Tutorial 10

keywords: time series, serial correlation, Breusch-Godfrey test, errors, residuals, dummy variables

estimated reading time: 37 minutes

Quang Bui

October 2, 2018

Question 1

Excel file: *SeattleElectric2005-06.xlxs*

To open an Excel file with EViews,

$$Right\ click\ Excel\ file \rightarrow Open\ with \rightarrow EViews$$

SeattleElectric2005-06.xlxs contains data on average hourly electricity usage in colds months in Seattle, USA, from 1 October 2005 to 31 March 2006. Information about each day during this time period is held in the following variables:

$$\begin{aligned} date- &\ date\ in\ American\ format\ mm-dd-yyyy \\ dow- &\ day\ of\ the\ week \\ pubhol- &\ =1\ if\ that\ day\ was\ a\ public\ holiday, =0\ otherwise \\ avetemp- &\ average\ daily\ temperature \\ aveload- &\ average\ hourly\ electricity\ load\ on\ that\ day \end{aligned}$$

	DATE	DOW	PUBHOL	AVETEMP	AVELOAD
10/01/2005	2005 − 10 − 01	7	0	63	1618
10/02/2005	2005 − 10 − 02	1	0	60	1572
10/03/2005	2005 − 10 − 03	2	0	55	1834
10/04/2005	2005 − 10 − 04	3	0	59	1813
10/05/2005	2005 − 10 − 05	4	0	58	1804
10/06/2005	2005 − 10 − 06	5	0	58	1824
10/07/2005	2005 − 10 − 07	6	0	55	1843
10/08/2005	2005 − 10 − 08	7	0	55	1704
10/09/2005	2005 − 10 − 09	1	0	57	1661
10/10/2005	2005 − 10 − 10	2	1	58	1863
10/11/2005	2005 − 10 − 11	3	0	57	1819
10/12/2005	2005 − 10 − 12	4	0	57	1821
10/13/2005	2005 − 10 − 13	5	0	57	1867
10/14/2005	2005 − 10 − 14	6	0	56	1850
10/15/2005	2005 − 10 − 15	7	0	58	1724

Table 1: Data for the first 15 days during the cold months in Seattle, USA, October 2005 to March 2006.

$dow = 7:\ Saturday,\ dow = 1:\ Sunday,\ dow = 2:\ Monday,\ \ldots\ , dow = 6:\ Friday$

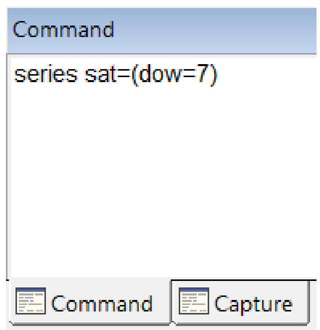
(a) Use *dow* to create dummy variables for different days of the week. Especially, create dummy variables for Saturday and Sunday and also a dummy variable called *wknd* which is 1 if the day is a Saturday or a Sunday, and 0 for any other day.

$$sat_i = \begin{cases} 1 & if\ i^{th}\ day\ is\ a\ Saturday \\ 0 & otherwise \end{cases}$$

$$sun_i = \begin{cases} 1 & if\ i^{th}\ day\ is\ a\ Sunday \\ 0 & otherwise \end{cases}$$

To create the Saturday dummy variables using *dow*,

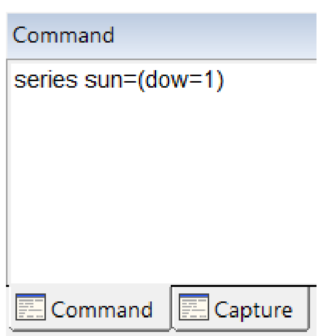
$$EViews\ Command\ window : series\ sat = (dow = 7)$$



$$(Press\ Enter\ to\ execute\ code)$$

To create the Sunday dummy variables using *dow*,

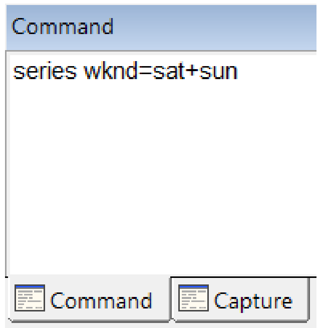
$$EViews\ Command\ window : series\ sun = (dow = 1)$$

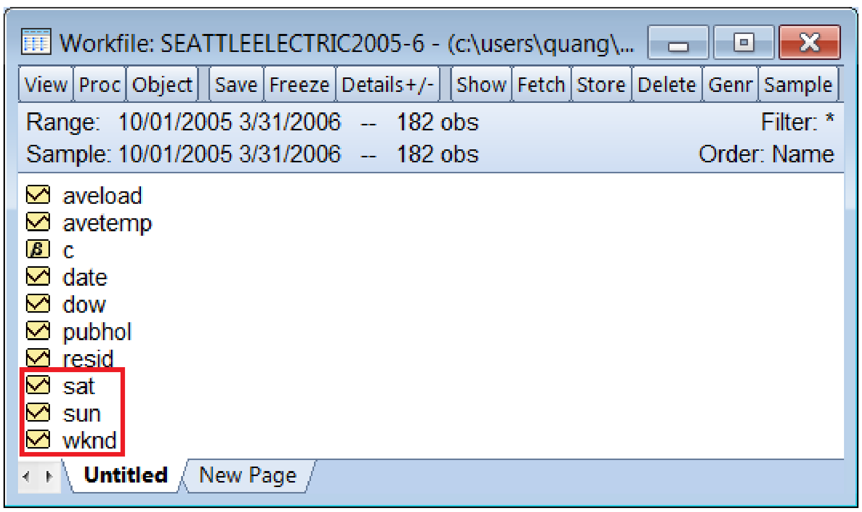


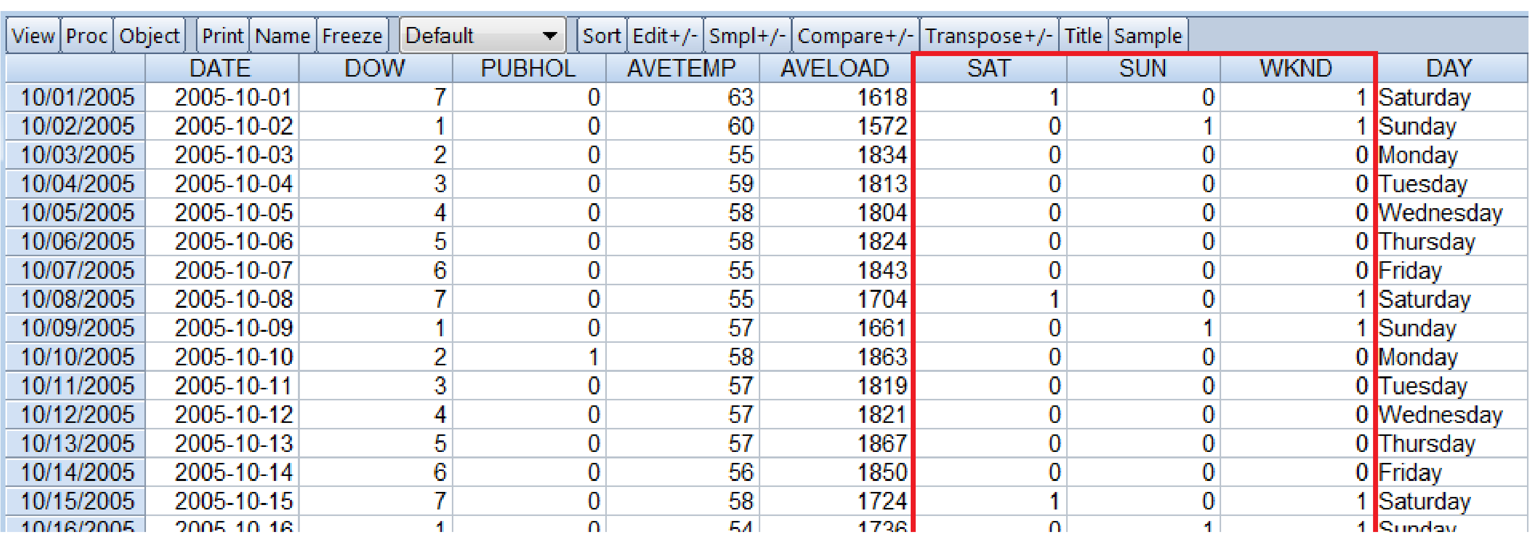
$$(Press\ Enter\ to\ execute\ code)$$

Using *sat* and *sun*, create the dummy variable *wknd*,

$$wknd_i = \begin{cases} 1 & if\ i^{th}\ day\ is\ a\ weekend \\ 0 & otherwise \end{cases}$$







	DATE	DOW	PUBHOL	AVETEMP	AVELOAD	SAT	SUN	WKND	DAY
10/01/2005	2005-10-01	7	0	63	1618	1	0	1	Saturday
10/02/2005	2005-10-02	1	0	60	1572	0	1	1	Sunday
10/03/2005	2005-10-03	2	0	55	1834	0	0	0	Monday
10/04/2005	2005-10-04	3	0	59	1813	0	0	0	Tuesday
10/05/2005	2005-10-05	4	0	58	1804	0	0	0	Wednesday
10/06/2005	2005-10-06	5	0	58	1824	0	0	0	Thursday
10/07/2005	2005-10-07	6	0	55	1843	0	0	0	Friday
10/08/2005	2005-10-08	7	0	55	1704	1	0	1	Saturday
10/09/2005	2005-10-09	1	0	57	1661	0	1	1	Sunday
10/10/2005	2005-10-10	2	1	58	1863	0	0	0	Monday
10/11/2005	2005-10-11	3	0	57	1819	0	0	0	Tuesday
10/12/2005	2005-10-12	4	0	57	1821	0	0	0	Wednesday
10/13/2005	2005-10-13	5	0	57	1867	0	0	0	Thursday
10/14/2005	2005-10-14	6	0	56	1850	0	0	0	Friday
10/15/2005	2005-10-15	7	0	58	1724	1	0	1	Saturday

