ECON 546: Spring 2010
Solution to Assignment 3

Q.1.(a) By independence –

$$L = c^n \prod_i x_i^{-(c+1)}$$

$$\log L_u = n\log(c) - (c+1)\sum_i \log(x_i)$$

$$\log L_R = n\log(2) - (2+1)\sum_i \log(x_i)$$

$$= n\log(2) - 3\sum_i \log(x_i)$$

$$(\partial \log L / \partial c) = n/c - \sum_i \log(x_i) = 0$$

$$\Rightarrow \tilde{c} = [n / \sum_i \log(x_i)]$$

$$(\partial^2 \log L / \partial c^2) = -n/c^2 < 0 \text{ (Everywhere)}$$

$$\Rightarrow \text{max.}$$

So, $LRT = 2[\log \tilde{L}_u - \log \tilde{L}_R]$

$$= 2\{[n\log(\tilde{c}) - (\tilde{c}+1)\sum_i \log(x_i)] - [n\log(2) - 3\sum_i \log(x_i)]\}$$

$$= 2\left\{n\log\left[\frac{n}{\sum_i x_i}\right] - \left(n + \sum_i \log(x_i)\right) - n\log(2) + 3\sum_i \log(x_i)\right\}$$

$$= 2 \left\{ n \left[\log\left(\frac{n}{\sum_i x_i}\right) - 1 - \log(2) \right] + 2 \sum_i \log(x_i) \right\}$$

$$= 2 \left\{ n \left[\log\left(n / \sum_i x_i \right) - 1.693148 \right] + 2 \sum_i \log(x_i) \right\}$$

(b) $\quad I = -E\left[\partial^2 \log L / \partial c^2 \right] = n/c^2$

$$IA = \lim_{n \to \infty} \left[\frac{1}{n} I \right] = 1/c^2$$

$$I^* = (n/\tilde{c}^2) \qquad \left[\text{So that } \operatorname{plim}\left(\tfrac{1}{n} I^*\right) = IA \right]$$

& $\text{Wald} = n(\tilde{c} - 2)^2 / \tilde{c}^2$

$$= n\left[1 - \frac{2}{\tilde{c}}\right]^2 = n\left[1 - \frac{2\sum_i \log x_i}{n}\right]^2$$

(c) $\quad LM = \left(D \log L_u \big|_{\tilde{c}_0} \right)^2 I^*(\tilde{c}_0)^{-1}$

where $\tilde{c}_0 = 2$.

So, $LM = 4\left[\frac{n}{2} - \sum_i \log(x_i)\right]^2 / n = W$!

(d) $\quad n = 200 \ ; \quad \prod_i x_i = 1.2204032 \times 10^{39}$

So, $\sum_i \log(x_i) = \log \prod_i x_i = 90$

Then, $LRT = 2.14388$

$\text{Wald} = 2.0$

$LM = 2.0$

Each statistic $\xrightarrow{d} \chi^2_{(1)}$. 5% critical value is 3.84, so in each case, DO NOT REJECT H_0.

