<u>Quantitative Methods in Finance</u>

Tutorial, Part 7:
Model diagnostics. Normality of the disturbances, multicollinearity, heteroscedasticity, and autocorrelation.

<u>Example:</u> We analyse money demand in the Slovenian economy for the period 1999–2006 (the data are provided in Stata Data file `moneydemand.dta`, while the programming code is given in Stata Do file `moneydemand-commands-t07.do`). We have monthly time-series data available for the following variables:

- year and month of observation (*time*; 1999m1, …, 2006m12);
- harmonized money aggregate M1 (*HM1*; in mil. EUR);
- income of households (*PPR*; in mil. EUR);
- interest rate on demand deposits (*RVP*; at the annual level);
- interest rate on short-term deposits (*RVV*; up to 90 days, at the annual level);
- consumer price index (*CZP*; 2000 = 100).

We estimate the following linear regression model of money demand:

$$HM1_t = \beta_1 + \beta_2 PPR_t + \beta_3 RVP_t + \beta_4 RVV_t + \beta_5 CZP_t + u_t .$$

a) Explore the data using different Stata commands. By using the scatter plots, examine the relationships of the above linear money demand function. Are the relationships among the dependent variable *HM1* and the explanatory variables expected?
b) Estimate the linear regression model of money demand by ordinary least squares, and interpret the Stata output; in particular the regression coefficients. Again, are the signs and magnitudes of the relationships expected?
c) Save the residuals and fitted values of the linear regression model of money demand, i.e. from point b). Check validity of the assumption on normality of the disturbances by using the Jarque–Bera test (perform the test both manually and with the appropriate command). What could you have done if the assumption had been violated?
d) Check validity of the assumption on (absence of) multicollinearity (perform the test both manually e.g. for explanatory variable *PPR* and with the appropriate command). What could you have done if the assumption had been violated?
e) Check validity of the assumption on homoscedasticity by using the White test (perform the test both manually and with the appropriate command). If you find presence of heteroscedasticity in the model, calculate the unbiased standard errors of parameter estimates using a robust estimator of variance, such as the Huber/White estimator of variance. Which assumption of the classical linear regression model was loosened in doing so, and how?
f) Check validity of the assumption on (absence of) autocorrelation by using the Breusch–Godfrey test (perform the test both manually e.g. for AR(4) and with the appropriate command for required lag length). If you find the presence of autocorrelation in the model, calculate the unbiased standard errors of parameter estimates using a HAC estimator of

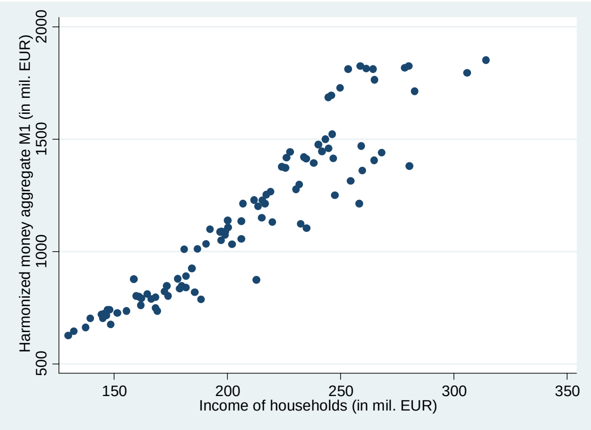
variance, such as the Newey–West robust estimator of variance. Which assumptions of the classical linear regression model were loosened in doing so, and how?

Computer printout of the results in Stata:

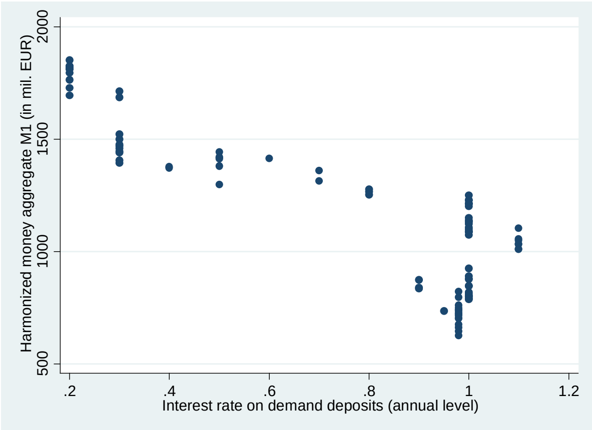
a) Data exploration

```
. tsset time
        time variable:  time, 1999m1 to 2006m12
                delta:  1 month

. sum

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        time |        96       515.5    27.85678        468        563
         hm1 |        96     1151.35     354.281    626.222       1853
         rvp |        96    .7678125    .3251533         .2        1.1
         rvv |        96    6.374062    2.826605        2.6       11.4
         ppr |        96    208.2853     43.2753    129.727    314.326
-------------+--------------------------------------------------------
         pdr |        96    186.4002    51.01813     87.064    301.542
         pgo |        96    5042.121    1808.664   2730.601       9719
         czp |        96    116.5073    14.38521       89.1      135.4
```

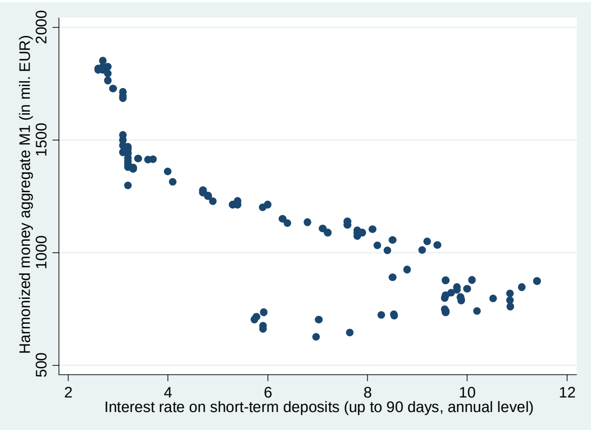
```
. scatter hm1 ppr
```



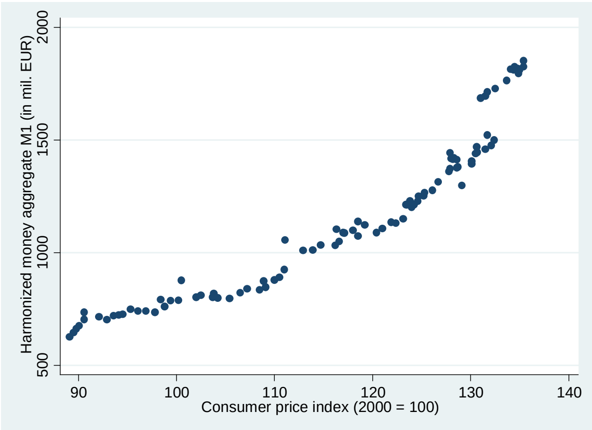

```
. scatter hm1 rvp
```




```
. scatter hm1 rvv
```




```
. scatter hm1 czp
```



b) Estimation of the linear money demand function

```
. regress hm1 ppr rvp rvv czp

      Source |       SS       df       MS              Number of obs =      96
-------------+------------------------------           F(  4,    91) =  527.72
       Model |  11431132.5     4  2857783.12           Prob > F      =  0.0000
    Residual |  492791.936    91    5415.296           R-squared     =  0.9587
-------------+------------------------------           Adj R-squared =  0.9569
       Total |  11923924.4    95  125514.994           Root MSE      =  73.589

------------------------------------------------------------------------------
         hm1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ppr |   1.697766    .513892     3.30   0.001     .6769831     2.71855
         rvp |  -311.6847   45.25178    -6.89   0.000    -401.5718   -221.7976
         rvv |  -11.57513    5.33166    -2.17   0.033    -22.16582     -.98444
         czp |   11.50168   1.472604     7.81   0.000     8.576535    14.42683
       _cons |  -229.2038   125.2134    -1.83   0.070    -477.9248    19.51725
------------------------------------------------------------------------------
```