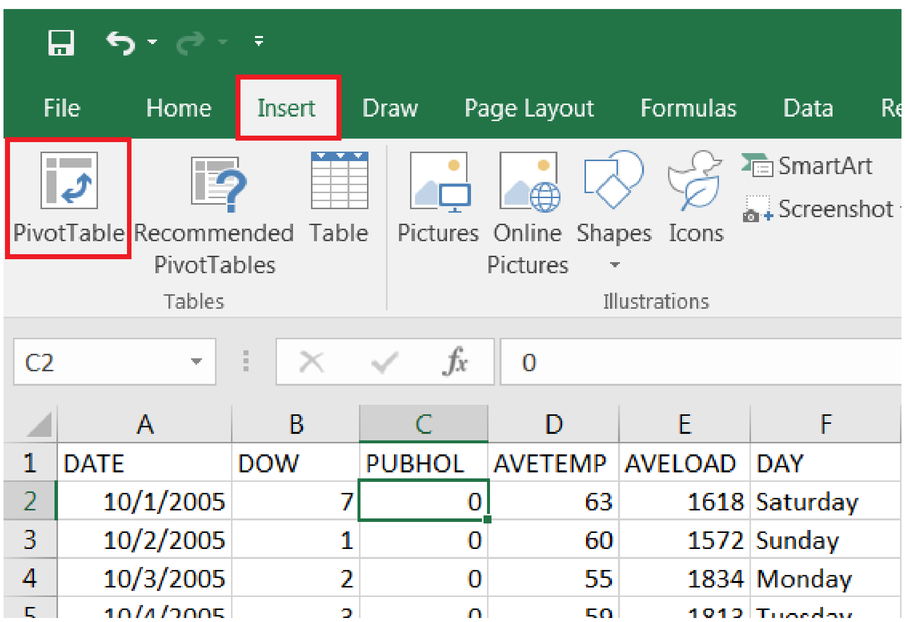
(b) The goal is to predict *aveload* in winter months (Seattle, USA, is in the Northern hemisphere) based on temperature and other available information (our assumption is that the weather bureau can produce temperature forecasts that are pretty accurate). Given this goal, have an appropriate look at the data. In particular, investigate differences in average load in different days of the week and look at the scatter plot of *aveload* against *avetemp* to familiarise yourself with the data.

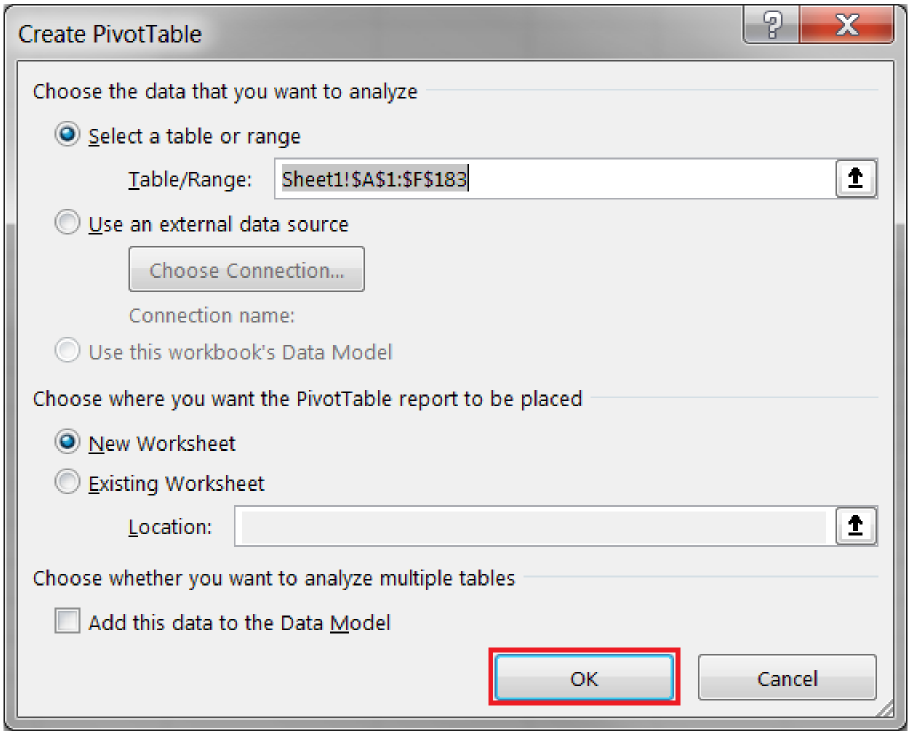
Remember that a good analyst does not confine himself or herself to any particular software. If a pivot chart would be better for visualising differences in average load in difference days of the week, then use it! If you have learnt another software (for example R, SPSS or STATA) in another unit that you think will be more helpful, use it!

Investigate the difference in average hourly electricity load in different days of the week

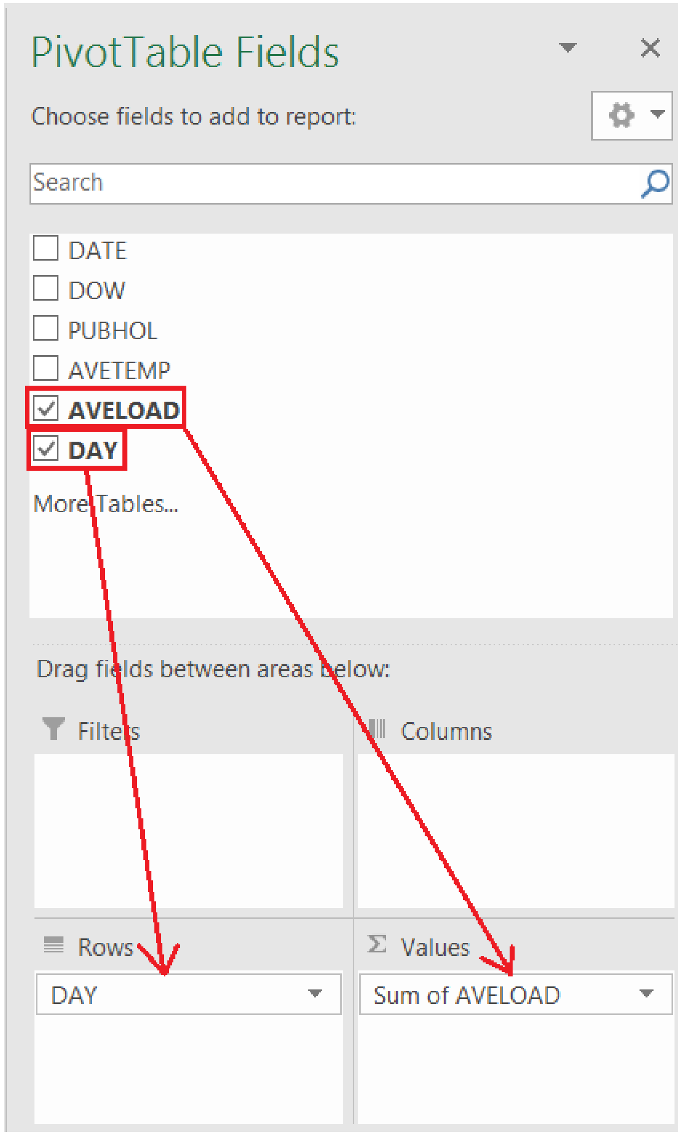
We can produce a column graph of the average of *aveload* for each day of the week using Excel's Pivot Chart. To produce the Pivot Chart, select any cell in the data set then,

$$Insert \rightarrow PivotTable \rightarrow OK$$





From the PivotTable Fields, drag and drop *day* and *aveload* into the Row and Value windows respectively,



We want the mean of the average hourly electricity load for each day in the Pivot Table,

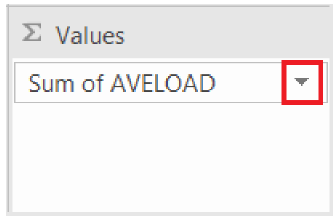
$$\overline{average\ Monday\ hourly\ electricity\ load} = \frac{\sum average\ Monday\ hourly\ electricity\ load}{no.\ of\ Mondays\ in\ sample}$$

$$\overline{average\ Tuesday\ hourly\ electricity\ load} = \frac{\sum average\ Tuesday\ hourly\ electricity\ load}{no.\ of\ Tuesdays\ in\ sample}$$

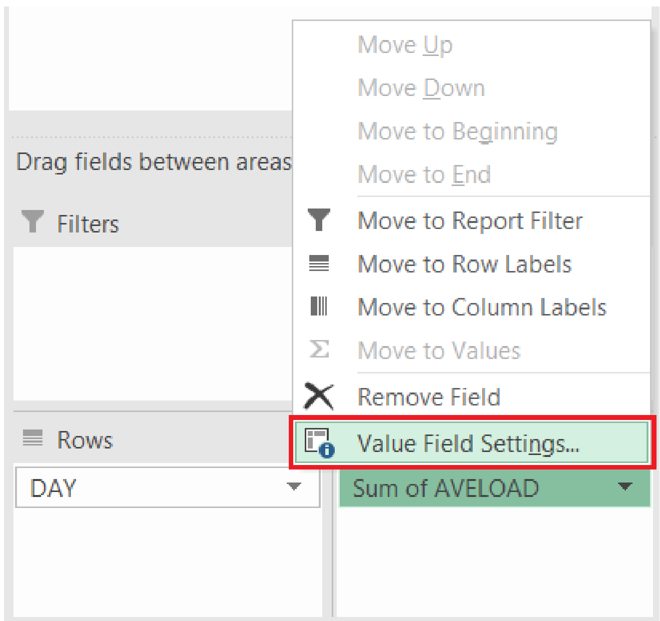
$$\vdots$$

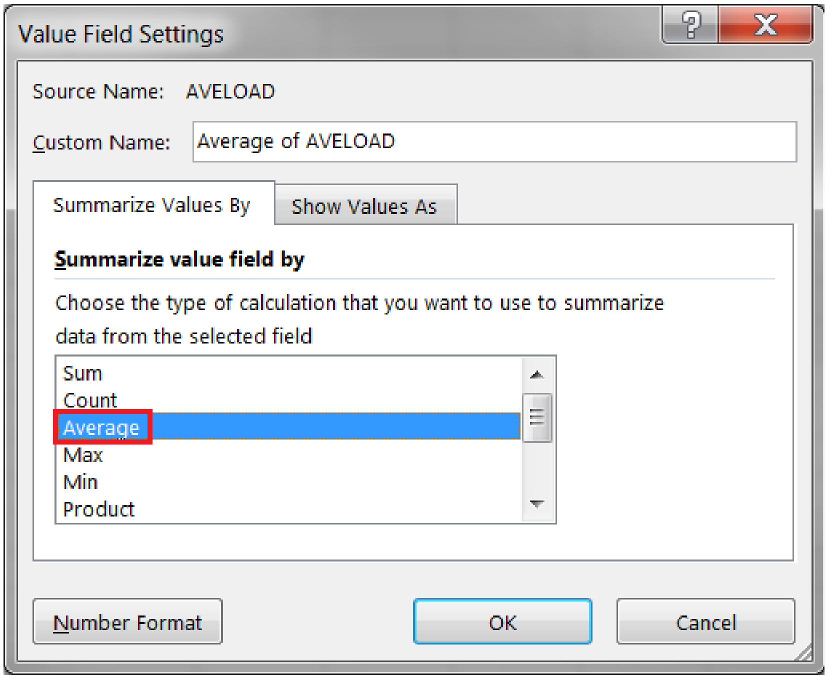
$$\overline{average\ Sunday\ hourly\ electricity\ load} = \frac{\sum average\ Sunday\ hourly\ electricity\ load}{no.\ of\ Sundays\ in\ sample}$$

To do this, change the Summarised Value of *avetemp* from Sum to Average. Click the drop-down arrow,



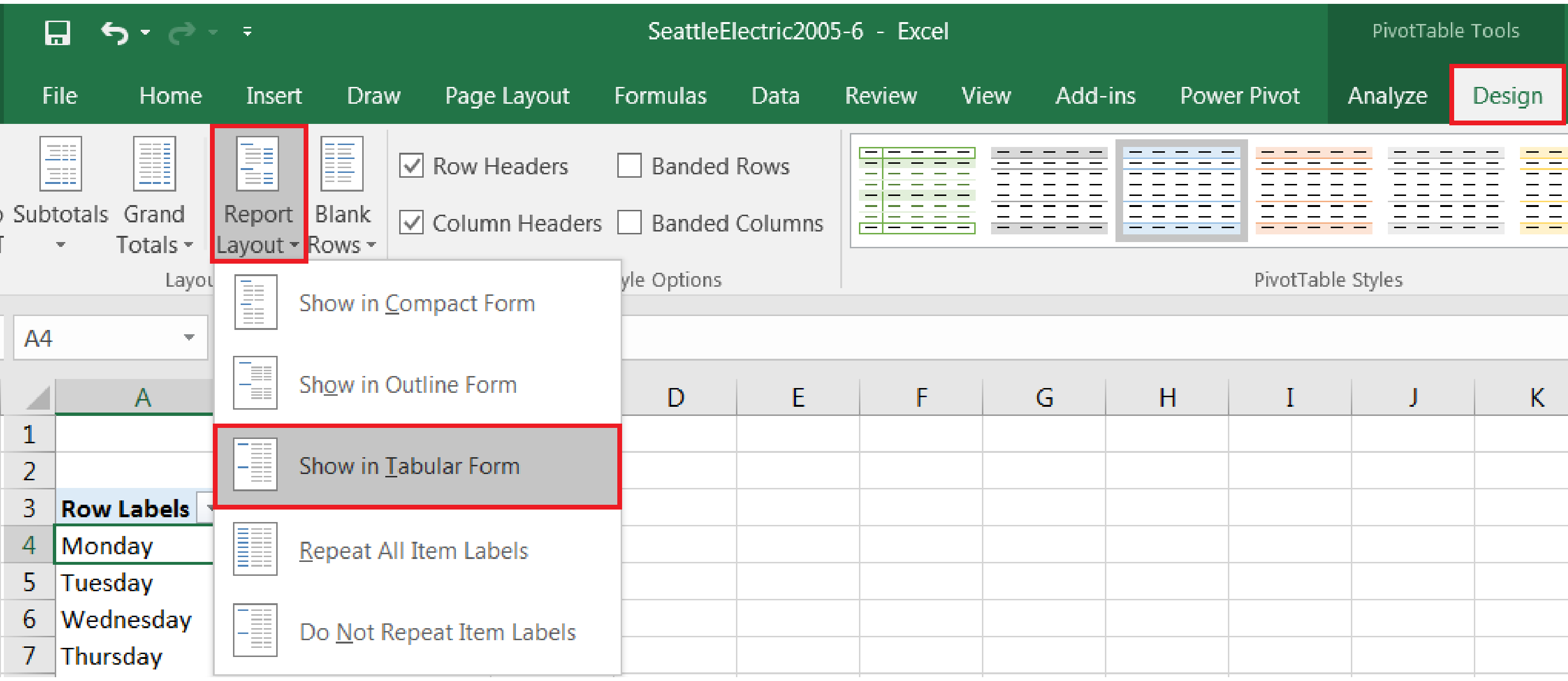
$$Value\ Field\ Settings \rightarrow Average \rightarrow OK$$





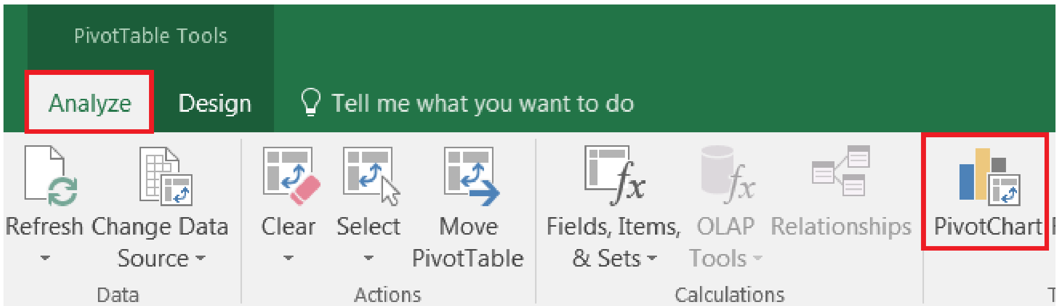
To automatically name the PivotChart headings with variables names,

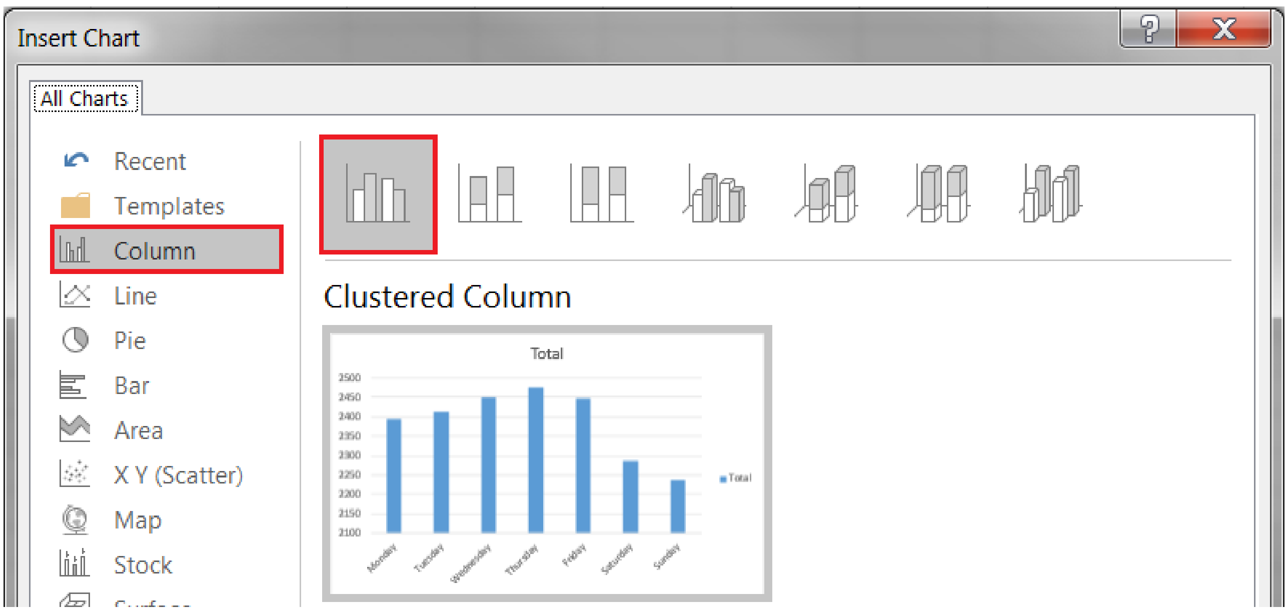
$$\textit{Design} \rightarrow \textit{Report Layout} \rightarrow \textit{Show in Tabular Form}$$

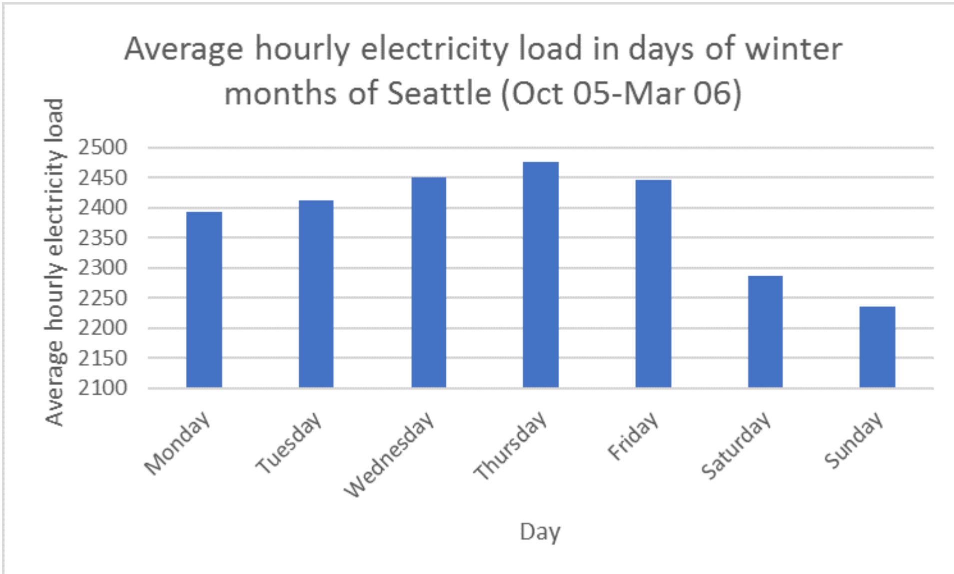


To produce a column graph from the Pivot Table,

$$Analyze \rightarrow PivotChart \rightarrow Clustered\ Column \rightarrow OK$$





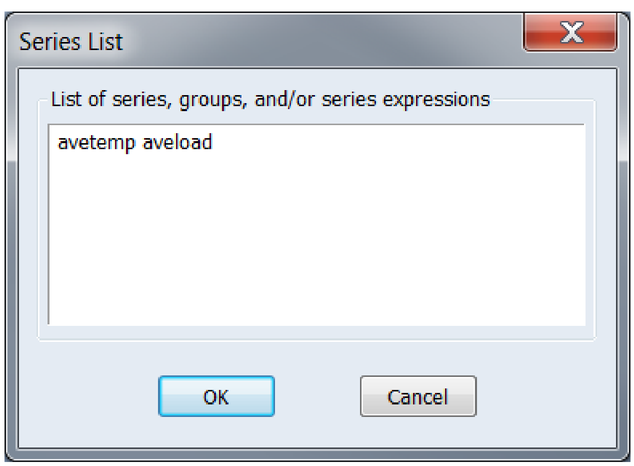


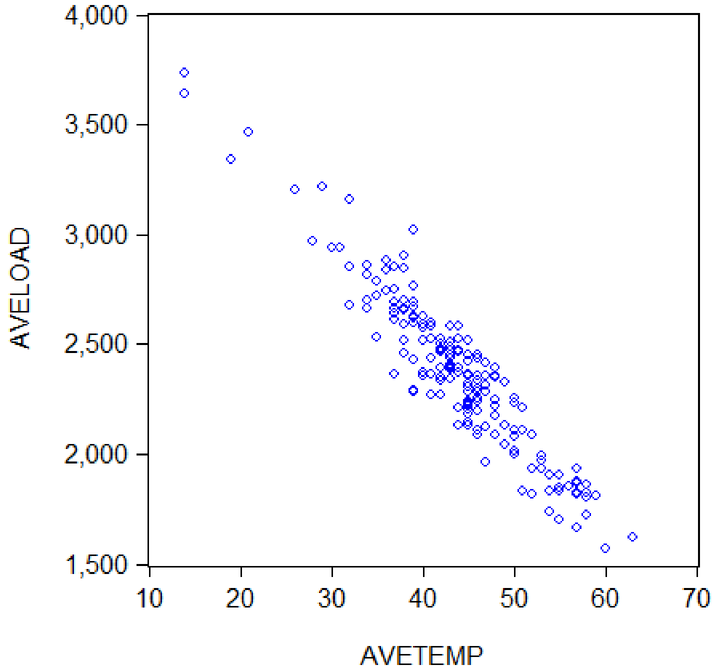
From the Pivot Chart, we find that, on average, electricity load during the winter months of Seattle is lowest during the weekends.

<u>Investigate the relationship between average hourly electricity load and average daily temperature</u>

Produce a scatter plot of *aveload* against *avetemp*,

$$Quick \rightarrow Graph \rightarrow avetemp\ aveload \rightarrow Specific : Scatter \rightarrow OK$$



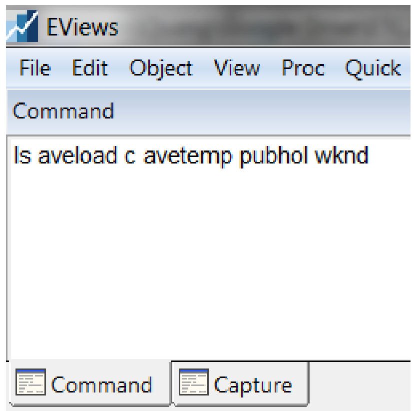


The scatter plot suggests a strong negative linear relationship between average hourly electricity load and average daily temperature. During the days of the winter months, as average temperate increases, electricity load decreases.

(c) Estimate a regression of *aveload* on a constant, *avetemp*, *wknd* and public holiday dummies,

$$aveload_t = \beta_0 + \beta_1 avetemp_t + \beta_2 wknd_t + \beta_3 pubhol_t + u_t$$

To estimate the model of average hourly electricity load from the Command Window in EViews,



Dependent Variable: AVELOAD
Method: Least Squares
Sample: 10/01/2005 3/31/2006
Included observations: 182

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4332.783	41.79892	103.6578	0.0000
AVETEMP	−43.58037	0.940259	−46.34931	0.0000
PUBHOL	−121.5734	31.00930	−3.920546	0.0001
WKND	−143.6717	16.35973	−8.782032	0.0000

R-squared	0.927222	Mean dependent var	2386.005
Adjusted R-squared	0.925995	S.D. dependent var	365.8870
S.E. of regression	99.53527	Akaike info criterion	12.06063
Sum squared resid	1763494.	Schwarz criterion	12.13105
Log likelihood	−1093.518	Hannan-Quinn criter.	12.08918
F-statistic	755.9290	Durbin-Watson stat	0.942970
Prob(F-statistic)	0.000000		

$$\widehat{aveload}_t = \underset{(41.7990)}{4332.783} - \underset{(0.9403)}{43.5804} avetemp_t - \underset{(16.3597)}{143.6717} wknd_t - \underset{(31.0093)}{121.5734} pubhol_t$$

We know that electricity usage, like anything else in life, is likely to be related to what the usage was in previous days.

$$\begin{aligned} corr(aveload_t, aveload_{t-1}) &\neq 0 \\ corr(aveload_t, aveload_{t-2}) &\neq 0 \\ &\vdots \end{aligned}$$

As a result, we suspect that part of the electricity load that cannot be explained by temperature,

$$u_t$$

is likely to be correlated over time,

$$\begin{aligned} corr(u_t, u_{t-1}) &\neq 0 \\ corr(u_t, u_{t-2}) &\neq 0 \\ &\vdots \end{aligned}$$

This is because

$$\begin{aligned} aveload_t &= \beta_0 + \beta_1 avetemp_t + \beta_2 wknd_t + \beta_3 pubhol_t + u_t \\ aveload_{t-1} &= \beta_0 + \beta_1 avetemp_{t-1} + \beta_2 wknd_{t-1} + \beta_3 pubhol_{t-1} + u_{t-1} \\ aveload_{t-2} &= \beta_0 + \beta_1 avetemp_{t-2} + \beta_2 wknd_{t-2} + \beta_3 pubhol_{t-2} + u_{t-2} \\ &\vdots \end{aligned}$$

When the errors are correlated with itself over different time periods, we say that the errors are serially correlated.

If that is the case, what consequences would that have for the OLS estimator? In such a case, can we use the model that you estimated to test that there is no difference in the intercept in weekdays and weekends?

Background

Consequences of serially correlated errors (autocorrelated errors)

When the assumption of serially uncorrelated errors is violated, the OLS estimator remains unbiased (if the unbiasedness assumptions hold) but will no longer be the most efficient unbiased estimator. More importantly, the OLS standard errors will be incorrect, and the t and F test-statistic will not follow a t and F distribution, even asymptotically.

- The OLS estimator of the parameters in our model of *aveload* is unbiased but no longer the most efficient unbiased estimator.
- Cannot use our estimated model of average hourly electricity load to test if there is a difference in intercept between weekdays and weekends.

(d) Use visual aids and a formal test to test the hypothesis that errors of this model are white noise against the alternative that they are generated by an AR(7). You should reject the null. Then investigate what kind of AR model would be sufficient for capturing the dynamics of the errors (the t-statistics in your BG auxiliary regression may give you a hint, and the partial autocorrelations of the residuals in the correlogram are informative as well). Re-estimate the model by adding an AR equation for u_t.

Background

Residuals and Errors

The residuals is the difference between the observed average hourly electricity load and the predicted/fitted average hourly electricity load,

$$\hat{u}_1 = aveload_1 - \widehat{aveload}_1$$

$$\hat{u}_2 = aveload_2 - \widehat{aveload}_2$$

$$\vdots$$

$$\hat{u}_n = aveload_n - \widehat{aveload}_n$$

Although the errors are unobserved, the residuals are observed. This means that we can obtain a line graph or the correlogram of the residuals to determine if there is evidence of serial correlation. Furthermore, we can formally test for serial correlation in the unobserved errors by performing a test on the observed residuals.

White Noise

If the errors are white noise, then they are serially uncorrelated

$$corr(u_t, u_{t-j}) = 0 \qquad for\ all\ j\ except\ j = 0$$

and its expectation equals to 0 and variance is constant (but unknown),

$$E(u_t) = 0$$
$$Var(u_t) = \sigma^2$$

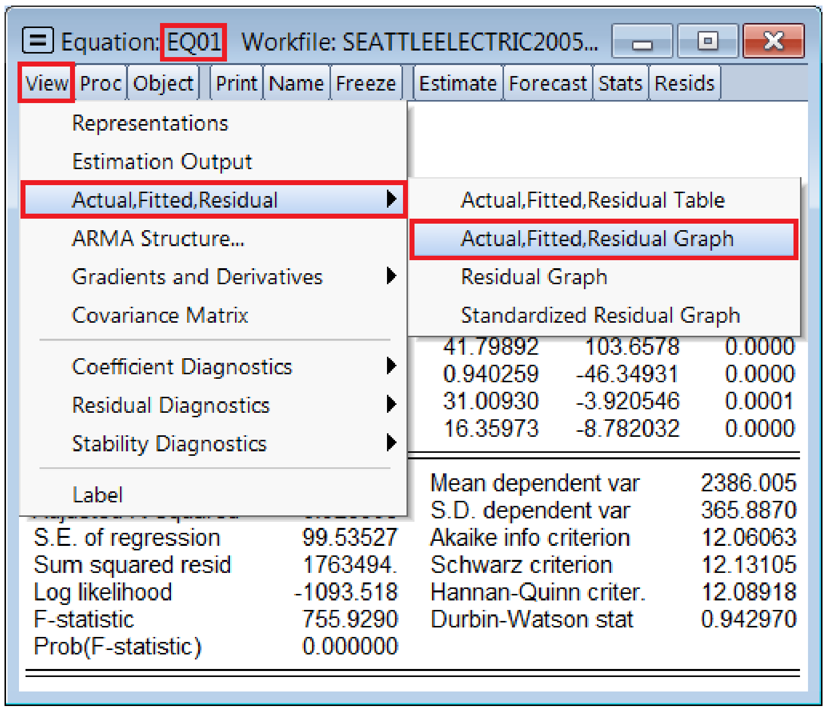
Visual aid to determine if the errors are white noise or serially correlated

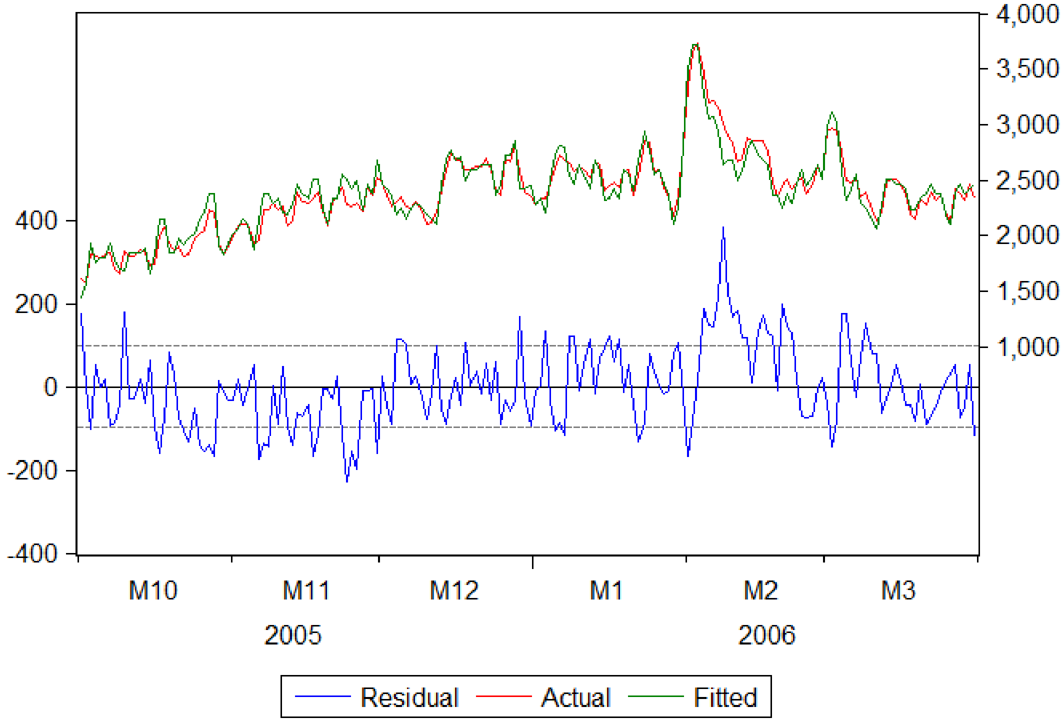
Since we cannot plot a line graph of the errors (because they are unobserved) to determine if there is presence of serial correlation in the errors, we instead plot a line graph of the residuals.

If the residuals follow a pattern, this suggests that the errors are serially correlated. On the other hand, a residual plot that does not exhibit any patterns i.e. one that looks like white noise would suggest that the errors are not serially correlated.

To obtain a line graph of $aveload$, $\widehat{aveload}$, and $\hat{u}$ from the estimated model of $aveload$ in EViews (I have named the equation $eq01$ in EViews),

$$eq01 \rightarrow View \ \rightarrow Actual, Fitted, Residual \rightarrow Residual\ Graph$$





The line graph of the residuals shows there is a slight cyclical pattern i.e.positive residuals are likely followed by positive residuals and negative residuals are likely followed by negative residuals. This indicates that there error is positively serially correlated.

After visually inspecting the residual plot, we should formally test for serially correlated errors.

Background

Breusch-Godfrey test for serial correlation in the errors

Since we are testing for serial correlation in u_t of order 7, our model of *aveload* is

$$aveload_t = \beta_0 + \beta_1 avetemp_t + \beta_2 wknd_t + \beta_3 pubhol_t + u_t \quad (1)$$
$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \cdots + \rho_7 u_{t-7} + e_t \quad (2)$$
$$\{e_t, t = 1, 2, \ldots, n \sim i.i.d(0, \sigma^2)\}$$

where e_t for $t = 1, 2, \ldots, n$ are independent and identically distributed with mean 0 and variance σ^2 (which means that e_t for $t = 1, 2, \ldots, n$ are unrelated to each other and have equal variance i.e. homoskedastic and serially uncorrelated).

The null and alternative hypothesis can be written as:

$$H_0 : \rho_1 = \rho_2 = \rho_3 = \cdots = \rho_7 = 0$$
$$H_1 : at\ least\ one\ of\ the\ 7\ autoregressive\ parameters\ \neq\ 0$$

Since we do not observe u_t, we cannot estimate equation (2) and test the above hypothesis.

To perform the Breusch-Godfrey test for serial correlation in the errors of model of *aveload* at any lag up to and include lag 7 (7th order serial correlation),

- Estimate the model of *aveload* (equation (1))
- Save the residuals from the estimated model of *aveload*
- Then estimate the following *auxiliary regression* . . .

  $$\hat{u}_t = \alpha_0 + \alpha_1 avetemp + \alpha_2 wknd + \alpha_3 pubhol + \rho_1 \hat{u}_{t-1} + \rho_2 \hat{u}_{t-2} + \cdots + \rho_7 \hat{u}_{t-7} + v_t$$

- We're testing the null hypothesis that there is no serial correlation in the errors at any lag up to and include lag 7, against the alternative hypothesis that there is serial correlation in the errors in at least one lag up to and including lag 7.
- Compare the calculated test statistics with the critical value and conclude if there is evidence of serial correlation in the errors in at least one lag up to and including lag 7.

There are two forms of the Breusch-Godfrey test (both are only valid asymptotically):

- The first form uses the F test statistic,

$$F = \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-k-1)} \overset{asy}{\sim} F_{q,n-k-1} \quad under\ H_0$$

- The second form uses the BG test statistic,

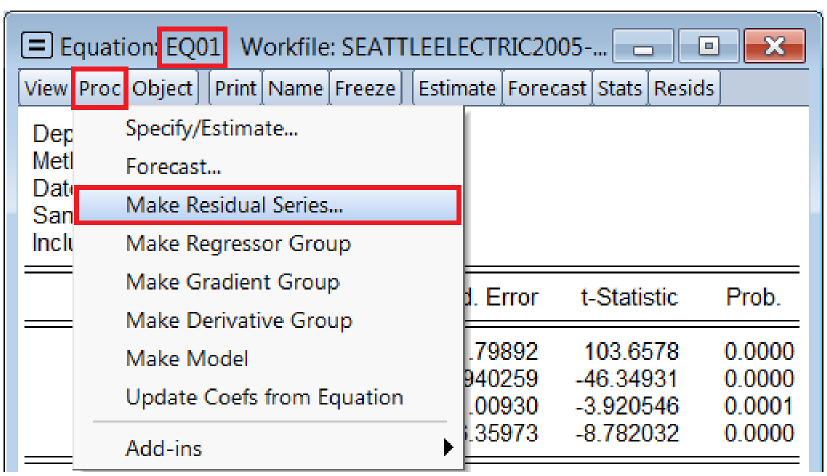
$$BG = (n-q) \times R^2_{\hat{u}} \overset{asy}{\sim} \chi^2_q \quad under\ H_0$$

where $q = 7$ represents the orders of lag that we are testing.

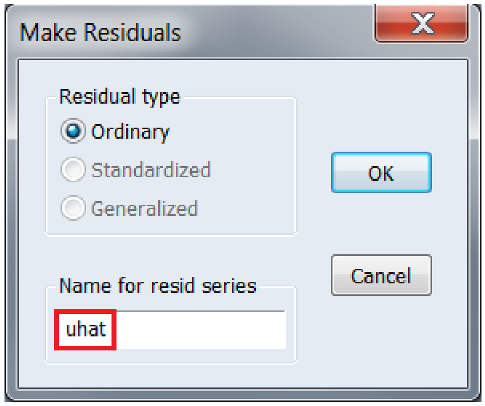
Note: In the BG test statistic, q is subtracted from n because we lose q observations when we include q lags of the residuals in the auxiliary regression.

To perform the Breusch-Godfrey test for 7th order serial correlation in the errors, generate the residuals from the estimated model of *aveload*,

$$eq01 \rightarrow Proc \rightarrow Make\ Residual\ Series$$



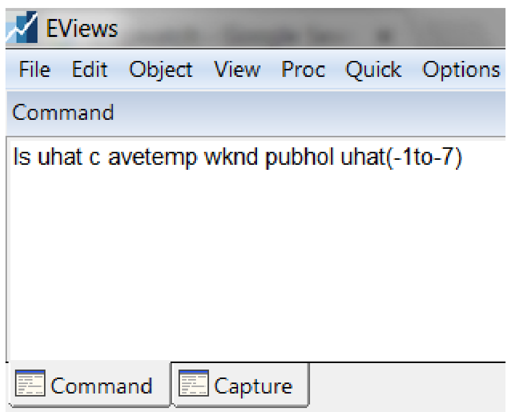
and name the residual series *uhat*,



then estimated the following auxiliary regression,

$$\hat{u}_t = \alpha_0 + \alpha_1 avetemp + \alpha_2 wknd + \alpha_3 pubhol + \rho_1 \hat{u}_{t-1} + \rho_2 \hat{u}_{t-2} + \cdots + \rho_7 \hat{u}_{t-7} + v_t$$

$$uhat\ c\ avetemp\ wknd\ pubhol\ uhat(-1to-7)$$



Dependent Variable: UHAT
Method: Least Squares
Sample (adjusted): 10/08/2005 3/31/2006
Included observations: 175 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	−25.17570	38.59715	−0.652268	0.5151
AVETEMP	0.537992	0.883942	0.608628	0.5436
WKND	−0.186538	14.09467	−0.013235	0.9895
PUBHOL	30.10940	26.60905	1.131547	0.2595
UHAT(-1)	0.493938	0.079677	6.199229	0.0000
UHAT(-2)	0.003029	0.087706	0.034533	0.9725
UHAT(-3)	−0.008167	0.086885	−0.094002	0.9252
UHAT(-4)	0.108065	0.086073	1.255496	0.2111
UHAT(-5)	0.153122	0.086524	1.769712	0.0786
UHAT(-6)	0.014457	0.087593	0.165050	0.8691
UHAT(-7)	−0.061279	0.078089	−0.784737	0.4337

R-squared	0.340335	Mean dependent var	−0.266614
Adjusted R-squared	0.300112	S.D. dependent var	99.16426
S.E. of regression	82.96015	Akaike info criterion	11.73539
Sum squared resid	1128711.	Schwarz criterion	11.93432
Log likelihood	−1015.847	Hannan-Quinn criter.	11.81608
F-statistic	8.461108	Durbin-Watson stat	1.928231
Prob(F-statistic)	0.000000		

$$n_{aux} = n - q = 182 - 7 = 175 \qquad R^2_{aux} = 0.3403$$

Auxiliary regression:

$$\hat{u}_t = \alpha_0 + \alpha_1 avetemp + \alpha_2 wknd + \alpha_3 pubhol + \rho_1 \hat{u}_{t-1} + \rho_2 \hat{u}_{t-2} + \cdots + \rho_7 \hat{u}_{t-7} + v_t$$

Estimated auxiliary regression:

$$\begin{aligned}\widehat{\hat{u}}_t &= -25.1757 + 0.5380 avetemp - 0.1866 wknd + 30.1094 pubhol + 0.4939 \hat{u}_{t-1} \\ &\quad + 0.0030 \hat{u}_{t-2} + \cdots - 0.0613 \hat{u}_{t-7} + v_t\end{aligned}$$

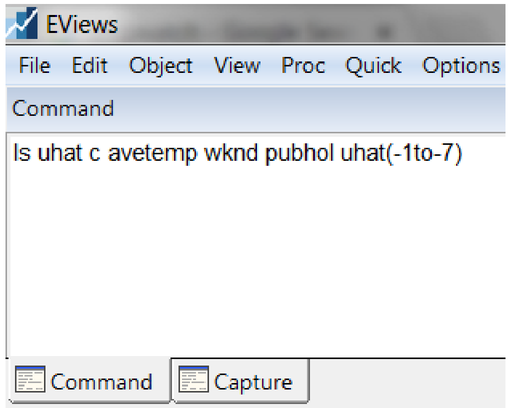
BG test statistic:

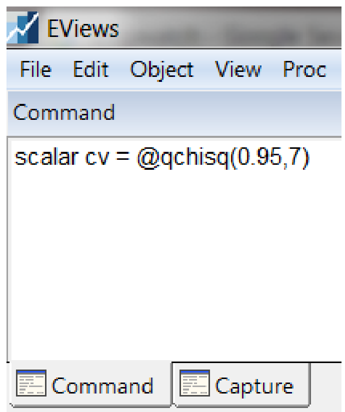
$$BG = (n-7) \times R^2_{\hat{u}} \overset{asy}{\sim} \chi^2_7 \quad under\ H_0$$

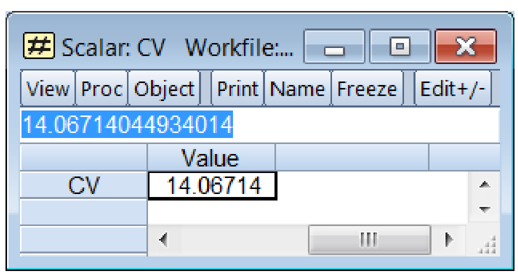
Calculate BG test statistic:

$$BG_{calc} = 175 \times 0.3403 = 59.5525$$

Critical value:







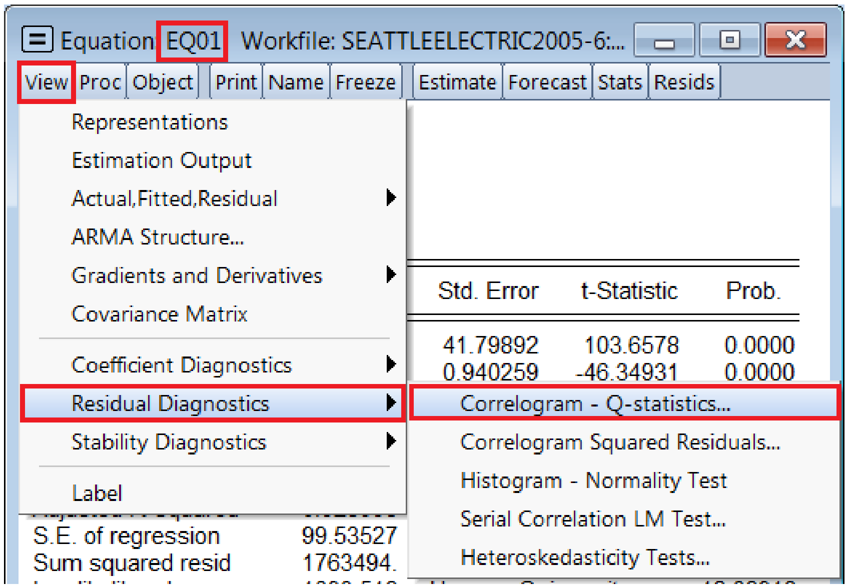
$$BG_{crit} = 14.0671$$

Since $BG_{calc} = 59.5525 > BG_{crit} = 14.0671$ we reject the null and conclude that there is sufficient evidence from our sample to suggesting that there is serial correlation in the errors in at least one lag up to and include the 7th lag.

Investigate what kind of AR model would be sufficient for capturing the dynamics of the errors (the t-statistics in your BG auxiliary regression may give you a hint, and the partial autocorrelations of the residuals in the correlogram are informative as well).

To obtain the correlogram of the residuals of the estimated model of *aveload*,

$$eq01 \rightarrow View \rightarrow Residual\ Diagnostics \rightarrow Correlogram\ Q - statistics$$



$$Lags\ to\ include: 36$$

Sample: 10/01/2005 3/31/2006
Included observations: 182

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.516	0.516	49.227	0.000
		2	0.288	0.030	64.680	0.000
		3	0.220	0.083	73.776	0.000
		4	0.277	0.175	88.159	0.000
		5	0.323	0.145	107.87	0.000
		6	0.237	-0.018	118.60	0.000
		7	0.116	-0.070	121.16	0.000
		8	0.058	-0.041	121.80	0.000
		9	0.193	0.173	129.01	0.000
		10	0.220	0.028	138.41	0.000
		11	0.224	0.085	148.26	0.000
		12	0.161	0.017	153.36	0.000

- Auxiliary regression shows that the first lag is statistically significant $t_{calc} = 6.1992$

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	−25.17570	38.59715	−0.652268	0.5151
AVETEMP	0.537992	0.883942	0.608628	0.5436
WKND	−0.186538	14.09467	−0.013235	0.9895
PUBHOL	30.10940	26.60905	1.131547	0.2595
UHAT(-1)	0.493938	0.079677	6.199229	0.0000
UHAT(-2)	0.003029	0.087706	0.034533	0.9725
UHAT(-3)	−0.008167	0.086885	−0.094002	0.9252
UHAT(-4)	0.108065	0.086073	1.255496	0.2111
UHAT(-5)	0.153122	0.086524	1.769712	0.0786
UHAT(-6)	0.014457	0.087593	0.165050	0.8691
UHAT(-7)	−0.061279	0.078089	−0.784737	0.4337

- Correlogram shows that 1st partial autocorrelation is statistically significant, with 4th partial autocorrelation also closely statistically significant
- Start by adding an AR(1) error in the model of *aveload*,

$$
\begin{aligned}
aveload_t &= \beta_0 + \beta_1 avetemp_t + \beta_2 wknd_t + \beta_3 pubhol_t + u_t \\
u_t &= \rho_1 u_{t-1} + e_t
\end{aligned}
$$

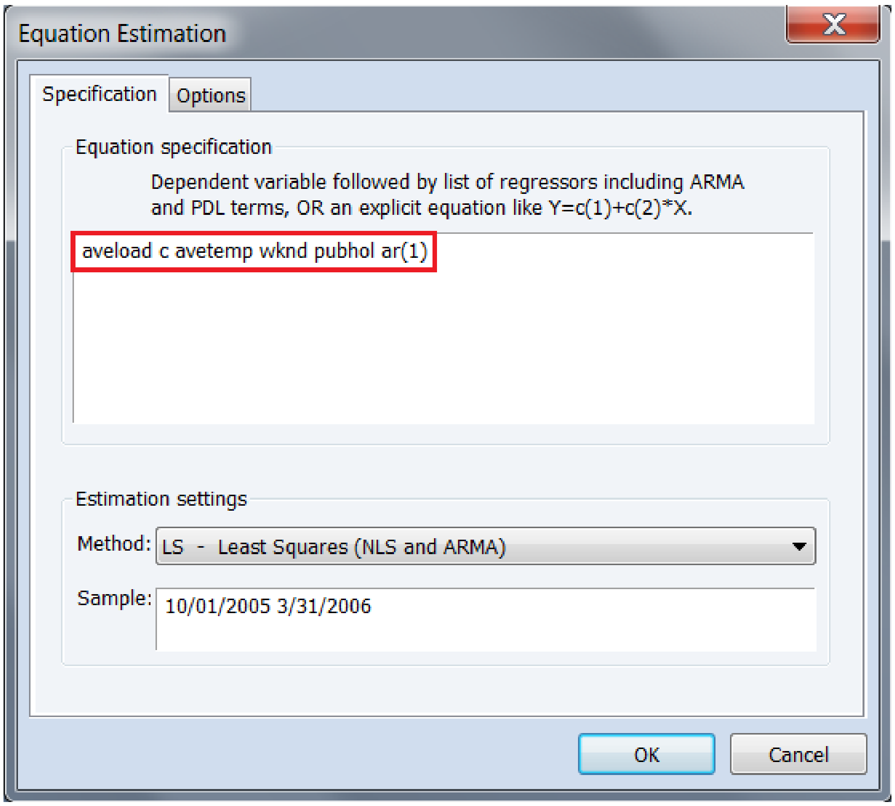
  then estimate this model and check the residual correlogram to see if there is a need to consider a higher order AR model for the errors (you could also inspect the residual plot to see if the amended model adequately captures the dynamics of errors).

To estimate model of *aveload* with an AR(1) error

$$aveload_t = \beta_0 + \beta_1 avetemp_t + \beta_2 wknd_t + \beta_3 pubhol_t + u_t$$
$$u_t = \rho_1 u_{t-1} + e_t$$

in EViews,

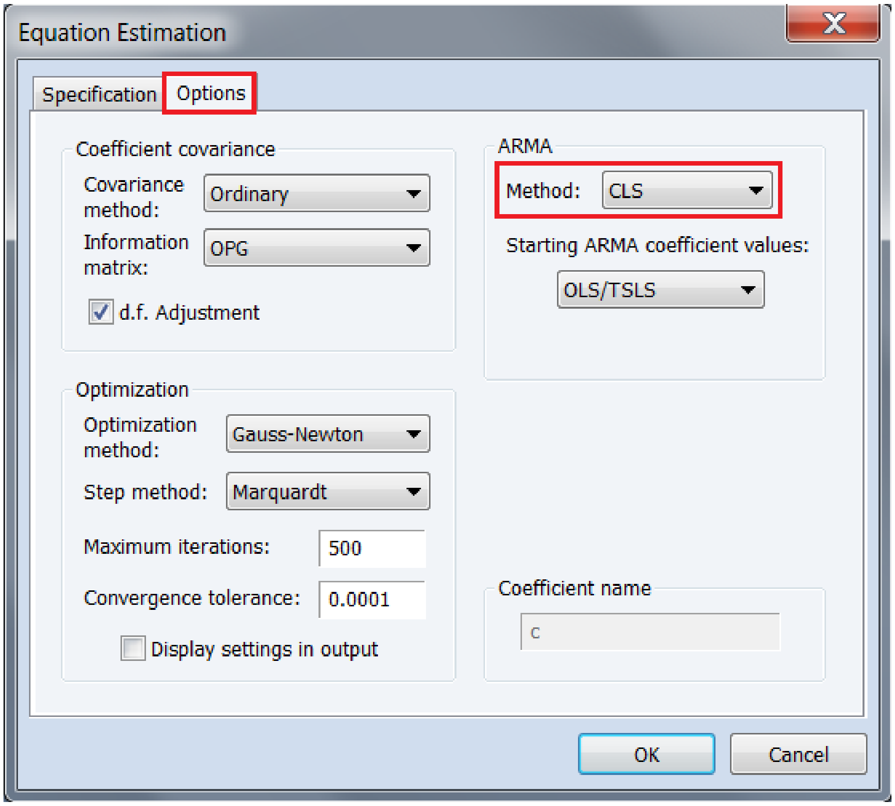
$$Quick \rightarrow Estimate\ Equation$$



then under the *Options* tab,

$$ARMA\ Method : CLS$$

A note from the Tutorial 10 Section in Moodle: Different versions of EViews have different default settings for FGLS estimation of a model with AR errors. Therefore, you may get results that are slightly different from the results given in the answer key depending on the version of EViews that you are using. However, the qualitative conclusions will be the same even if the estimates slightly differ numerically.


Equation Estimation
Specification
Options
Coefficient covariance
Covariance method:
Ordinary
Information matrix:
OPG
d.f. Adjustment
ARMA
Method:
CLS
Starting ARMA coefficient values:
OLS/TSLS
Optimization
Optimization method:
Gauss-Newton
Step method:
Marquardt
Maximum iterations:
500
Convergence tolerance:
0.0001
Display settings in output
Coefficient name
c
OK
Cancel

Dependent Variable: AVELOAD
Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps)
Sample (adjusted): 10/02/2005 3/31/2006
Included observations: 181 after adjustments
Convergence achieved after 12 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3582.213	85.97658	41.66498	0.0000
AVETEMP	−26.35015	1.466074	−17.97327	0.0000
WKND	−146.2892	10.02363	−14.59443	0.0000
PUBHOL	−48.19995	19.12197	−2.520658	0.0126
AR(1)	0.914550	0.030277	30.20613	0.0000

R-squared	0.965152	Mean dependent var	2390.249
Adjusted R-squared	0.964360	S.D. dependent var	362.3839
S.E. of regression	68.41268	Akaike info criterion	11.31623
Sum squared resid	823732.0	Schwarz criterion	11.40459
Log likelihood	−1019.119	Hannan-Quinn criter.	11.35205
F-statistic	1218.633	Durbin-Watson stat	2.183122
Prob(F-statistic)	0.000000		

Inverted AR Roots	.91

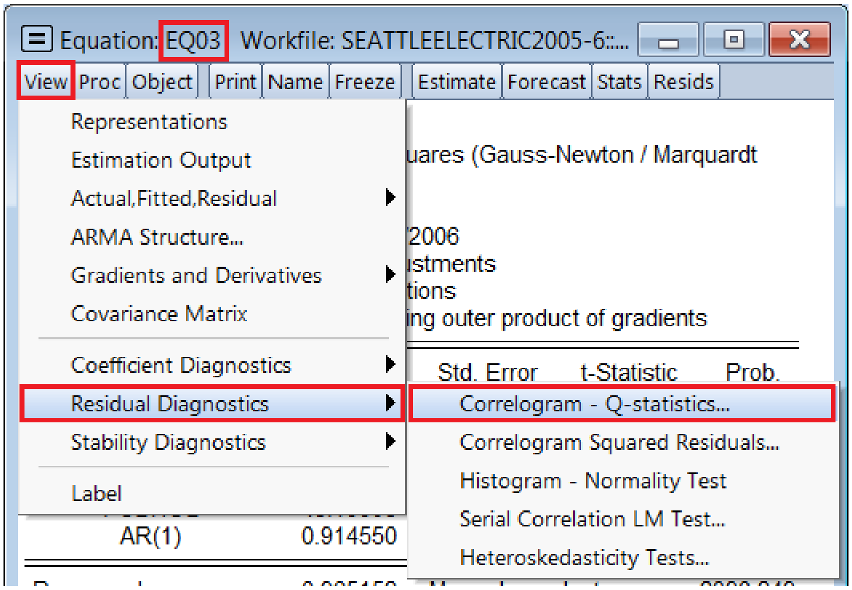
$$aveload_t = \underset{(85.9766)}{3582.213} - \underset{(1.4661)}{26.3502} avetemp_t - \underset{(10.0236)}{146.2892} wknd_t - \underset{(19.1220)}{48.2000} pubhol_t + \hat{u}_t$$

$$\hat{u}_t = \underset{(0.0303)}{0.9146} \hat{u}_{t-1} + \hat{e}_t \qquad R^2 = 0.9652$$

Note the way that the estimated model is reported when incorporating AR errors.

To obtain the correlogram of the residual of the estimated *aveload* model with AR(1) errors (I've named this equation eq03),

$$eq03 \rightarrow View \rightarrow Residual\ Diagnostics \rightarrow Correlogram\ Q - statistics$$



Lags to include : 36

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob*
		1	-0.117	-0.117	2.5324	
		2	-0.006	-0.020	2.5383	0.111
		3	-0.025	-0.028	2.6530	0.265
		4	-0.018	-0.025	2.7140	0.438
		5	0.087	0.082	4.1232	0.390
		6	-0.013	0.006	4.1545	0.527
		7	-0.056	-0.057	4.7558	0.575
		8	-0.082	-0.093	6.0298	0.536
		9	0.060	0.042	6.7204	0.567
		10	-0.000	0.000	6.7204	0.666
		11	0.105	0.105	8.8875	0.543
		12	0.061	0.100	9.6243	0.564

We can check for evidence of serial correlation in the errors by inspecting the correlogram from the residuals. If the sample autocorrelation coefficients are within the significance bands, then they are not statistically significantly different from 0, which would suggest that there is insufficient evidence of serial correlation in the errors.

As we can see from the correlogram of the residuals, all the sample autocorrelation coefficients lie within the significance bands $\therefore$ there is insufficient evidence to suggest that the errors are serially correlated so the regression with AR(1) errors has taken care of serial correlation in the errors.

This means that the model of *aveload* with AR(1) errors does not have serially correlated errors.

Notes about the estimates of the parameters in the *aveload* model with AR(1) errors:

- The feasible generalised least squares (FGLS) estimator is used to estimate this parameters.
- Since the parameters are non-linear (see page 20-22 of Week 9 lecture notes for derivation of model), the FGLS is not unbiased.
- FGLS is consistent which means that as the sample size increases, the FGLS estimator converges to the parameter (it converges to the truth).
- In large samples, the FGLS estimator is more efficient that the OLS estimator.
- Hypothesis tests based on the FGLS standard errors are valid when the sample is large (i.e. asymptotically valid).

(e) Using the regression model with AR errors, investigate if the sensitivity of electricity load to temperature (i.e. the coefficient of temperature) is different in weekends relative to the rest of the week.

If we suspect that the sensitivity of electricity load to temperature is different in weekends relative to the rest of the week then we should include the interaction term

$$wknd_t \times avetemp_t$$

in our regression model of $aveload$,

$$aveload_t = \beta_0 + \beta_1 avetemp_t + \beta_2 wknd_t + \beta_3 pubhol_t + \beta_4 wknd_t \times avetemp_t + u_t$$

Taking the partial derivative of $aveload_t$ with respect to $avetemp_t$ gives the partial effect of $avetemp_t$ on $aveload_t$:

$$\frac{\partial aveload_t}{\partial avetemp_t} = \beta_1 + \beta_4 wknd_t$$

that is, it is an equation that describes the effects of temperature on electricity load holding the other variables constant.

The effect on electricity load for a 1 Fahrenheit degree increase in temperature during the weekend (and holding public holiday constant) is given by,

$$\beta_1 + \beta_4$$

while the effect on electricty load for a 1 Fahrenheit degree increase in temperature during the weekday (and holding public holiday constant) is given by,

$$\beta_1$$

As such, if the effect of temperature on electricity is not different between weekends and weekdays then the coefficient β_4 equals to 0,

$$H_0 : \beta_4 = 0$$

but if there is a difference then,

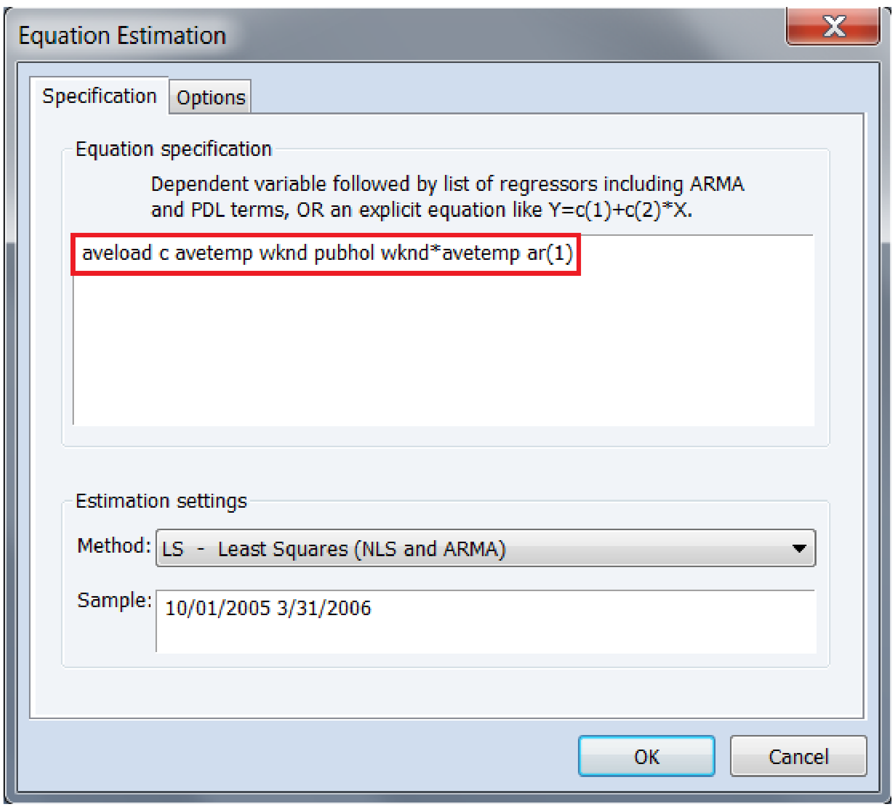
$$H_1 : \beta_4 \neq 0$$

To estimate model of *aveload* with an AR(1) error and with the interaction term $wknd_t \times pubhol_t$

