}_t$, from this regression.

- Remember (from Chapter 6) that first step in doing the unit root test is to correctly select the lag length.

- Use the sequential strategy, turns out that AR(1) specification for the residuals is appropriate.

- Dickey-Fuller strategy says we should regress $\Delta\widehat{\varepsilon}_t$ on $\widehat{\varepsilon}_{t-1}$.

- t-statistic on $\widehat{\varepsilon}_{t-1}$ in the resulting regression is -14.54.

- Since sample size is 180 and Table 7.2 says that the 5% critical value is between -3.39 and -3.33, we reject the unit root hypothesis and conclude that the residuals do not have a unit root.

- Thus, the two price series are indeed cointegrated.

- Since cointegrated, do not need to worry about the spurious regressions problem.

- Hence, estimate of the long run multiplier is 0.996.

More Practical Issues in Cointegration Testing

- Note: we have focussed on two variables, Y and X. In practice, you may have many more variables.

- Example: consider the three variables: income (Y), consumption (C) and investment (I).

- Some macroeconomists claim that the ratios $\frac{C}{Y}$ and $\frac{I}{Y}$ are roughly stable in the long-run.

- Common to take logs, so:

$$\ln(C) - \ln(Y) \approx \text{constant}$$

and

$$\ln(I) - \ln(Y) \approx \text{constant}.$$

- If $\ln(C), \ln(Y)$ and $\ln(I)$ all contain unit roots, reasoning above suggests that two cointegrating relationships might occur.

- Engle-Granger test (based on a cointegrating regression involving all three variables), would only find whether cointegration is/is not present (not tell you how many cointegrating relationships)

- What should you do in this case? One option is to use the Johansen test (not covered in course)

- Or you could do multiple Engle-Granger tests using different combinations of your variables.

- E.g. do an Engle-Granger test with all three variables, $\ln(C), \ln(Y)$ and $\ln(I)$.

- If you find cointegration with this test, then at least one cointegrating relationship exists.

- Then you could do three more Engle-Granger tests: i) using $\ln(C)$ and $\ln(Y)$, i) using $\ln(I)$ and $\ln(Y)$ and iii) using $\ln(C)$ and $\ln(I)$. If two cointegrating relationships exist, then these latter tests will indicate it.

- Another issue: Often the researcher has a suspicion as to what the cointegrating relationship should be.

- E.g. if C/Y roughly constant the regression:

$$\ln(C) = \alpha + \beta \ln(Y) + \varepsilon$$

should have coefficient $\beta = 1$.

- Step 1 of the Engle-Granger test uses OLS to estimate β.

- But you could set $\beta = 1$ if you wanted to test whether $\ln(C)$ and $\ln(Y)$ are cointegrated with a cointegrating coefficient of $\beta = 1$.

- You test this by constructing a new variable, Z, where

$$Z = \ln(C) - \ln(Y)$$

and then test whether Z has a unit root using Dickey-Fuller test.

- If Z is found to be stationary, then you know $\ln(C) - \ln(Y)$ is stationary and this is a cointegrating relationship.

- This is intended only as a brief introduction to cointegration. The textbook includes much more, including a discussion of error correction models which are commonly used when cointegration occurs.

1.3 What if Y and X Have Unit Roots but are NOT Cointegrated?

- Do not run regression of Y on X (spurious regression problem).

- Maybe you should rethink your basic model.

- E.g. instead of working with Y and X themselves, e.g., difference them. Remember that if Y and X have unit roots, then ΔY and ΔX should be stationary.

- Then you could estimate the original ADL model, but with changes in the variables:

$$\Delta Y_t = \alpha + \delta t + \gamma_1 \Delta Y_{t-1} + .. + \gamma_{p-1} \Delta Y_{t-p+1}$$
$$+\omega_0 \Delta X_t + \omega_1 \Delta X_{t-1} + .. + \omega_{q-1} \Delta X_{t-q+1} + \varepsilon_t$$

- If Y and X have unit roots, then all the variables in the regression above will be stationary and OLS methods for estimation and testing can be used.

- Problem: sometime you end up with a regression where coefficients do not have interpretation you want. But sometimes, this is good solution.

- E.g. suppose $Y =$ log wages and $X =$ log prices and both have unit roots but are not cointegrated.

- If you work with ΔY and ΔX, then your variables have a nice interpretation as being wage inflation and price inflation

1.4 Summary and Further Directions

- So far we have shown how to build time series regression models for the three main cases:

i) when all variables are stationary,

ii) when all variables have unit roots and are cointegrated

iii) when all variables have unit roots but are not stationary.

- But what do you use these models for?

- One answer is the usual regression one (e.g. coefficients measure marginal effects)

- But there are lots of other things such as Granger causality, forecasting and issues which arise with Vector autoregressions.

- These are discussed in textbook (and may be relevant for your future work, e.g., on a dissertation). But we will not have time to cover them in this course, so you are not responsible for this material

1.5 Chapter Summary

1. If all variables are stationary, then an ADL(p, q) model can be estimated using OLS. Econometric techniques are all standard.

2. A variant on the ADL model is often used to avoid potential multicollinearity problems. It provides a straightforward estimate of the long run multiplier.

3. If all variables are nonstationary, great care must be taken in the analysis due to the spurious regression problem.

4. If all variables are nonstationary but the regression error is stationary, then cointegration occurs.

5. If cointegration is present, the spurious regression problem does not occur.

6. Cointegration is an attractive concept for economists since it implies that an equilibrium relationship exists.

7. Cointegration can be tested using the Engle-Granger test. This test is a Dickey-Fuller test on the residuals from the cointegrating regression.

8. If the variables have unit roots but are not cointegrated, you should not work with them directly. Rather you should difference them and estimate an ADL model using the differenced variables.