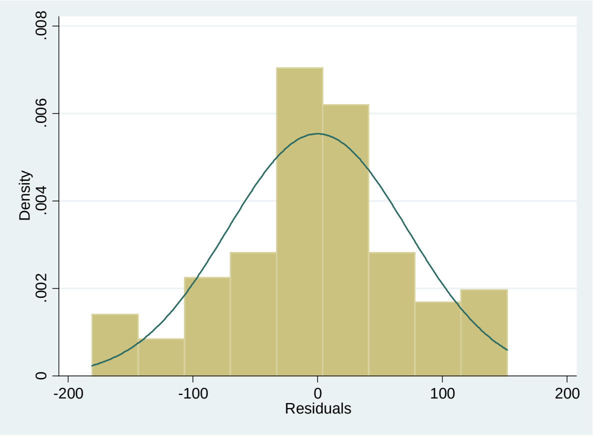
c) Model diagnostics - Normality of the disturbances

```
. predict res, resid
. predict fit, xb

. list hm1 fit res

     +--------------------------------+
     |     hm1        fit         res |
     |--------------------------------|
  1. | 626.222   629.7126   -3.490638 |
  2. | 645.085   630.6459    14.43913 |
  3. | 662.675   663.2455   -.5705134 |
  4. | 675.893   685.4155   -9.522521 |
  5. |  703.11    687.326    15.78391 |
     |--------------------------------|
  6. | 735.355   735.4176   -.0626146 |
  7. | 714.333   706.8221    7.510875 |
  8. | 702.177   689.3663    12.81066 |
  9. | 719.821   688.3902    31.43076 |
 10. | 723.418   698.9282    24.48987 |
     |--------------------------------|
 11. | 726.128   710.6638    15.46422 |
 12. | 748.817   736.8136    12.00344 |
 13. |  740.66   709.7551    30.90489 |
 14. | 740.068   712.9952    27.07281 |
 15. | 735.536   743.3943   -7.858349 |
     |--------------------------------|
 16. | 792.003   751.5343    40.46867 |
 17. | 761.324   750.6394    10.68459 |
 18. | 785.896   807.9182   -22.02216 |
 19. | 787.823   768.5693    19.25373 |
 20. | 878.068   773.7431    104.3249 |
     |--------------------------------|
 21. | 801.827   789.3091    12.51789 |
 22. | 810.073    806.821    3.251977 |
 23. | 802.401   832.7275   -30.32648 |
 24. | 818.129   842.4316   -24.30257 |
 25. | 799.304   820.4866   -21.18257 |
     |--------------------------------|
 26. | 797.012   841.6254   -44.61332 |

...
```

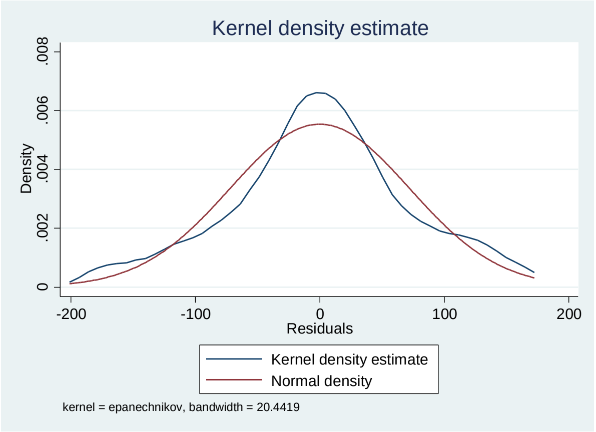
```
 83. |    1523    1574.57   -51.57023 |
 84. |    1714   1636.204    77.79575 |
 85. |    1687   1563.568    123.4316 |
     |--------------------------------|
 86. |    1694   1602.508    91.49198 |
 87. |    1728   1623.145    104.8554 |
 88. |    1765   1663.847    101.1526 |
 89. |    1795   1746.891    48.10896 |
 90. |    1825   1698.695    126.3048 |
     |--------------------------------|
 91. |    1814   1663.368    150.6319 |
 92. |    1813   1661.113    151.8867 |
 93. |    1826   1673.886    152.1144 |
 94. |    1813   1671.924    141.0764 |
 95. |    1817   1701.468    115.5324 |
     |--------------------------------|
 96. |    1853   1768.186    84.81373 |
     +--------------------------------+
```

. histogram res, normal

```
(bin=9, start=-180.69301, width=36.978604)
```



. kdensity res, normal



. sum res, detail

```
                          Residuals
-------------------------------------------------------------
      Percentiles      Smallest
 1%     -180.693       -180.693
 5%    -147.8331      -158.7813
10%    -93.55871       -156.471       Obs                  96
25%    -37.31053      -153.6954       Sum of Wgt.          96

50%     1.594681                      Mean          -5.57e-08
                        Largest       Std. Dev.       72.0228
75%     39.02831       141.0764
90%     101.1526       150.6319       Variance       5187.283
95%     126.3048       151.8867       Skewness      -.1189608
99%     152.1144       152.1144       Kurtosis       3.034316
```

