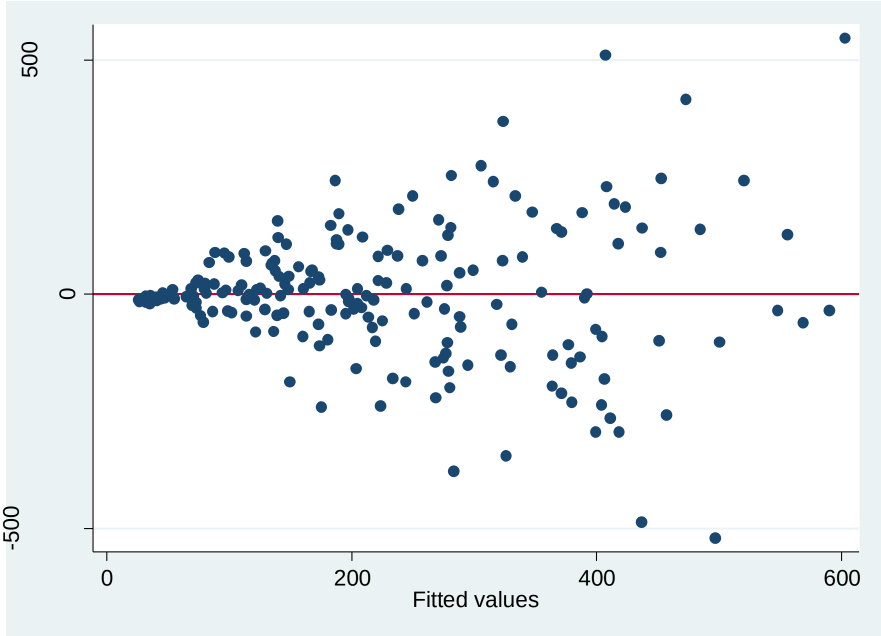
# 4. Heteroscedasticity Problem

## Simulated Example

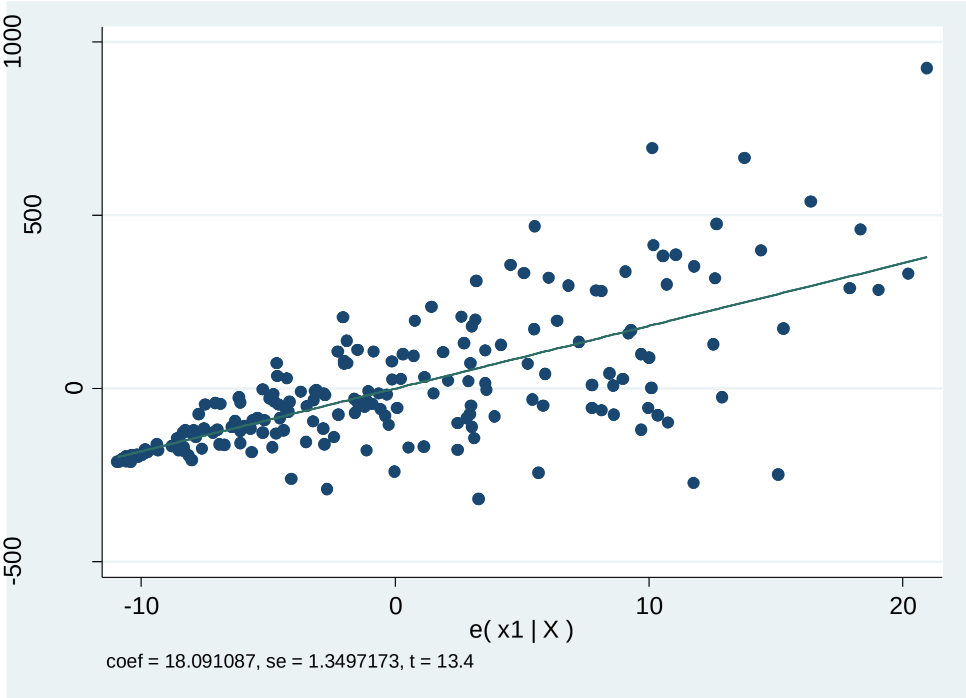
```
. set obs 200
number of observations (_N) was 0, now 200

. g x1=abs(rnormal(10,10))

. g sigma2=100*x1^2+rnormal(0,1)

. g s=sqrt(sigma2)
(1 missing value generated)

. g u=rnormal(0,s)
(1 missing value generated)

. g y=10+20*x1+u
(1 missing value generated)
```