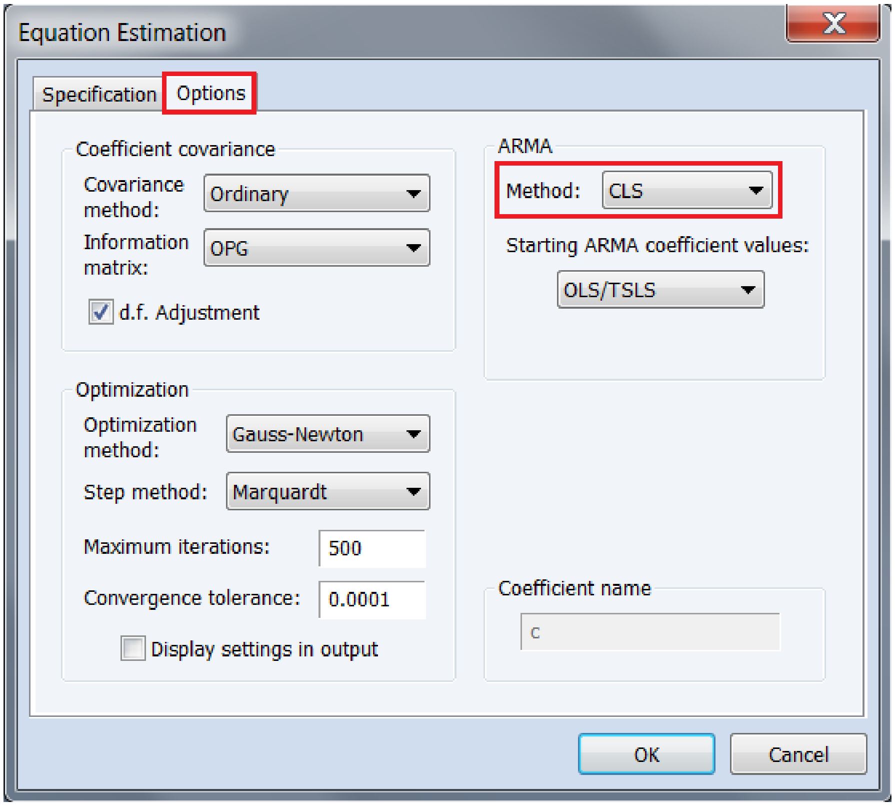
$$\begin{aligned} aveload_t &= \beta_0 + \beta_1 avetemp_t + \beta_2 wknd_t + \beta_3 pubhol_t + \beta_4 wknd_t \times pubhol_t + u_t \\ u_t &= \rho_1 u_{t-1} + e_t \end{aligned}$$

in EViews,

$$Quick \to Estimate\ Equation$$



then under the *Options* tab, choose $ARMA\ Method : CLS$



A note from the Tutorial 9 tab in Moodle: Different versions of EViews have different default settings for FGLS estimation of a model with AR errors. Therefore, you may get results that are slightly different from the results given in the answer key depending on the version of EViews that you are using. However, the qualitative conclusions will be the same even if the estimates slightly differ numerically.

Dependent Variable: AVELOAD
Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps)
Date: 10/01/18 Time: 22:00
Sample (adjusted): 10/02/2005 3/31/2006
Included observations: 181 after adjustments
Convergence achieved after 13 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3575.808	87.10534	41.05153	0.0000
AVETEMP	−26.21469	1.494040	−17.54617	0.0000
WKND	−118.7532	56.30826	−2.108984	0.0364
PUBHOL	−48.73944	19.19379	−2.539333	0.0120
WKND*AVETEMP	−0.632066	1.271752	−0.497004	0.6198
AR(1)	0.914551	0.030417	30.06718	0.0000

R-squared	0.965201	Mean dependent var	2390.249
Adjusted R-squared	0.964207	S.D. dependent var	362.3839
S.E. of regression	68.55950	Akaike info criterion	11.32587
Sum squared resid	822570.9	Schwarz criterion	11.43190
Log likelihood	−1018.991	Hannan-Quinn criter.	11.36885
F-statistic	970.7848	Durbin-Watson stat	2.186007
Prob(F-statistic)	0.000000		
Inverted AR Roots	.91		

$$
\begin{aligned}
aveload_t &= \underset{(87.1053)}{3575.808} - \underset{(1.4940)}{26.32147} avetemp_t - \underset{(56.3083)}{118.7532} wknd_t - \underset{(19.1938)}{48.7394} pubhol_t \\
&\quad - \underset{(1.2718)}{0.6321} wknd \times avetemp_t + \hat{u}_t \\
\hat{u}_t &= \underset{(0.0304)}{0.9146} \hat{u}_{t-1} + \hat{e}_t \qquad R^2 = 0.9652
\end{aligned}
$$

Note the way that the estimated model is reported when incorporating AR errors.

$$
\begin{aligned}
H_0 &: \beta_4 = 0 \\
H_1 &: \beta_4 \neq 0
\end{aligned}
$$

$$p - value = 0.6198$$

There is insufficient evidence from our sample to suggest that the sensitivity electricity load to temperature differs between weekends and weekdays.

(f) Using the regression model with AR errors, investigate if having dummies for Saturday and Sunday separately instead of *wknd* would improve the model. In other words, test the hypothesis $H_0 : \beta_{SAT} = \beta_{SUN}$.

$$\begin{aligned} aveload_t &= \beta_0 + \beta_1 avetemp_t + \beta_2 pubhol_t + \beta_{SAT} sat_t + \beta_{SUN} sun_t + u_t \\ u_t &= \rho u_{t-1} + e_t \end{aligned}$$

$$OR$$

$$\begin{aligned} aveload_t &= \beta_0 + \beta_1 avetemp_t + \beta_2 pubhol_t + \beta_{WKND} wknd_t \\ u_t &= \rho u_{t-1} + e_t \end{aligned}$$

The null hypothesis that $\beta_{SAT} = \beta_{SUN}$ suggests that the difference in average electricity usage between Saturday the weekdays is the same as the difference in average electricity usage between Sunday and the weekdays (since Sat and Sun dummies are included as regressors, weekday is the base dummy variable) i.e. that on either days of the weekend average electricity usage differs from the weekday by the same amount.

Imposing the restriction,

$$\beta_{SAT} = \beta_{SUN}$$

we have,

$$\begin{aligned} aveload_t &= \beta_0 + \beta_1 avetemp_t + \beta_2 pubhol_t + \beta_{SAT} sat_t + \boldsymbol{\beta_{SAT}} sun_t + u_t \\ &= \beta_0 + \beta_1 avetemp_t + \beta_2 pubhol_t + \boldsymbol{\beta_{SAT}}(sat_t + sun_t) + u_t \\ &= \beta_0 + \beta_1 avetemp_t + \beta_2 pubhol_t + \boldsymbol{\beta_{SAT}}(wknd_t) + u_t \end{aligned}$$

Unrestricted model:

$$aveload_t = \beta_0 + \beta_1 avetemp_t + \beta_2 pubhol_t + \beta_{SAT} sat_t + \beta_{SUN} sun_t + u_t$$

Restricted model:

$$aveload_t = \beta_0 + \beta_1 avetemp_t + \beta_2 pubhol_t + \boldsymbol{\beta_{SAT}} wknd_t + u_t$$

$$H_0 : \beta_{SAT} = \beta_{SUN} \quad (restricted\ model)$$

$$H_1 : \beta_{SAT} \neq \beta_{SUN} \quad (unrestricted\ model)$$

To estimate the unrestricted model in EViews,

$$aveload\ c\ avetemp\ pubhol\ sat\ sun\ AR(1)$$

$$ARMA\ Method : CLS$$

Dependent Variable: AVELOAD
Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps)
Sample (adjusted): 10/02/2005 3/31/2006
Included observations: 181 after adjustments
Convergence achieved after 12 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3574.174	86.51799	41.31134	0.0000
AVETEMP	−26.14050	1.441350	−18.13612	0.0000
PUBHOL	−43.72776	18.83209	−2.321981	0.0214
SAT	−130.6781	11.43118	−11.43172	0.0000
SUN	−160.9819	11.25608	−14.30177	0.0000
AR(1)	0.918613	0.029535	31.10215	0.0000

R-squared	0.966522	Mean dependent var	2390.249
Adjusted R-squared	0.965565	S.D. dependent var	362.3839
S.E. of regression	67.24636	Akaike info criterion	11.28719
Sum squared resid	791362.9	Schwarz criterion	11.39322
Log likelihood	−1015.491	Hannan-Quinn criter.	11.33018
F-statistic	1010.449	Durbin-Watson stat	2.114403
Prob(F-statistic)	0.000000		

Inverted AR Roots	.92

$$SSR_{ur} = 791362.9$$

We have estimated the restricted model earlier,

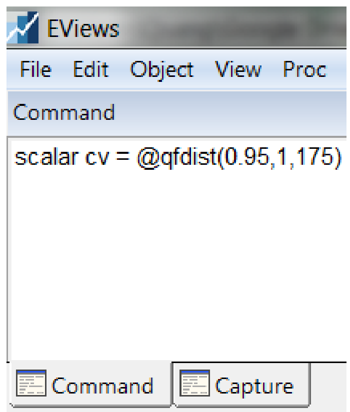
$$SSR_r = 823732.0$$

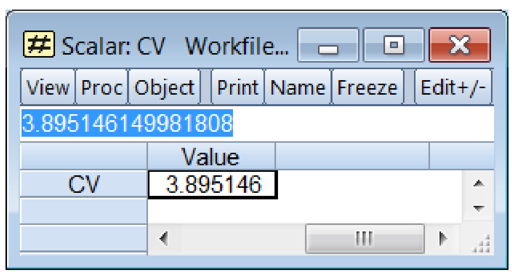
$$F = \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-k-1)} = \frac{(SSR_r - SSR_{ur})/1}{SSR_{ur}/(181-5-1)} \sim F_{1,175} \quad under\ H_0$$

$$F_{calc} = \frac{(823732.0 - 791362.9)/1}{791362.9/(175)} = 7.158$$

To obtain the critical value from the Command window in EViews,

$$scalar\ cv\ = @qfdist(0.95, 1, 175)$$





$$F_{crit} = 3.92$$

Since $F_{calc} = 7.158 > F_{crit} = 3.92$ we reject the null at the 5% significance level and conclude that there is sufficient evidence from our sample in support for the model with separate dummies of Saturday and Sunday.