HW_5, 2010 Solutions

Question 1.

. clear

. (4 vars, 64 obs pasted into editor)
regress cm flr pgnp tfr

```
      Source |       SS       df       MS              Number of obs =      64
-------------+------------------------------           F(  3,    60) =   59.17
       Model |  271802.616     3  90600.8721           Prob > F      =  0.0000
    Residual |  91875.3836    60  1531.25639           R-squared     =  0.7474
-------------+------------------------------           Adj R-squared =  0.7347
       Total |      363678    63  5772.66667           Root MSE      =  39.131

------------------------------------------------------------------------------
          cm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         flr |  -1.768029   .2480169    -7.13   0.000    -2.264137   -1.271921
        pgnp |  -.0055112   .0018782    -2.93   0.005    -.0092682   -.0017542
         tfr |   12.86864   4.190533     3.07   0.003     4.486323    21.25095
       _cons |   168.3067   32.89166     5.12   0.000     102.5136    234.0998
------------------------------------------------------------------------------
```

SSE $\sum(\hat{Y}-\overline{Y})^2 = 271802.616$

SST= $\sum(Y-\overline{Y})^2 = 363678$

SSR=e'e= 91875.3836

R^2 =1-SSR/SST

. di 271802.616/ 363678
.74737162

Adjusted R-squared=1-(SSR/SST)(n-1)/(n-k)=

. di 1-(91875.3836/ 363678)*((64-1)/(64-4))
.7347402

F-stat=(SSE/SSR)(n-k)/(n-1)=

. di (271802.616/ 91875.3836)*(64-4)/(4-1)
59.167669

Standard error of the regression=SSR/(n-k)

. di 91875.3836/(64-4)
1531.2564

Standard error of each coefficient= $s\sqrt{a_{ii}}$
where a_{ii} is the ith diagonal term from $(X'X)^{-1}$

95% CI for each coefficient:

$$\text{Coefficient value} \pm t_{.025,\,24} \cdot s\sqrt{a_{ii}} = \hat{\beta}_i \pm 2.064 \cdot s.e._i$$

	std. error	95% confidence interval	
flr |	.2480169	-2.264137	-1.271921
pgnp	.0018782	-.0092682	-.0017542
tfr	4.190533	4.486323	21.25095
interc	32.89166	102.5136	234.0998

Notice that none of the confidence intervals contain zero. The reason is that all coefficients are different from zero at a 95% level of confidence.

Question 2,

A: After creating the necessary variables (real and log conversions):
. summm

Variable	Obs	Mean	Std. Dev.	Min	Max
year	28	1993.5	8.225975	1980	2007
gdp	28	7124.579	2893.528	2795.6	12308.3
m2	28	4286.168	2146.37	1600.4	8848.1
cpi	28	143.0571	33.857	82.4	194.6
ltrate	28	8.049286	2.345291	5.34	13.45
tbrate	28	6.023607	2.757683	3.02	14.029
rgdp	28	4778.529	899.0091	3377.409	6324.923
rm2	28	2841.849	795.3379	1931.903	4546.813
ln_rgdp	28	8.454416	.191755	8.124865	8.752254
ln_rm2	28	7.918381	.2579255	7.566261	8.422182
ln_ltrate	28	2.048653	.2702965	1.675226	2.598979
ln_tbrate	28	1.710238	.4065248	1.105257	2.641127

Then, I run the regression:

. regress ln_rm2 ln_rgdp ln_ltrate

Source	SS	df	MS		
Model	1.55442216	2	.77721	mber of obs =	28
Residual	.241767603	25	.0096	F(2, 25) =	80.37
Total	1.79618976	27	.06652	Prob > F =	0.0000
				R-squared =	0.8654
				Adj R-squared =	0.8546
				Root MSE =	.09834

ln_rm2	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]

```
    ln_rgdp |   1.839521   .2203321     8.35   0.000     1.385739    2.293303
  ln_ltrate |   .4987116    .156309     3.19   0.004     .1767871    .8206361
      _cons |  -8.655382    2.15394    -4.02   0.000    -13.09151   -4.219259
------------------------------------------------------------------------------
```

As we can see both the log of real GDP and log of long term interests are statistically significant at the 5% level. In double log form these are the elasticities. Thus, a 1% increase in real GDP leads to a 1.83% increase in M2 and a 1% increase in the long term rate of interest leads to a .498 increase in M2.

B.) Evaluate (using a t-test) the null hypothesis that $\beta_2 = 1$

t*=(1.8395-1.0)/0.2203321=

```
. di   .839521/.2203321
3.8102528
```

Which is significant at the 5% level. Thus, we conclude that B2 n.e. 1.