| | |
|---|---|
| **rvfplot** | residual-versus-fitted plot |
| **avplot** | added-variable plot |
| **avplots** | all added-variable plots in one image |
| **cprplot** | component-plus-residual plot |
| **acprplot** | augmented component-plus-residual plot |
| **rvpplot** | residual-versus-predictor plot |
| **lvr2plot** | leverage-versus-squared-residual plot |

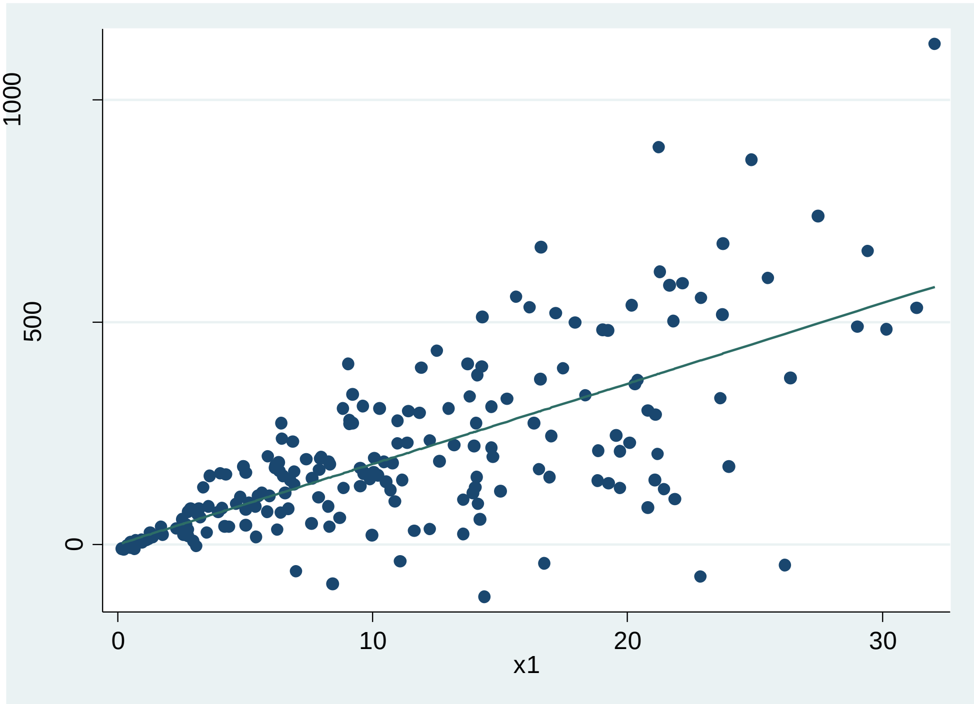
```
. reg y x1

      Source |       SS           df       MS      Number of obs   =       199
-------------+----------------------------------   F(1, 197)       =    179.66
       Model |  3688449.96         1  3688449.96   Prob > F        =    0.0000
    Residual |  4044514.53       197  20530.5306   R-squared       =    0.4770
-------------+----------------------------------   Adj R-squared   =    0.4743
       Total |  7732964.49       198  39055.3762   Root MSE        =    143.28

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   18.09109   1.349717    13.40   0.000     15.42934    20.75284
       _cons |   23.02303   18.09761     1.27   0.205     -12.6669    58.71295
------------------------------------------------------------------------------

. rvfplot, yline(0)
```



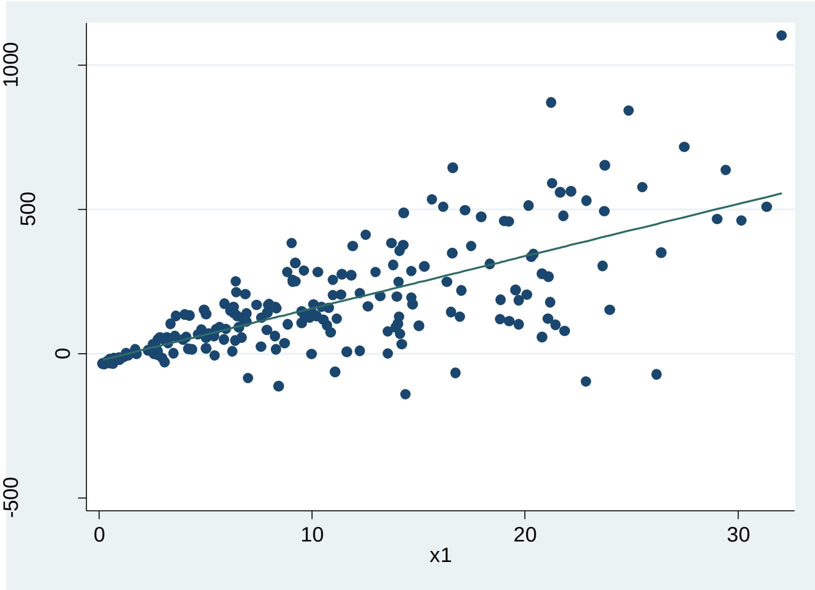
```
. avplot x1, recast(scatter)
```




```
. cprplot x1, recast(scatter)
```



```
. acprplot x1, recast(scatter)
```




```
. estat hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of y

         chi2(1)      =   110.22
         Prob > chi2  =   0.0000

. whitetst
White's general test statistic :  50.00066  Chi-sq( 2)  P-value =  1.4e-11

. predict uhat, r

. g uhat2=uhat^2

. g x12=x1^2

. reg uhat2 x1 x12

      Source |       SS           df       MS      Number of obs   =       199
-------------+----------------------------------   F(2, 196)       =     32.89
       Model |  9.6535e+10         2  4.8268e+10   Prob > F        =    0.0000
    Residual |  2.8767e+11       196  1.4677e+09   R-squared       =    0.2513
-------------+----------------------------------   Adj R-squared   =    0.2436
       Total |  3.8420e+11       198  1.9404e+09   Root MSE        =     38311

------------------------------------------------------------------------------
       uhat2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |    623.791   1203.673     0.52   0.605    -1750.022    2997.604
         x12 |   83.09351   43.00584     1.93   0.055    -1.720077    167.9071
       _cons |  -1537.417    6966.23    -0.22   0.826    -15275.81    12200.97
------------------------------------------------------------------------------

. sca r2=e(r2)

. sca n=e(N)
```

```
. sca list r2 n
        r2 =  .25125958
         n =        199

. sca whitetst=r2*n

. sca list whitetst
  whitetst =  50.000656

. reg uhat2 x12, nocon

      Source |       SS           df       MS      Number of obs   =       199
-------------+----------------------------------   F(1, 198)       =    122.27
       Model |  1.7806e+11         1  1.7806e+11   Prob > F        =    0.0000
    Residual |  2.8834e+11       198  1.4563e+09   R-squared       =    0.3818
-------------+----------------------------------   Adj R-squared   =    0.3787
       Total |  4.6641e+11       199  2.3437e+09   Root MSE        =     38161

------------------------------------------------------------------------------
       uhat2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         x12 |   108.0187   9.768714    11.06   0.000     88.75461    127.2828
------------------------------------------------------------------------------

. predict sigma2hat, xb

. g shat=sqrt(sigma2hat)

. g ys=y/shat
(1 missing value generated)

. g x1s=x1/shat

. g x0s=1/shat

. reg ys x1s x0s, nocon

      Source |       SS           df       MS      Number of obs   =       199
-------------+----------------------------------   F(2, 197)       =    490.96
       Model |  1080.51297         2  540.256485   Prob > F        =    0.0000
    Residual |  216.780575       197  1.10040901   R-squared       =    0.8329
-------------+----------------------------------   Adj R-squared   =    0.8312
       Total |  1297.29355       199  6.51906304   Root MSE        =     1.049

------------------------------------------------------------------------------
          ys |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         x1s |   20.31127   .8363783    24.28   0.000     18.66187    21.96068
         x0s |   9.930732   1.101445     9.02   0.000     7.758594    12.10287
------------------------------------------------------------------------------

. vwls y x1, sd(shat)

Variance-weighted least-squares regression      Number of obs     =       199
Goodness-of-fit chi2(197)  =  216.78            Model chi2(1)     =    648.97
Prob > chi2                =  0.1591            Prob > chi2       =    0.0000
------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   20.31127   .7973072    25.47   0.000     18.74858    21.87397
       _cons |   9.930732   1.049992     9.46   0.000     7.872786    11.98868
------------------------------------------------------------------------------
```

## Example

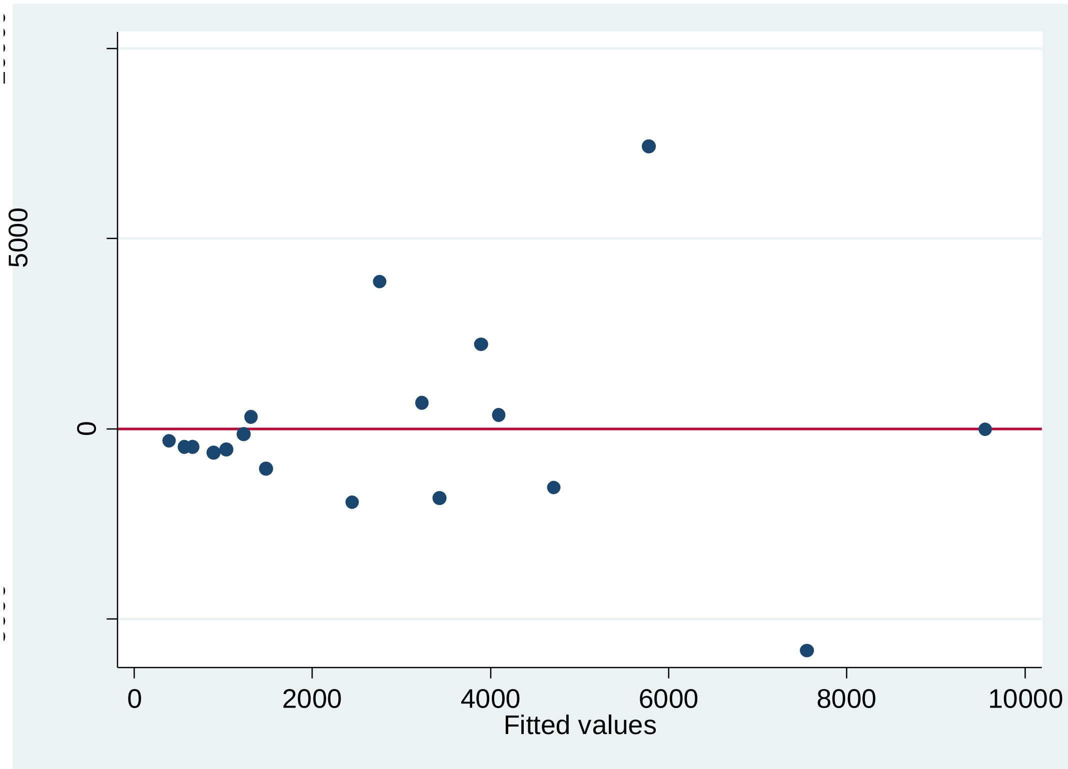
### Detecting Heteroscedasticity Problem

