Computer Handout 4: In-sample Forecast Diego Escobari Econ 3342

This Computed Handout 4 will provide an example of the following concepts:

- 1) Simple and multiple regression.
- 2) In-sample forecast.
- 3) In-sample forecast errors

We will be using the data set Handout04 from the class website. The data set is already formatted for EViews (or gretl) and contains for key components of U.S. real GDP: Manufacturing, retail, services, and agriculture. The series correspond to annual data from 1960 to 2001 measured in millions of dollars.

We want to estimate the following model to see how the agricultural GDP has been changing over the years:

$$agriculture_t = \beta_0 + \beta_1 year_t + \varepsilon_t \tag{1}$$

The variable "year" is just takes the value of the corresponding year: 1960, 1961, ..., 2001.

To generate the variable year you have to type the following command: genr year = @year

Now, to estimate the model in equation (1), you have to type the command: "LS agriculture c year" to obtain the following regression output:

Dependent Variable: Method: Least Square Date: 09/24/10 Time Sample: 1960 2001 Included observations	es e: 00:03	E		
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C YEAR	-6119.273 3.126007	165.4182 0.083522	-36.99275 37.42740	0.0000 0.0000
R-squared Adjusted R-squared S.E. of regression Sum squared resid Log likelihood Durbin-Watson stat	0.972238 0.971544 6.560855 1721.793 -137.5779 1.298698	Mean deper S.D. depend Akaike info Schwarz cri F-statistic Prob(F-stati	dent var criterion terion	71.78352 38.89304 6.646567 6.729313 1400.810 0.000000

Notice that the interpretation of the slope coefficient β_1 is the same as before: If year increases by one unit, then the agricultural GDP (aGDP) will increase by 3.12 million dollars. This means that in a given year the aGDP is 3.12 million dollars greater than the aGDP the year before. The p-value indicates that

the variable "year" is highly significant and the R² shows that time (year) explains 97% of the variation in aGDP.

What happened in the year zero? The aGDP is estimated to be negative 6119 million dollars. Does that make sense? No! That's why you have to be very careful in using these type of models to predict out-of-sample values.

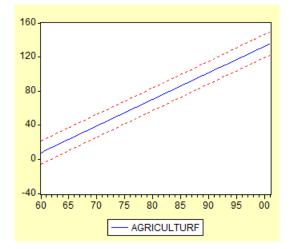
Let's obtain the in-sample forecasted values for aGDP (agriculture):

 $agriculture_{t} = \hat{\beta}_{0} + \hat{\beta}_{1}year_{t}$ $agriculture_{t} = -6119.273 + 3.126 \ year_{t}$

This can be done by simply selecting the "Forecast" icon:

View Proc Object Print Name	Vorkfile: DATAHANDOUT04EVIEWS\Unt
M Forecast	
S Equation: UNTITLE	D Series: AGRICULTURE
Series names Forecast name: S.E. (optional): GARCH(optional):	Griculturi Griculturi (no dynamics in equation) Structural (ignore ARMA)
R A S S S I 1960 2001 S I Insert actuals for o	Output Forecast graph Forecast evaluation ut-of-sample observations
	OK Cancel

To obtain:



and most importantly, Eviews generated the variable "agricurturf" that contains the in-sample forecasted values. The difference between "agriculture" and "agriculturf" corresponds to the forecasting errors and this variable is automatically stored in "resid." You can obtain a graph of all these three components (actual value = agriculture, fitted value = agriculturf, forecasting error = resid) by selecting the following option:

w Proc Object Print Name Freez	e Est	imate Forecast Stats Resids
Representations Estimation Output		lE
Actual, Fitted, Residual	•	Actual,Fitted,Residual Table
ARMA Structure		Actual,Fitted,Residual Graph
Gradients and Derivatives	•	Residual Graph
Covariance Matrix		Standardized Residual Graph
Coefficient Tests	×	
Residual Tests	•	165.4182 -36.99275 0.0000
Stability Tests	+	0.083522 37.42740 0.0000
Label		Mean dependent var 71.78352
		S.D. dependent var 38.89304

To obtain:

