Gretl Workshop 9
Autocorrelation

Ramanathan data9-7

New Car Sales

Naïve Model

- Start gretl, and use File/Open Data/sample file/ to load data9-7 under Ramanathan tab.
- Notice variable definitions, sample
- Choose Model/Ordinary least squares
- Double-click on QNC (dependent variable)
- Add the following 6 variables (PRICE – POP) as explanatory variables. Then Ok.

Naïve Model

- **Model 1: OLS estimates using the 64 observations 1975:1-1990:4**
- **Dependent variable: QNC**

VARIABLE	COEFFICIENT	STDERROR	T STAT	P-VALUE
const	25531.7	6606.09	3.865	0.00029 ***
PRICE	50.1164	22.9729	2.182	0.03328 **
INCOME	630.491	310.333	2.032	0.04686 **
PRIME	-44.3828	14.0302	-3.163	0.00250 ***
UNEMP	-41.8123	73.7499	-0.567	0.57298
STOCK	14.0646	47.5273	0.296	0.76836
POP	-150.679	39.3046	-3.834	0.00032 ***

- **Sum of squared residuals = 3.53659e+006**
- **Standard error of residuals = 249.089**
- **Unadjusted R-squared = 0.493523**
- **Adjusted R-squared = 0.44021**
- **F-statistic (6, 57) = 9.25702 (p-value < 0.00001)**
- **Durbin-Watson statistic = 1.65514**
- **First-order autocorrelation coeff. = 0.159997**

Durbin-Watson Test

- Find the value of the Durbin-Watson statistic.
- From the DW table find d-lower and d-upper for k'=6, n=65
- You should find that d-lower = 1.404 d-upper = 1.805.
- What do you conclude about autocorrelation?

Breusch-Godfrey LM Test

- In equation window choose Tests/autocorrelation and allow 4 lags (why 4 in this case?).
- Notice form of the regression for this test.
- What are null and alternative hypotheses?
- What do you conclude?
- Why such a strong difference between the two test results?

Breusch-Godfrey test for autocorrelation up to order 4

- OLS estimates using the 60 observations 1976:1-1990:4
- Dependent variable: uhat

VARIABLE	COEFFICIENT	STDERROR	T STAT	P-VALUE
const	-15324.9	6140.02	-2.496	0.01598 **
PRICE	-41.7727	20.3664	-2.051	0.04563 **
INCOME	-86.8977	246.579	-0.352	0.72604
PRIME	1.98653	11.1825	0.178	0.85973
UNEMP	31.0575	59.0998	0.526	0.60160
STOCK	-18.7118	37.3808	-0.501	0.61891
POP	94.3712	35.9822	2.623	0.01159 **
uhat_1	0.0653533	0.112756	0.580	0.56484
uhat_2	-0.0575221	0.113925	-0.505	0.61588
uhat_3	0.0818782	0.114339	0.716	0.47733
uhat_4	0.623711	0.115089	5.419	<0.00001 ***

- Unadjusted R-squared = 0.441439
- Test statistic: LMF = 9.681371,
- with p-value = P(F(4,49) > 9.68137) = 7.51e-006

- Alternative statistic: TR^2 = 26.486364,
- with p-value = P(Chi-square(4) > 26.4864) = 2.52e-005

- Ljung-Box Q' = 24.4712 with p-value = P(Chi-square(4) > 24.4712) = 6.43e-005

A more serious model

- Based on discussion of multicollinearity, heteroscedasticity, what would be a better model specification?
- Consider also the meaning of the dummy variables. What is their likely use in this model?

A more serious model

- From main window choose

  Add/Define new variable

  and enter the following variable definitions for the per capita variables.
- QNCPC=QNC/POP
- STOCKPC=STOCK/POP

Model 2

- Choose Model/Ordinary least squares.
- Double-click on QNCPC (dependent)
- Add as explanatory variables: PRICE, INCOME, PRIME, UNEMP, STOCKPC.
- Test for autocorrelation with Durbin-Watson test (k'=5, n=65)

    d-lower=1.438 d-upper=1.767

What do you conclude?

Breusch-Godfrey LM Test

- In equation window choose Tests/autocorrelation and allow 4 lags
- Notice the form of this test regression
- What do you conclude? At what lag is autocorrelation most severe? What is reason for this?

- **Breusch-Godfrey test for autocorrelation up to order 4**
- **OLS estimates using the 60 observations 1976:1-1990:4**
- **Dependent variable: uhat**

VARIABLE	COEFFICIENT	STDERROR	T STAT	P-VALUE
const	37.1736	17.4744	2.127	0.03835 **
PRICE	0.0270202	0.0454096	0.595	0.55450
INCOME	1.00203	0.973909	1.029	0.30849
PRIME	-0.0504756	0.0440464	-1.146	0.25727
UNEMP	-0.154806	0.249258	-0.621	0.53738
STOCKPC	-103.577	36.6543	-2.826	0.00676 ***
uhat_1	0.138346	0.109687	1.261	0.21306
uhat_2	-0.0202363	0.111373	-0.182	0.85655
uhat_3	0.0542624	0.111670	0.486	0.62915
uhat_4	0.628663	0.109797	5.726	<0.00001 ***

- **Unadjusted R-squared = 0.471973**

- **Test statistic: LMF = 11.173029,**
- **with p-value = P(F(4,50) > 11.173) = 1.49e-006**

- **Alternative statistic: TR^2 = 28.318376,**
- **with p-value = P(Chi-square(4) > 28.3184) = 1.08e-005**

- **Ljung-Box Q' = 26.5193 with p-value = P(Chi-square(4) > 26.5193) = 2.49e-005**

Treating Autocorrelation

- One approach to treating autocorrelation is to respecify the model. Since seasonality appears to be source of problem, add the seasonal dummies to the regression.
- Choose Model/ordinary least squares from main window and add 3 of the four seasonal dummies (why not all 4?).
- Perform both tests for autocorrelation.

Autoregressive Correction

- Unable to solve problem by adding variables.
- Try generalized Cochrane-Orcutt procedure with lags 1 and 4.
- From main window choose Model/time series/autoregressive estimation and enter lags 1 and 4 (why?). Then Ok.
- Apply DW test for autocorrelation again.

Exercise

Here the error term is modeled as an AR(1,4) process: $u_t = \rho_1 u_{t-1} + \rho_4 u_{t-4} + \varepsilon_t$.

1. Present the form of the quasi-differenced equation for this case.
2. Describe the steps of the Cochrane-Orcutt procedure for this case, presenting explicitly any equations used at each step.

Construct LM Test

- From equation window choose

  Save/residuals

  and give them the name “ar_res”
- From main window choose Model/ordinary least squares, and double-click on ar_res
- Click on the Lags box (bottom left) and choose lags of dependent variable; then select lags 1 to 4; close this window. Then ok.
- Test overall significance with general F-statistic.

Explanation of gretl steps

- Double click on ar_res to set it as the dependent variable in the test equation
- The explanatory variables from the previous regression automatically appear.
- The Lags box allows you to add lags of any variables; in this case you want lags 1 to 4 of ar_res (the dependent variable).
- Now test overall significance of regression

Lessons

- Durbin-Watson test has limitations:
  - For AR(1) only
  - Region of indeterminacy
  - Model cannot contain lagged dependent variables or endogenous variables as regressors.
- Quarterly data with seasonality may result in AR(4) process.