QUESTION 2 (a)

```
-----------------------
'INITIALIZE VARIOUS VALUES
'-------------------------
scalar nrep=5000
'CC DENOTES THE TRUE VALUE OF THE PARAMETER, "C"
'IN THE NEXT LINE OF CODE WE IMPOSE THE NULL HYPOTHESIS TO BE TRUE, SO IN
THIS CASE THE "POWERS" OF THE TESTS THAT ARE CALCULATED BELOW ARE
ACTUALLY THE TRUE SIGNIFICANCE LEVELS FOR THIS SAMPLE SIZE

'IF WE ALTER "CC" TO OTHER VALUES, THE NULL WILL BE FALSE & THEN THE POWER
VALUES REALLY WILL BE POWERS, ALTHOUGH ONES THAT ARE NOT ADJUSTED FOR
ANY SIZE-DISTORTION
scalar cc=2
scalar n=10
scalar sumlrt=0
scalar sumw=0
scalar sumlm=0
rndseed 121212
series(nrep)  y
'CALCULATE 5% CRITICAL VALUE FOR CHI-SQUARE WITH 1 DEGREE OF FREEDOM:
'-----------------------------------------------------------------
scalar crit=@qchisq(0.95,1)
smpl 1 10

'START THE MONTE CARLO LOOP
'NOTE: THE PARETO RANDOM NUMBER GENERATOR IN EVIEWS APPEARS TO WORK
FINE FOR SOME CHOICES OF "C", BUT NOT FOR OTHERS - IF C=2, FOR EXAMPLE,
THERE IS A TENDENCY TO PRODUCE LARGE OUTLIERS, WHICH THEN MESS UP ALL OF
THE CALCULATIONS AND RESULTS.
'SO, GENERATE THE PARETO RANDOM VALUES USING THE RESULT THE (1/U)^(1/C),
WHERE U IS UNiFORM ON (0,1), WILL BE PARETO.
'--------------------------
for !i=1 to nrep
y=(1/@runif(0,1))^(1/cc)
scalar sumly=@sum(log(y))
scalar mle=n/sumly
scalar lrt=2*(n*(log(mle)-1.693148)+2*sumly)
scalar wald=n*(mle-2)^2/(mle^2)
scalar lm=(4/n)*((n/2)-sumly)^2
if(lm>crit) then
sumlm=sumlm+1
endif
if(lrt>crit) then
sumlrt=sumlrt+1
endif
if(wald>crit) then
sumw=sumw+1
endif

next

'END OF MONTE CARLO LOOP
scalar power_lrt=sumlrt/nrep
scalar power_w=sumw/nrep
scalar power_lm=sumlm/nrep
```

(b)	Sizes when n=10	n=50	n=200	n=500
LRT	0.0484	0.0500	0.0518	0.0518
LM	0.0420	0.0468	0.0516	0.0498
Wald	0.0420	0.0468	0.0516	0.0498

The LRT has the least size distortion for small sample sizes. When n > 50 the other two tests have slightly less size distortion. **NOTE THAT IN FACT THE WALD AND LM TEST STATISTICS ARE IDENTICAL FOR THIS PARTICULAR PROBLEM!**

(c) Some illustrative "powers" are given below. The asymptotic critical values have been used – so these powers are not "size-adjusted". We can still see the following, though:

(i) True size of LRT exceeds those of LM and Wald when n = 10 (see above), and yet when c < 2, the "raw" (unadjusted) powers of LM and Wald exceed LRT. So LM and Wald definitely have greater power than LRT in this situation.

(ii) Conversely, when c > 2, the "raw" (unadjusted) powers of LM and Wald are less than that of the LRT. So in this case we can't say which test has the greater (size-adjusted) power, strictly speaking. However, given the very slight size distortions and the big raw power differences, the LRT is probably the superior test in this case.

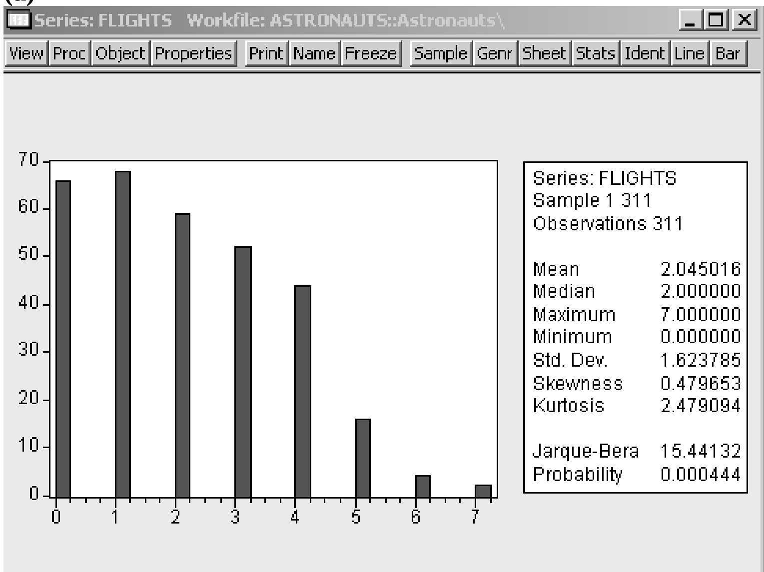
"Powers" when n=10 (not size-adjusted):

c	LRT	Wald	LM
1.6	0.1094	0.1662	0.1662
1.4	0.2198	0.2988	0.2988
1.0	0.6194	0.7046	0.7046
0.6	0.9614	0.9732	0.9732
0.4	0.9968	0.9982	0.9982
2.4	0.0880	0.0258	0.0258
2.8	0.1682	0.0418	0.0418
3.2	0.2870	0.0896	0.0896
3.6	0.4132	0.1524	0.1524
4.0	0.5422	0.2316	0.2316
4.8	0.7636	0.4274	0.4274
6.0	0.9338	0.7058	0.7058
8.0	0.9940	0.9396	0.9396

(d) Clearly, we need to "size-adjust" the tests. For each test separately, find the actual critical value (by trial and error) that ensures that the rejection rate is exactly 5% when n=10 and c=2. Then use *these* critical values with different values of c to get the appropriate rejection rates (powers) of the tests.

QUESTION 3

(a)



Clearly the variance exceeds the mean – there is over-dispersion – perhaps the Poisson model is not the best choice. Let's see.

(b) The following is a reasonable model – you may have found a better one:

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: FLIGHTS
Method: ML/QML - Poisson Count (Quadratic hill climbing)
Date: 03/18/08 Time: 09:20
Sample: 1 311
Included observations: 310
Convergence achieved after 8 iterations
QML (Huber/White) standard errors & covariance

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.616915	0.083739	7.367157	0.0000
SELECTED*MASTER	-0.018627	0.004881	-3.816373	0.0001
MASTER	36.70969	9.672076	3.795430	0.0001
DECEASED	-0.757208	0.150722	-5.023886	0.0000
MILITARY	0.258431	0.095619	2.702729	0.0069
CSMH	0.449001	0.139126	3.227290	0.0012
EVA	0.021593	0.002556	8.449300	0.0000

R-squared	0.201108	Mean dependent var	2.045161
Adjusted R-squared	0.185289	S.D. dependent var	1.626408
S.E. of regression	1.468018	Akaike info criterion	3.475896
Sum squared resid	652.9882	Schwarz criterion	3.560270
Log likelihood	-531.7639	Hannan-Quinn criter.	3.509625
Restr. log likelihood	-573.5634	Avg. log likelihood	-1.715367
LR statistic (6 df)	83.59901	LR index (Pseudo-R2)	0.072877
Probability(LR stat)	6.66E-16		

(c) Military background is significant and has a positive marginal effect. Date of birth (BORN) and gender are not significant. Academic achievement is significant if the astronaut has a Masters degree rather than a Bachelors of Doctorate. The Masters effect impacts on both the intercept and also on the marginal effect of the year of selection – the latter is significant only if the astronaut has a Masters, and not otherwise.

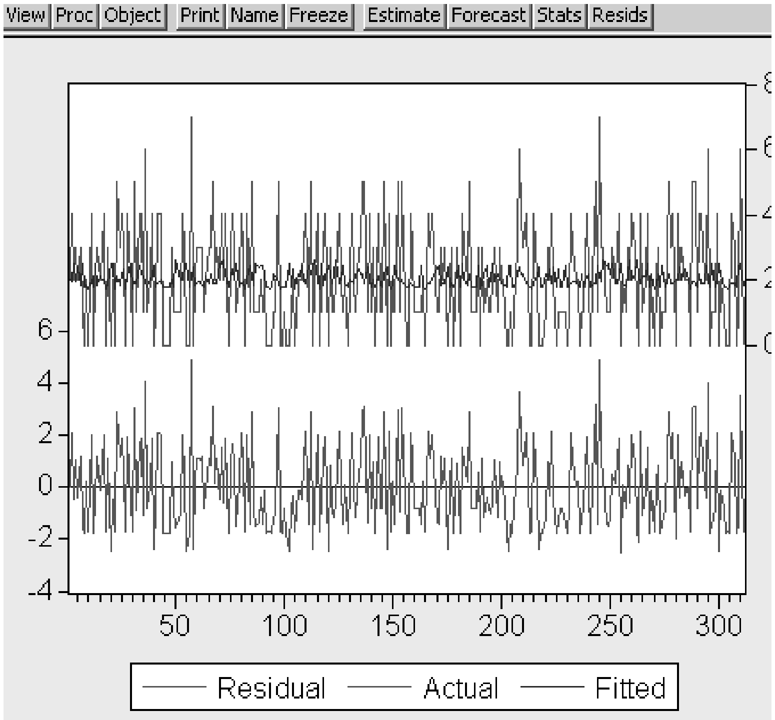
(d)

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: FLIGHTS
Method: ML/QML - Poisson Count (Quadratic hill climbing)
Date: 03/18/08 Time: 09:15
Sample: 1 311
Included observations: 311
Convergence achieved after 8 iterations
QML (Huber/White) standard errors & covariance

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	20.31522	6.407385	3.170594	0.0015
SELECTED	-0.009884	0.003235	-3.055659	0.0022

R-squared	0.016077	Mean dependent var	2.045016
Adjusted R-squared	0.012893	S.D. dependent var	1.623785
S.E. of regression	1.613283	Akaike info criterion	3.677581
Sum squared resid	804.2286	Schwarz criterion	3.701631
Log likelihood	-569.8638	Hannan-Quinn criter.	3.687194
Restr. log likelihood	-574.8707	Avg. log likelihood	-1.832360
LR statistic (1 df)	10.01385	LR index (Pseudo-R2)	0.008710
Probability(LR stat)	0.001554		



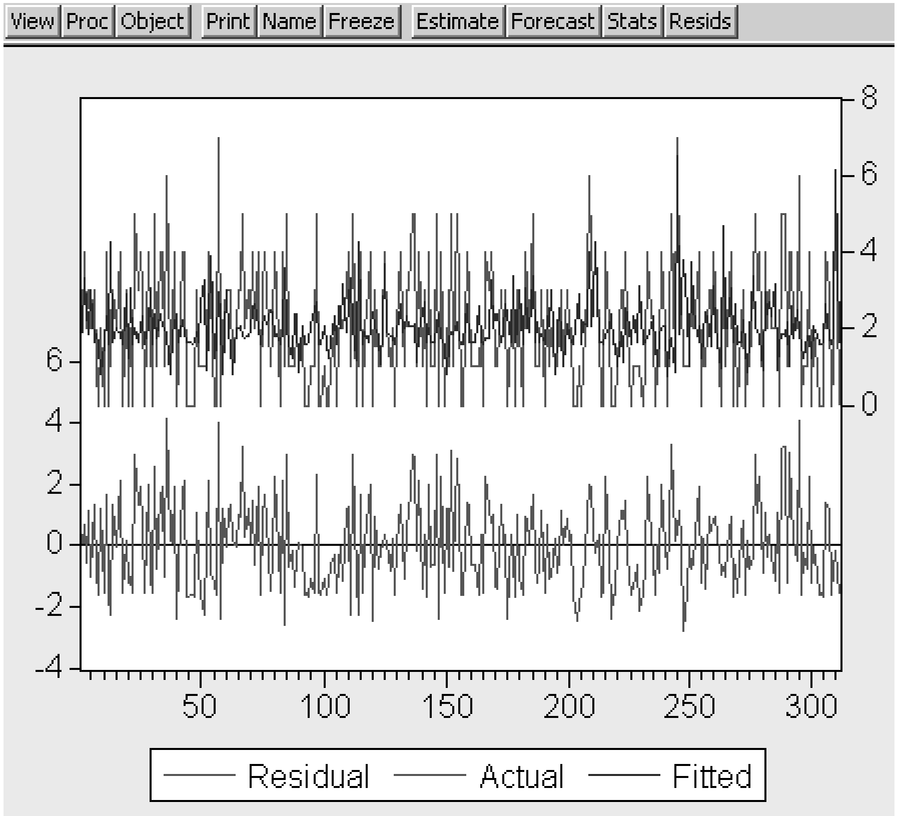
Model does not fit very well – does not explain the variation in the data.

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: FLIGHTS
Method: ML/QML - Poisson Count (Quadratic hill climbing)
Date: 03/18/08 Time: 09:18
Sample: 1 311
Included observations: 310
Convergence achieved after 8 iterations
QML (Huber/White) standard errors & covariance

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	25.49079	6.726606	3.789548	0.0002
SELECTED	-0.012520	0.003393	-3.690369	0.0002
CSMH	0.484414	0.143501	3.375695	0.0007
MOONC	-0.288915	0.134803	-2.143234	0.0321
DECEASED	-0.729454	0.150855	-4.835459	0.0000
EVA	0.020268	0.002507	8.084374	0.0000

R-squared	0.182686	Mean dependent var	2.045161
Adjusted R-squared	0.169244	S.D. dependent var	1.626408
S.E. of regression	1.482403	Akaike info criterion	3.503677
Sum squared resid	668.0459	Schwarz criterion	3.575998
Log likelihood	-537.0700	Hannan-Quinn criter.	3.532588
Restr. log likelihood	-573.5634	Avg. log likelihood	-1.732484
LR statistic (5 df)	72.98684	LR index (Pseudo-R2)	0.063626



This is much better. Note that by the SC values, this model is not as good as the one in part (b) above.

(e) Estimate the corresponding Negative Binomial model and apply Wald test:

Wald Test:
Equation: EQ01

Test Statistic	Value	df	Probability
F-statistic	0.550059	(1, 303)	0.4589
Chi-square	0.550059	1	0.4583

Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
EXP(C(7))	0.033778	0.045544

Delta method computed using analytic derivatives.

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

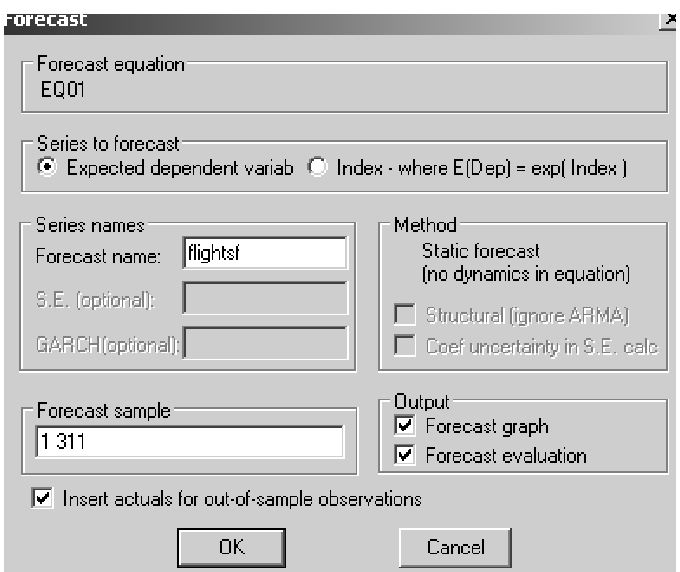
Dependent Variable: FLIGHTS
Method: ML - Negative Binomial Count (Quadratic hill climbing)
Date: 06/18/04 Time: 15:49
Sample: 1 311
Included observations: 310
Excluded observations: 1
Convergence achieved after 9 iterations
QML (Huber/White) standard errors & covariance

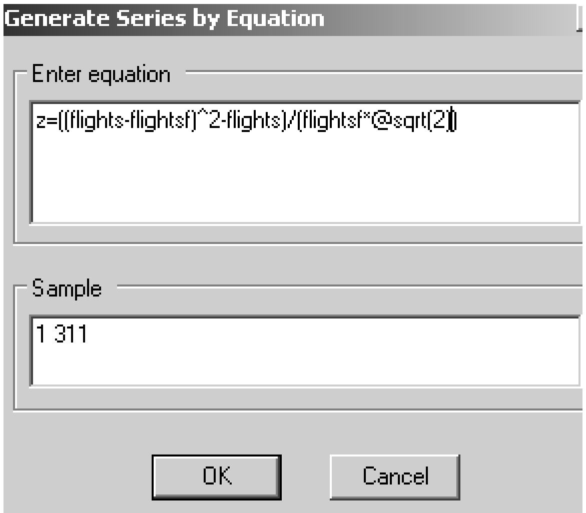
	Coefficient	Std. Error	z-Statistic	Prob.
C	26.28630	7.167289	3.667538	0.0002
CSMH	0.484820	0.144979	3.344080	0.0008
MOONC	-0.297669	0.136000	-2.188733	0.0286
SELECTED	-0.012921	0.003615	-3.573835	0.0004
DECEASED	-0.733482	0.151823	-4.831170	0.0000
EVA	0.020284	0.002579	7.864618	0.0000
Mixture Parameter				
SHAPE:C(7)	-3.387936	1.348327	-2.512696	0.0120

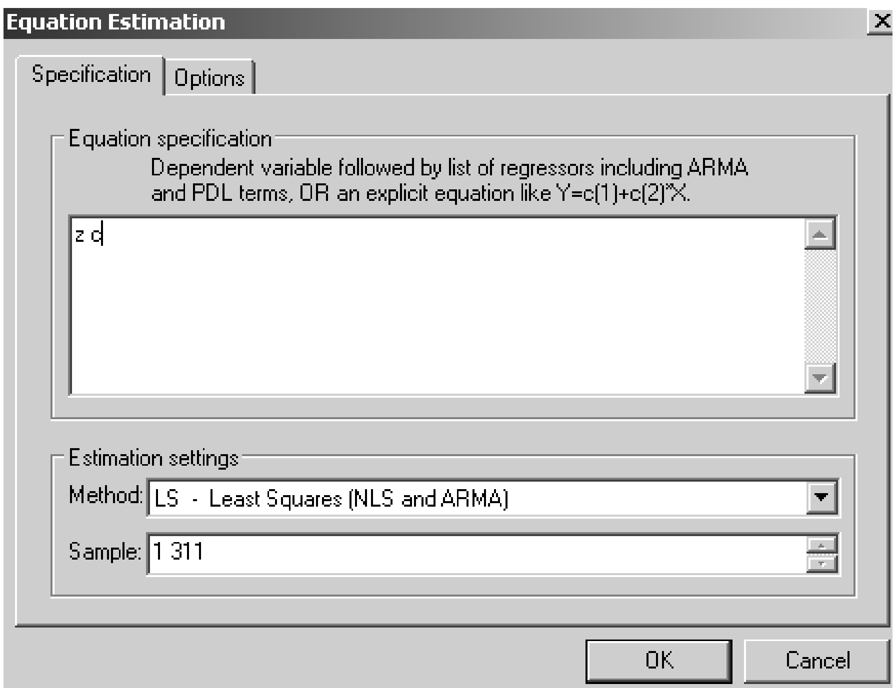
R-squared	0.181493	Mean dependent var	2.045161
Adjusted R-squared	0.165285	S.D. dependent var	1.626408
S.E. of regression	1.485931	Akaike info criterion	3.508366
Sum squared resid	669.0208	Schwarz criterion	3.592741
Log likelihood	-536.7968	Hannan-Quinn criter.	3.542096
Restr. log likelihood	-573.5634	Avg. log likelihood	-1.731603
LR statistic (6 df)	73.53322	LR index (Pseudo-R2)	0.064102
Probability(LR stat)	7.69E-14		

Cannot reject the null that the restriction holds– the test supports the Poisson model.