C) Show that the F-statistic from the command "test ln_rgdp" is equal to the squared t-statistic calculated in b).

```
. test ln_rgdp=1

 ( 1)  ln_rgdp = 1

       F(  1,    25) =   14.52
            Prob > F =    0.0008

. di   (.839521/.2203321)^2
14.518026
```

D): Restricted version of the model: Under the restriction that B2=1, we have what amounts to extra data on the right side or the model. To estimate this model, we must move that data to the left side:

```
. gen newy=0
. replace newy=ln_rm2-ln_rgdp
(28 real changes made)

. regress newy  ln_ltrate

      Source |       SS       df       MS              Number of obs =      28
-------------+------------------------------           F(  1,    26) =    0.15
       Model |  .002249785     1  .002249785           Prob > F      =  0.6988
    Residual |   .38216717    26  .014698737           R-squared     =  0.0059
-------------+------------------------------           Adj R-squared = -0.0324
       Total |  .384416955    27  .014237665           Root MSE      =  .12124

------------------------------------------------------------------------------
        newy |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
```

```
   ln_ltrate |  -.0337713   .0863213    -0.39   0.699    -.2112072    .1436646
       _cons |  -.4668489   .1783204    -2.62   0.015    -.8333917    -.100306
```

E) Using the regression results from **both** a) and d), construct an F-test from equation (6.26) on page 84 of the notes to test the null hypothesis $\beta_2 = 1$. Get your degrees of freedom correct. Verify that this equals the answer in c).

Part A) reports the unrestricted model. Part D) reports the restricted version of the model.

From A) Residual | SSR_u = .241767603
From D) Residual | SSR_r = .38216717

Equation 6.26 in notes shows that:

. di ((.38216717-.241767603)/1)/((.241767603/(28-3)))
14.518029

F) The reason why you need to be careful using 6.28 in a test of this nature is that your restricted model's total sum of squares is not equal to the SST in the unrestricted model.
We derived 6.27 and 6.28 presuming that the LHS remained unaltered. In many cases, restrictions meet this criteria and any of the 6.26, 6.27 or 6.28 generate the same results.

G) After reading about the definitions of U.S. money supply (M1, M2, M3). I see no reason to exclude either the long bond rate or the t-bill rate. Thus, it is my belief that both should be included in the model and perhaps introduced as a full quadratic structure between the two. So, one might consider including variables for lt, st, lt^2 st^2 and lt*st, lt is ln_rltrate and st is ln_rtbrate.

H) Calculate the Jarque-Bera test statistic (see equation 5.12.1 in Gujarati) and determine if it is significant at the 5% level.

Put the residuals in a named variable:

. predict errs, resid

. summarize errs, detail

Skewness .1114822 (looks pretty good!!)
Kurtosis 3.183538 (also looks pretty good)

JB = . di 28*((.11148^2/6)+((3.1835-3)^2/24))
.09728065

The joint hypothesis is: H0: S=0 and K=3.
The critical chi square level is: $\chi^2|_{0.05,2} = 5.99$

Thus, we easily fail to reject the null of normality in the errors.

I) Run three STATA tests for normality: a) *sktest errs* b) *swilks errs* and c) *sfrancia errs* . .

. *sktest errs*

```
                   Skewness/Kurtosis tests for Normality
                                                 ------- joint ------
```

Variable	Pr(Skewness)	Pr(Kurtosis)	adj chi2(2)	Prob>chi2
errs	0.777	0.448	0.69	0.7084

. *swilk errs*

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
errs	28	0.96755	0.980	-0.042	0.51669

. *sfrancia errs*

Shapiro-Francia W' test for normal data

Variable	Obs	W'	V'	z	Prob>z
errors	28	0.96422	1.193	0.325	0.37264

Discussion: In all three versions of the tests about normality of the errors, we fail to reject the null of normality (p-values=.70, .51, .37). Tests for normality can alert you to problems about your model. Residuals that fail these tests suggest potential problems with functional form of your model, outliers in your data, and other violations of the CNLRM. Students can easily read and learn about these tests for normality at the STATA website and from other sources.