. return list

```
scalars:
                  r(N) =  96
              r(sum_w) =  96
               r(mean) =  -5.57241340478e-08
                r(Var) =  5187.283463078505
                 r(sd) =  72.02279821749849
```

```
            r(skewness) =  -.1189607731123617
            r(kurtosis) =  3.034315649013452
                 r(sum) =  -5.34951686859e-06
                 r(min) =  -180.6930084228516
                 r(max) =  152.1144256591797
                  r(p1) =  -180.6930084228516
                  r(p5) =  -147.8331298828125
                 r(p10) =  -93.55870819091797
                 r(p25) =  -37.31053161621094
                 r(p50) =  1.594681181013584
                 r(p75) =  39.02831268310547
                 r(p90) =  101.1526336669922
                 r(p95) =  126.3047637939453
                 r(p99) =  152.1144256591797
```

```
. scalar obs=r(N)
. scalar sk=r(skewness)
. scalar ku=r(kurtosis)

. scalar jb=obs*(sk^2/6 + (ku-3)^2/24)
. display jb
.2311369

. display chi2tail(2,jb)
.89085959

. jb6 res
Jarque-Bera normality test:  .2311 Chi(2)  .8909
Jarque-Bera test for Ho: normality: (res)
```

d) Model diagnostics - Multicollinearity

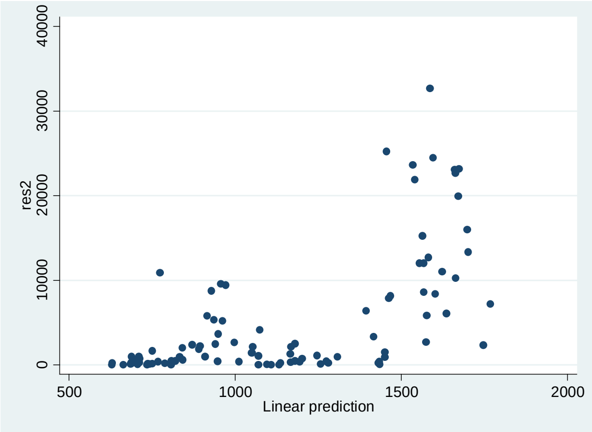
```
. regress ppr rvp rvv czp /* Example of calculation for variable PPR */

      Source |       SS       df       MS              Number of obs =      96
-------------+------------------------------           F(  3,    92) =  235.40
       Model |  157405.554     3  52468.5178           Prob > F      =  0.0000
    Residual |  20505.8858    92  222.890063           R-squared     =  0.8847
-------------+------------------------------           Adj R-squared =  0.8810
       Total |  177911.439    95  1872.75199           Root MSE      =   14.93

------------------------------------------------------------------------------
         ppr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         rvp |  -24.07317   8.830847    -2.73   0.008    -41.61199    -6.53434
         rvv |  -.3925855     1.0809    -0.36   0.717    -2.539347    1.754175
         czp |   2.343588   .1719202    13.63   0.000      2.00214    2.685037
       _cons |  -43.77383   24.98969    -1.75   0.083    -93.40551    5.857857
------------------------------------------------------------------------------

. scalar R2ppr=e(r2)

. scalar VIFppr=1/(1-R2ppr)
. display VIFppr
8.6761158

. qui regress hm1 ppr rvp rvv czp
. estat vif

    Variable |       VIF       1/VIF
-------------+----------------------
         ppr |      8.68    0.115259
         czp |      7.87    0.127027
```

```
         rvv |      3.98    0.250982
         rvp |      3.80    0.263300
-------------+----------------------
    Mean VIF |      6.08
```

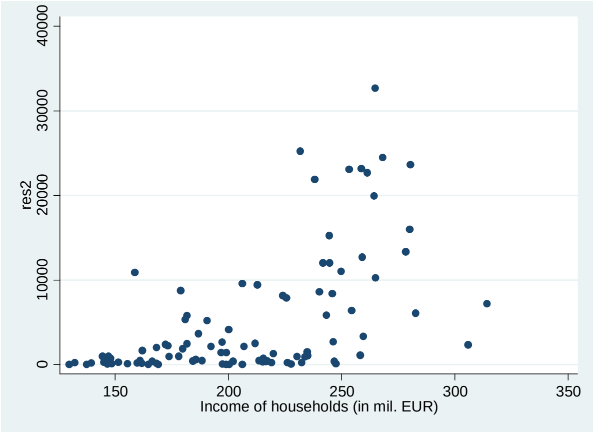
e) Model diagnostics - Homoscedasticity

```
. gen res2=res^2
```

