Introductory Econometrics
Tutorial 10

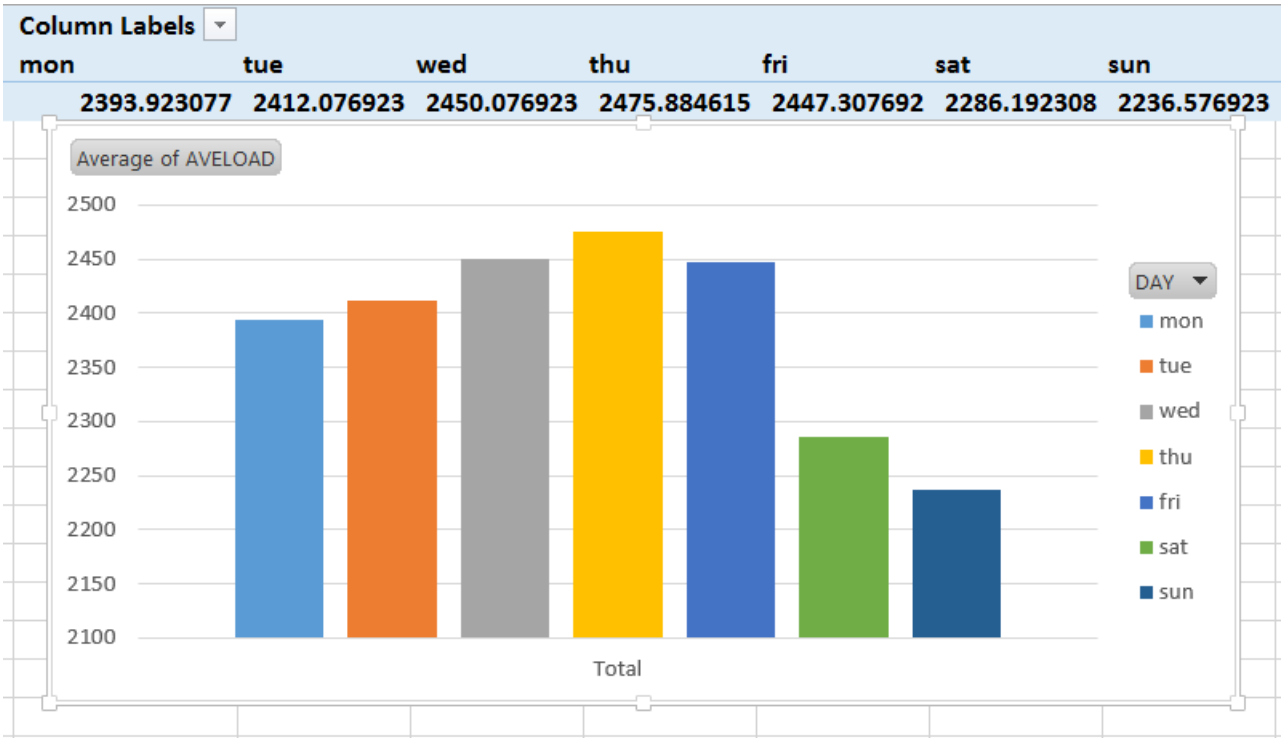
<u>Part B:</u> This part will be covered in the tutorial. It is still a good idea to attempt these questions before the tutorial.

The purpose of this tutorial is to use dummy variables when using time series data.

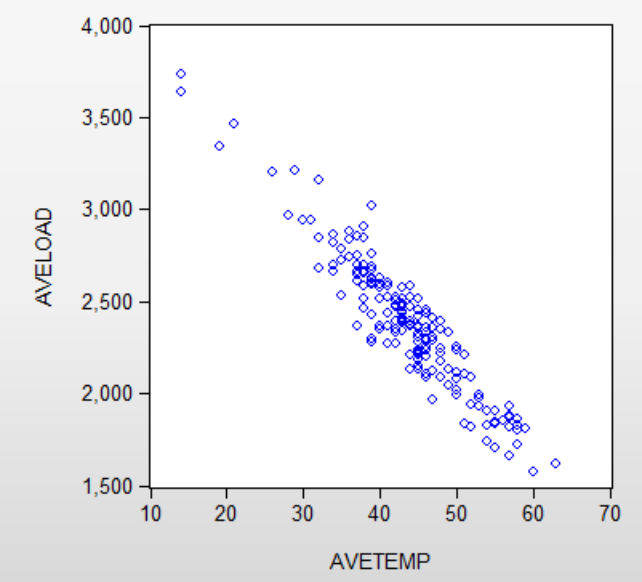
1. (a) *For this tutorial, I recommend that you ask assignment groups to sit together and do this together, and you just walk around and answer their questions. By now, in particular after the first assignment, they should be able to be pretty self-sufficient in transforming variables and running regressions in Eviews. If you feel that many are having similar issues, then interject and present a general explanation to the tutorial class. Just a suggestion, please feel free to ignore if you assess that baby-birds are not ready for unassisted flights yet.*
   In Eviews, "`series sat = (dow=7)`*" and "*`series sun = (dow=1)`*" will create dummy variables for Saturday and Sunday. Then "*`series wknd = sat + sun`*" creates a dummy for weekends. Of course there are several different ways to generate this.*

   (b) *As the question recommends, students are free to use any software that they know to get the best look at their data. Here is what I have done: I used vlookup to change the dow to names for each day, and then used pivot chart and pivot table to get a bar chart for aveload in each day of the week in Excel. I generated the scatter plot in Eviews.*

Column Labels						
mon	tue	wed	thu	fri	sat	sun
2393.923077	2412.076923	2450.076923	2475.884615	2447.307692	2286.192308	2236.576923



   (c) OLS will still be unbiased but no longer BLUE. Also, the OLS standard errors will not be correct and cannot be used for inference. So, the model above cannot be used.

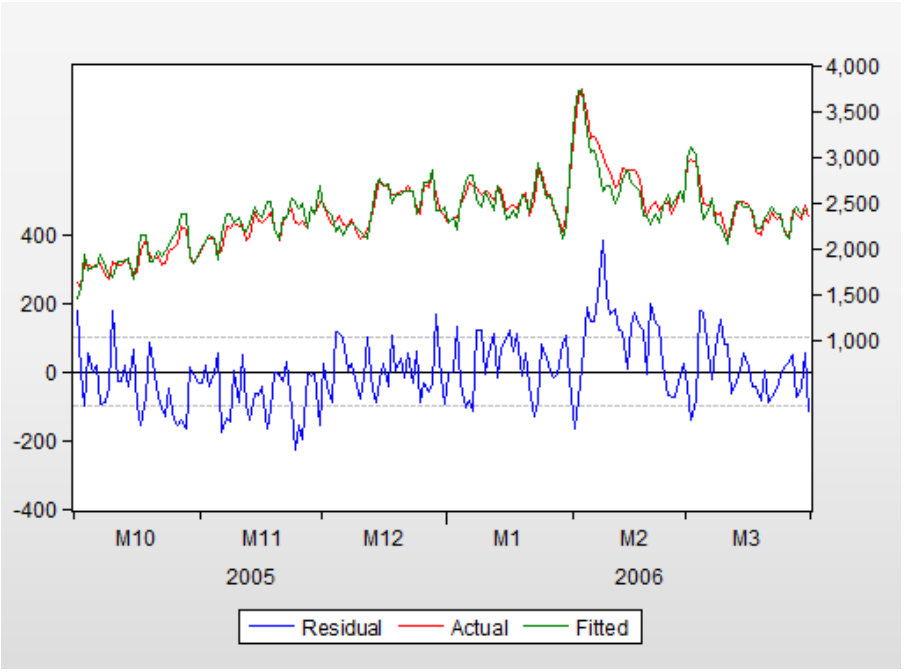


Dependent Variable: AVELOAD
Method: Least Squares
Sample: 10/01/2005 3/31/2006
Included observations: 182

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
WKND	-143.6717	16.35973	-8.782032	0.0000
PUBHOL	-121.5734	31.00930	-3.920546	0.0001

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		

(d)



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

Autocorrelation	Partial Correlation		AC	PAC
		1	0.516	0.516
		2	0.288	0.030
		3	0.220	0.083
		4	0.277	0.175
		5	0.323	0.145
		6	0.237	-0.018
		7	0.116	-0.070
		8	0.058	-0.041
		9	0.193	0.173
		10	0.220	0.028
		11	0.224	0.085
		12	0.161	0.017

$$
\begin{aligned}
aveload_t &= \beta_0 + \beta_1 avetemp_t + \beta_2 wknd_t + \beta_3 pubhol_t + u_t \\
u_t &= \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \rho_4 u_{t-4} + \rho_5 u_{t-5} + \rho_6 u_{t-6} + \rho_7 u_{t-7} + e_t \\
H_0 &: \rho_1 = \rho_2 = \rho_3 = \rho_4 = \rho_5 = \rho_6 = \rho_7 = 0 \\
H_1 &: \text{at least one of the AR parameters is not zero}
\end{aligned}
$$

Estimate of the aux. regression

```
uhat c avetemp wknd pubhol uhat(-1 to -7)
```

$$
\begin{aligned}
R^2_{\hat{u}} &= 0.340, \quad n_{\hat{u}} = n - 7 = 182 - 7 = 175 \\
BG &= n_{\hat{u}} \times R^2_{\hat{u}} \sim \chi^2_7 \quad \text{under } H_0 \\
BG_{calc} &= 175 \times 0.340 = 59.5 > BG_{crit} = 14.07
\end{aligned}
$$

We reject the null and conclude the error are serially correlated

In the BG auxiliary regression, only the first AR term is significant. The correlogram also shows a significant first order autocorrelation, with the 4th partial autocorrelation also close to being significant. We start with adding AR(1) errors and check the residual correlogram again to see if there is a need to consider a longer AR model for the errors. Note that unlike OLS which has an exact analytical formula ($(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$), feasible GLS does not have a fixed formula and relies on minimising sum of squared residuals numerically, and there are numerous algorithms for minimising a function numerically, each of which works best in some situations and fails in others. Therefore, although you get the same answer for OLS estimates no matter which software you choose, you may get different answers from different software when using FGLS. Even different versions of the same software may give you different answers if the default estimation method is different in different versions. The following two tables give the results of the same estimation command in EViews 8 and EViews 10:

Dependent Variable: AVELOAD
Method: Least Squares
Date: 09/23/18 Time: 00:18
Sample (adjusted): 10/02/2005 3/31/2006
Included observations: 181 after adjustments
Convergence achieved after 13 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3582.163	85.98445	41.66059	0.0000
AVETEMP	-26.34896	1.466071	-17.97250	0.0000
WKND	-146.2891	10.02355	-14.59454	0.0000
PUBHOL	-48.19952	19.12181	-2.520656	0.0126
AR(1)	0.914579	0.030275	30.20876	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	2.183140
Prob(F-statistic)	0.000000		

Inverted AR Roots	.91

Dependent Variable: AVELOAD
Method: ARMA Maximum Likelihood (OPG - BHHH)
Sample: 10/01/2005 3/31/2006
Included observations: 182
Convergence achieved after 17 iterations
Coefficient covariance computed using outer product of gradients

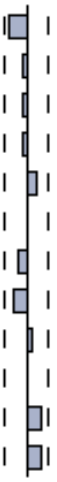
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3576.414	71.92891	49.72151	0.0000
AVETEMP	-26.59983	1.099957	-24.18261	0.0000
WKND	-146.6367	11.46094	-12.79447	0.0000
PUBHOL	-48.23138	15.66494	-3.078938	0.0024
AR(1)	0.912171	0.029491	30.93090	0.0000
SIGMASQ	4545.861	520.2919	8.737135	0.0000

R-squared	0.965856	Mean dependent var	2386.005
Adjusted R-squared	0.964886	S.D. dependent var	365.8870
S.E. of regression	68.56262	Akaike info criterion	11.33559
Sum squared resid	827346.7	Schwarz criterion	11.44121
Log likelihood	-1025.538	Hannan-Quinn criter.	11.37841
F-statistic	995.7266	Durbin-Watson stat	2.186760
Prob(F-statistic)	0.000000		

Inverted AR Roots	.91

Note that the results are close, but are not the same. In fact version 10 does not use FGLS as its default, but it uses the maximum likelihood estimation method (you will be introduced to maximum likelihood in ETC3400 and ETC3410). Note that maximum likelihood uses all 182 observations. Since inference with time series is all based on large samples, whether a method uses all observations or all observations minus 1 does not matter at all.
Residuals are now white noise. So the regression with AR(1) errors has taken care of serial

Sample: 10/02/2005 3/31/2006
Included observations: 181
Q-statistic probabilities adjusted for 1 ARMA term(s)

Autocorrelation	Partial Correlation		AC	PAC
		1	-0.117	-0.117
		2	-0.006	-0.020
		3	-0.025	-0.028
		4	-0.018	-0.025
		5	0.086	0.082
		6	-0.013	0.006
		7	-0.056	-0.057
		8	-0.082	-0.093
		9	0.060	0.042
		10	-0.000	0.000
		11	0.105	0.105
		12	0.061	0.100

correlation in errors. Compare these results with OLS results that we started with.