(f) From the Poisson model –




Generate Series by Equation
Enter equation
z=((flights-flightsf)^2-flights)/(flightsf*@sqrt(2))
Sample
1 311
OK
Cancel


Equation Estimation
Specification
Options
Equation specification
Dependent variable followed by list of regressors including ARMA and PDL terms, OR an explicit equation like Y=c(1)+c(2)*X.
z c
Estimation settings
Method: LS - Least Squares (NLS and ARMA)
Sample: 1 311
OK
Cancel

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: Z
Method: Least Squares
Date: 03/18/08 Time: 09:29
Sample: 1 311
Included observations: 310

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.081785	0.050531	1.618530	0.1066

R-squared	0.000000	Mean dependent var	0.081785
Adjusted R-squared	0.000000	S.D. dependent var	0.889682
S.E. of regression	0.889682	Akaike info criterion	2.607316
Sum squared resid	244.5843	Schwarz criterion	2.619370
Log likelihood	-403.1340	Durbin-Watson stat	1.860553

Equation: UNTITLED Workfile: ASTRONAUTS::Astronauts\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: Z
Method: Least Squares
Date: 03/18/08 Time: 09:30
Sample: 1 311
Included observations: 310

Variable	Coefficient	Std. Error	t-Statistic	Prob.
FLIGHTSF	0.016451	0.023336	0.704929	0.4814

R-squared	-0.006859	Mean dependent var	0.081785
Adjusted R-squared	-0.006859	S.D. dependent var	0.889682
S.E. of regression	0.892728	Akaike info criterion	2.614152
Sum squared resid	246.2618	Schwarz criterion	2.626205
Log likelihood	-404.1935	Durbin-Watson stat	1.853656

In each case, we cannot reject the null that the coefficient is zero (at the 10% level) – this again supports the Poisson model.

(g)

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: FLIGHTS
Method: ML/QML - Poisson Count (Quadratic hill climbing)
Date: 03/18/08 Time: 09:27
Sample: 1 311
Included observations: 310
Convergence achieved after 8 iterations
QML (Huber/White) standard errors & covariance

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	25.49079	6.726606	3.789548	0.0002
CSMH	0.484414	0.143501	3.375695	0.0007
MOONC	-0.288915	0.134803	-2.143234	0.0321
SELECTED	-0.012520	0.003393	-3.690369	0.0002
DECEASED	-0.729454	0.150855	-4.835459	0.0000
EVA	0.020268	0.002507	8.084374	0.0000

R-squared	0.182686	Mean dependent var	2.045161
Adjusted R-squared	0.169244	S.D. dependent var	1.626408
S.E. of regression	1.482403	Akaike info criterion	3.503677
Sum squared resid	668.0459	Schwarz criterion	3.575998
Log likelihood	-537.0700	Hannan-Quinn criter.	3.532588
Restr. log likelihood	-573.5634	Avg. log likelihood	-1.732484
LR statistic (5 df)	72.98684	LR index (Pseudo-R2)	0.063626
Probability(LR stat)	2.44E-14		

(i) The hours of extra-vehicular activity is a continuous variable. The marginal effect is calculated as series marg_eff_eva=flightsf*c(6)

Series: MARG_EFF_EVA Workfile: ASTRU

View | Proc | Object | Properties | Print | Name | Freeze

	MARG_EF...
Mean	0.041451
Median	0.038533
Maximum	0.132455
Minimum	0.015735
Std. Dev.	0.014890
Skewness	1.925981
Kurtosis	10.61127

A representative m.e. would be 0.04. So, 10 extra hours of EVA implies 4 extra expected flights, other things being equal.

(ii) The medal effect is via a dummy variable – obtain the prediction of conditional mean with this dummy set to unity and then to zero, and take the difference to get the marginal effect. I'll set the other regressors to their sample medians. (There are other options here.)

```
scalar marg_eff_csmh=@exp(c(1)+c(2)*1+c(3)*@median(moonc)+c(4)*@median(selected)
+c(5)*@median(deceased)+c(6)*@median(eva))-
@exp(c(1)+c(2)*0+c(3)*@median(moonc)+c(4)*@median(selected)+
c(5)*@median(deceased)+c(6)*@median(eva))
```

The marginal effect is 1.180856. So, being awarded the medal increases the expected number of flights by 1,2, other things being equal.