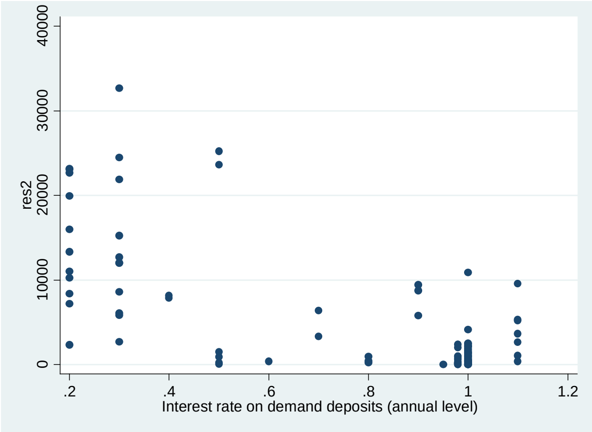
```
. scatter res2 fit
```



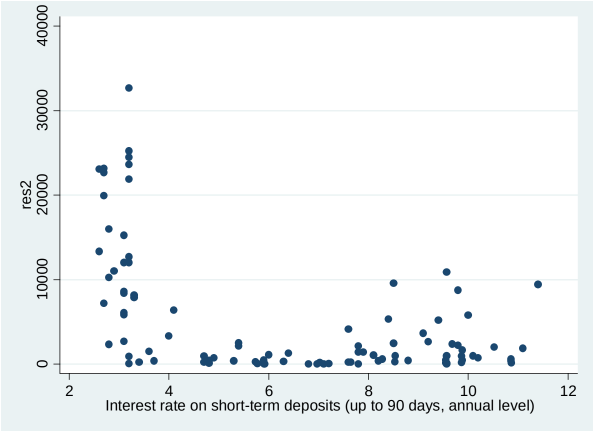

```
. scatter res2 ppr
```



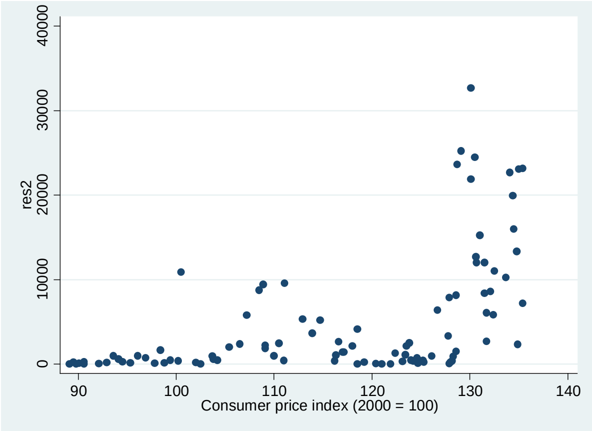

```
. scatter res2 rvp
```




```
. scatter res2 rvv
```




```
. scatter res2 czp
```



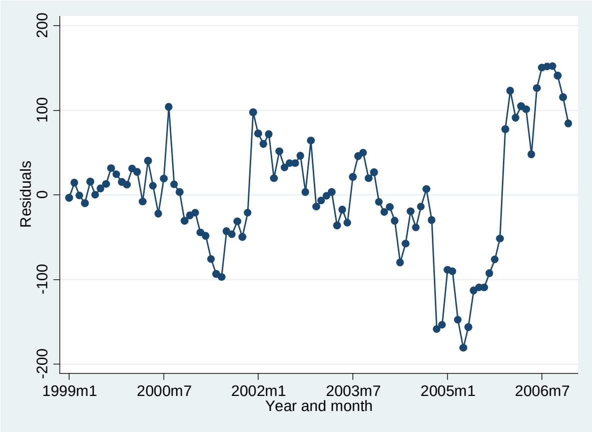
```
. gen ppr2=ppr^2 /* Perform White test manually */
. gen rvp2=rvp^2
. gen rvv2=rvv^2
. gen czp2=czp^2

. gen pprrvp=ppr*rvp
. gen pprrvv=ppr*rvv
. gen pprczp=ppr*czp
. gen rvprvv=rvp*rvv
. gen rvpczp=rvp*czp
. gen rvvczp=rvv*czp

. regress res2 ppr rvp rvv czp ppr2 rvp2 rvv2 czp2 pprrvp pprrvv pprczp rvprvv
  rvpczp rvvczp

      Source |       SS       df       MS              Number of obs =      96
-------------+------------------------------           F( 14,    81) =    7.39
       Model |  2.8855e+09    14   206110493           Prob > F      =  0.0000
    Residual |  2.2605e+09    81  27907485.4           R-squared     =  0.5607
-------------+------------------------------           Adj R-squared =  0.4848
       Total |  5.1461e+09    95  54168981.3           Root MSE      =  5282.8

------------------------------------------------------------------------------
        res2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ppr |  -584.4684   845.2599    -0.69   0.491     -2266.27    1097.334
         rvp |  -250020.7   248329.1    -1.01   0.317    -744117.8    244076.4
         rvv |  -8164.354   8897.928    -0.92   0.362    -25868.44    9539.733
         czp |   2868.217   5063.871     0.57   0.573    -7207.297    12943.73
        ppr2 |  -5.120904   2.440723    -2.10   0.039    -9.977177   -.2646304
        rvp2 |   32406.94   28668.18     1.13   0.262    -24633.75    89447.64
        rvv2 |   143.0258   266.7427     0.54   0.593    -387.7084      673.76
        czp2 |  -45.08782   25.93511    -1.74   0.086    -96.69055    6.514916
      pprrvp |  -141.8024   242.1751    -0.59   0.560     -623.655    340.0501
      pprrvv |   -9.72262   31.09498    -0.31   0.755    -71.59188    52.14664
      pprczp |   25.84947   13.61795     1.90   0.061    -1.245982    52.94493
      rvprvv |   3603.482   6050.429     0.60   0.553    -8434.973    15641.94
      rvpczp |   1636.692   1837.513     0.89   0.376    -2019.382    5292.766
      rvvczp |    37.0222   88.06732     0.42   0.675    -138.2041    212.2485
       _cons |   63626.61   355214.1     0.18   0.858    -643138.1    770391.3
------------------------------------------------------------------------------

. ereturn list

scalars:
                  e(N) =  96
               e(df_m) =  14
               e(df_r) =  81
                  e(F) =  7.385491376162329
                 e(r2) =  .5607300930346307
               e(rmse) =  5282.753584703605
                e(mss) =  2885546902.323607
                e(rss) =  2260506320.372603
               e(r2_a) =  .4848068992381471
                 e(ll) =  -950.9944097108637
               e(ll_0) =  -990.4811888750949
               e(rank) =  15

macros:
            e(cmdline) : "regress res2 ppr rvp rvv czp ppr2 rvp2 rvv2 czp2 pprrvp
                          pprrvv pprczp rvprvv rvpczp rvvczp"
              e(title) : "Linear regression"
         e(marginsok) : "XB default"
                e(vce) : "ols"
             e(depvar) : "res2"
```

```
                 e(cmd) : "regress"
          e(properties) : "b V"
             e(predict) : "regres_p"
               e(model) : "ols"
           e(estat_cmd) : "regress_estat"

matrices:
                   e(b) :  1 x 15
                   e(V) :  15 x 15

functions:
             e(sample)
```

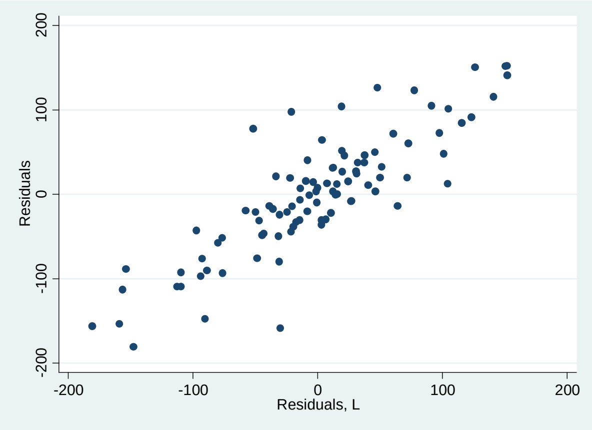
```
. scalar theta=e(N)*e(r2)

. display theta, chi2tail(e(rank)-1,theta)
53.830089 5.712e-11

. qui regress hm1 ppr rvp rvv czp
. estat imtest, white

White's test for Ho: homoskedasticity
         against Ha: unrestricted heteroskedasticity

         chi2(14)     =     53.83
         Prob > chi2  =    0.0000

Cameron & Trivedi's decomposition of IM-test

---------------------------------------------------
              Source |       chi2     df      p
---------------------+-----------------------------
  Heteroskedasticity |      53.83     14    0.0000
            Skewness |       4.68      4    0.3213
            Kurtosis |       0.01      1    0.9136
---------------------+-----------------------------
               Total |      58.53     19    0.0000
---------------------------------------------------

. regress hm1 ppr rvp rvv czp, robust

Linear regression                                      Number of obs =      96
                                                       F(  4,    91) = 1000.25
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.9587
                                                       Root MSE      =  73.589

------------------------------------------------------------------------------
             |               Robust
         hm1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ppr |   1.697766   .5633882     3.01   0.003     .5786649    2.816868
         rvp |  -311.6847   44.33028    -7.03   0.000    -399.7413   -223.6281
         rvv |  -11.57513   3.532513    -3.28   0.001    -18.59203   -4.558225
         czp |   11.50168    1.32376     8.69   0.000     8.872196    14.13117
       _cons |  -229.2038   58.25138    -3.93   0.000     -344.913   -113.4945
------------------------------------------------------------------------------
```