Question 2. Uses the following

Y = Per Capita Consumption of Chickens, Pounds
X2 = Real Disposable Income Per Capita, $
X3 = Real Retail Price of Chicken Per Pound, Cents
X4 = Real Retail Price of Pork Per Pound, Cents
X5 = Real Retail Price of Beef Per Pound, Cents
X6 = Composite Real Price of Chicken Substitutes Per Pound, Cents

a) Done
b) Estimate the double log model using only the first three independent variables. <u>Interpret thoroughly the model result. Do the results jive with demand theory?</u>

. regress ln_y ln_x2 ln_x3 ln_x4

Source	SS	df	MS		
				Number of obs =	23
				F(3, 19) =	336.18
Model	.760427328	3	.253475776	Prob > F =	0.0000
Residual	.014325762	19	.000753987	R-squared =	0.9815
				Adj R-squared =	0.9786
Total	.77475309	22	.03521605	Root MSE =	.02746

ln_y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_x2	.4059242	.0447915	9.06	0.000	.3121745	.4996738
ln_x3	-.4388253	.0833325	-5.27	0.000	-.6132422	-.2644084
ln_x4	.106656	.0878384	1.21	0.240	-.0771919	.2905039
_cons	2.125498	.1378822	15.42	0.000	1.836907	2.414089

Results from this regression are given above. The F-statistic tests the null hypothesis that all coefficients other than the intercept coefficient are jointly equal to zero. The F-statistic is highly significant (p-value=0.0000) suggesting that RHS information offers explanatory power. The R-squared and Adjusted R-squared are both extremely high suggesting a good fit has been achieved. The coefficients on income and own-price are of the correct sign, both highly significant and are in elasticity form. Thus, the income elasticity is estimated to be 0.40. This implies a 10% decrease in income results in a 4% drop in quantity demanded for chicken. The own price elasticity is estimated to be .44. This implies a 10% decrease in price results in a 4.4% increase in the demand for chicken. The model suggests aggregate chicken demand is inelastic with respect to own price and income. The coefficient on the price of pork, a presumed substitute good, is of the correct sign but not statistically significant.

c) . regress ln_y ln_x2 ln_x3 ln_x4 ln_x5 ln_x6

```
      Source |       SS       df       MS              Number of obs =      23
-------------+------------------------------           F(  5,    17) =  270.47
       Model |  .765134808     5  .153026962           Prob > F      =  0.0000
    Residual |  .009618282    17  .000565781           R-squared     =  0.9876
-------------+------------------------------           Adj R-squared =  0.9839
       Total |   .77475309    22   .03521605           Root MSE      =  .02379

------------------------------------------------------------------------------
        ln_y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       ln_x2 |   .3464648   .0717982     4.83   0.000     .1949838    .4979458
       ln_x3 |  -.5880747   .1005235    -5.85   0.000    -.8001608   -.3759887
       ln_x4 |    .365334   .1178702     3.10   0.007     .1166497    .6140183
       ln_x5 |   .3403395   .1270562     2.68   0.016     .0722744    .6084046
       ln_x6 |  -.4531944   .1686692    -2.69   0.016    -.8090554   -.0973334
       _cons |    2.43254   .1618114    15.03   0.000     2.091148    2.773932
------------------------------------------------------------------------------
```

d)

```
. test ln_x5 ln_x6

 ( 1)  ln_x5 = 0
 ( 2)  ln_x6 = 0

       F(  2,    17) =    4.16
            Prob > F =    0.0338
```

Test rejects the null given in (1) and (2) that both are jointly equal to zero.

e) Using the equation 6.26 in the notes, I calculate:

j=2 (# of restrictions)
n-k=17 (from unrestricted model)

```
. di (( .014325762 - .009618282 )/2)/( .009618282/(23-6))
4.1601587
```

This confirms what the test command actually does. It essentially tests the restriction that the coefficients on ln_x5 and ln_x6 are jointly equal to zero.

Many other test options are available in STATA. For example, to test the absurd joint hypothesis that ln_x5=1 and ln_x6=9, there are two ways:

First Way:

```
. test ln_x5=1, notest

. test ln_x6=9, accumulate

 ( 1)  ln_x5 = 1
 ( 2)  ln_x6 = 9

       F(  2,    17) = 3846.83
            Prob > F =    0.0000
```

Second Way:

```
. test (ln_x5=1) (ln_x6=9)

 ( 1)  ln_x5 = 1
 ( 2)  ln_x6 = 9

       F(  2,    17) = 3846.83
            Prob > F =    0.0000
```

The notest option suppresses the output. If you need to test ln_x5=1 before conducting the joint test, the first way (without the notest option) is the best way. If you only need the joint test, the second option is clearly the fastest way to go. Finally, you could have set the coefficient on ln_x5=1 and the coefficient on ln_x6=9, rearranged the regression and ran a restricted version of the model leading you to a manual calculation of equation 6.26 to yield the same F-value.