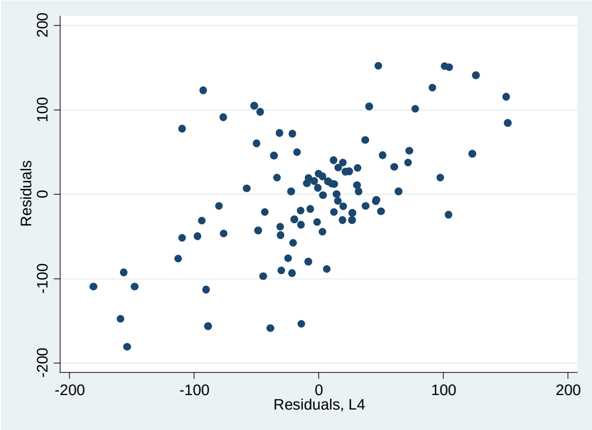
f) Model diagnostics - Autocorrelation

```
. twoway connected res time
```




```
. scatter res l.res                         . scatter res l4.res
```






```
. gen res_l1=res[_n-1] /* Perform Breusch-Godfrey test manually, AR(4) */
(1 missing value generated)

. gen res_l2=res[_n-2]
(2 missing values generated)

. gen res_l3=res[_n-3]
(3 missing values generated)

. gen res_l4=res[_n-4]
(4 missing values generated)

. regress res ppr rvp rvv czp res_l1 res_l2 res_l3 res_l4

      Source |       SS       df       MS              Number of obs =      92
-------------+------------------------------           F(  8,    83) =   30.24
       Model |  366687.863     8  45835.9829           Prob > F      =  0.0000
    Residual |  125792.381    83  1515.57086           R-squared     =  0.7446
-------------+------------------------------           Adj R-squared =  0.7200
       Total |  492480.244    91  5411.87081           Root MSE      =   38.93

------------------------------------------------------------------------------
         res |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ppr |  -.2646584   .3019436    -0.88   0.383    -.8652121    .3358952
         rvp |  -18.67495   24.82208    -0.75   0.454    -68.04508    30.69517
```

```
         rvv |   .1185893   3.161154     0.04   0.970    -6.168819    6.405997
         czp |   .5051443   .8727978     0.58   0.564    -1.230815    2.241104
      res_l1 |   .8674026   .1120089     7.74   0.000     .6446214    1.090184
      res_l2 |  -.1409334   .1454018    -0.97   0.335    -.4301317     .148265
      res_l3 |   .1761789   .1480089     1.19   0.237     -.118205    .4705627
      res_l4 |  -.0063491   .1139522    -0.06   0.956    -.2329954    .2202971
       _cons |   11.09089   79.46617     0.14   0.889    -146.9641    169.1459
------------------------------------------------------------------------------

. scalar lm=e(N)*e(r2)

. display lm, chi2tail(4,lm)
68.500785 4.703e-14

. qui regress hm1 ppr rvp rvv czp
. estat bgodfrey, lags(4) nomiss0

Breusch-Godfrey LM test for autocorrelation
---------------------------------------------------------------------------
    lags(p)  |          chi2               df                 Prob > chi2
-------------+-------------------------------------------------------------
       4     |         68.501               4                   0.0000
---------------------------------------------------------------------------
                        H0: no serial correlation

. estat bgodfrey, lags(1/12)

Breusch-Godfrey LM test for autocorrelation
---------------------------------------------------------------------------
    lags(p)  |          chi2               df                 Prob > chi2
-------------+-------------------------------------------------------------
       1     |         70.771               1                   0.0000
       2     |         70.773               2                   0.0000
       3     |         71.397               3                   0.0000
       4     |         71.398               4                   0.0000
       5     |         71.639               5                   0.0000
       6     |         71.641               6                   0.0000
       7     |         73.403               7                   0.0000
       8     |         73.536               8                   0.0000
       9     |         74.066               9                   0.0000
      10     |         74.112              10                   0.0000
      11     |         74.164              11                   0.0000
      12     |         76.742              12                   0.0000
---------------------------------------------------------------------------
                        H0: no serial correlation

. estat bgodfrey, lags(1/90)

Breusch-Godfrey LM test for autocorrelation
---------------------------------------------------------------------------
    lags(p)  |          chi2               df                 Prob > chi2
-------------+-------------------------------------------------------------
       1     |         70.771               1                   0.0000
       2     |         70.773               2                   0.0000
       3     |         71.397               3                   0.0000
       4     |         71.398               4                   0.0000
       5     |         71.639               5                   0.0000
       6     |         71.641               6                   0.0000
       7     |         73.403               7                   0.0000
       8     |         73.536               8                   0.0000
       9     |         74.066               9                   0.0000
      10     |         74.112              10                   0.0000

     ...
```

```
     41      |         88.902            41                 0.0000
     42      |         89.282            42                 0.0000
     43      |         89.287            43                 0.0000
     44      |         89.390            44                 0.0001
     45      |         89.410            45                 0.0001
     46      |         89.685            46                 0.0001
     47      |         90.072            47                 0.0002
     48      |         90.076            48                 0.0002
     49      |         90.143            49                 0.0003
     50      |         90.540            50                 0.0004

    ...

     71      |         93.846            71                 0.0361
     72      |         93.849            72                 0.0429
     73      |         93.995            73                 0.0496
     74      |         94.011            74                 0.0582
     75      |         94.011            75                 0.0680
     76      |         94.012            76                 0.0789
     77      |         94.379            77                 0.0869
     78      |         94.403            78                 0.0996
     79      |         95.140            79                 0.1042
     80      |         95.440            80                 0.1147
     81      |         95.442            81                 0.1303
     82      |         95.442            82                 0.1472
     83      |         95.491            83                 0.1646
     84      |         95.585            84                 0.1823
     85      |         95.601            85                 0.2026
     86      |         95.607            86                 0.2244
     87      |         95.651            87                 0.2465
     88      |         95.920            88                 0.2644
     89      |         95.920            89                 0.2893
     90      |         95.919            90                 0.3152
---------------------------------------------------------------------------
                       H0: no serial correlation
```

. newey hm1 ppr rvp rvv czp, lag(78)

```
Regression with Newey-West standard errors          Number of obs  =        96
maximum lag: 78                                     F(  4,    91)  =    859.25
                                                    Prob > F       =    0.0000

------------------------------------------------------------------------------
             |             Newey-West
         hm1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ppr |   1.697766   .3108188     5.46   0.000     1.080363     2.31517
         rvp |  -311.6847   64.70783    -4.82   0.000    -440.2189   -183.1505
         rvv |  -11.57513   6.148327    -1.88   0.063    -23.78802     .637769
         czp |   11.50168    .672376    17.11   0.000     10.16609    12.83727
       _cons |  -229.2038   71.30645    -3.21   0.002    -370.8453   -87.56224
------------------------------------------------------------------------------
```

■