Model: $RD = \beta_0 + \beta_1 SALE + u$

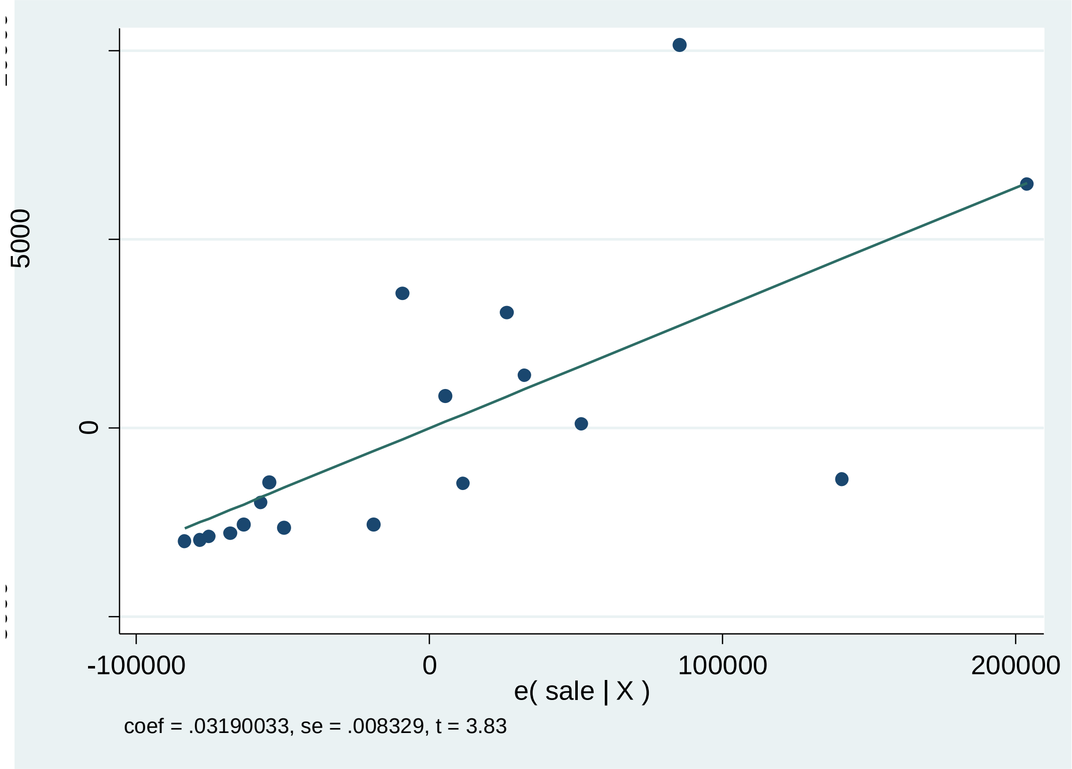
```
. reg rd sale

      Source |       SS       df       MS              Number of obs =      18
-------------+------------------------------           F(  1,    16) =   14.67
       Model |   111675212     1   111675212           Prob > F      =  0.0015
    Residual |   121806834    16  7612927.12           R-squared     =  0.4783
-------------+------------------------------           Adj R-squared =  0.4457
       Total |   233482046    17    13734238           Root MSE      =  2759.2

------------------------------------------------------------------------------
          rd |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        sale |   .0319003    .008329     3.83   0.001     .0142436     .049557
       _cons |   192.9932   990.9858     0.19   0.848    -1907.803    2293.789
------------------------------------------------------------------------------
```

Informal Test:

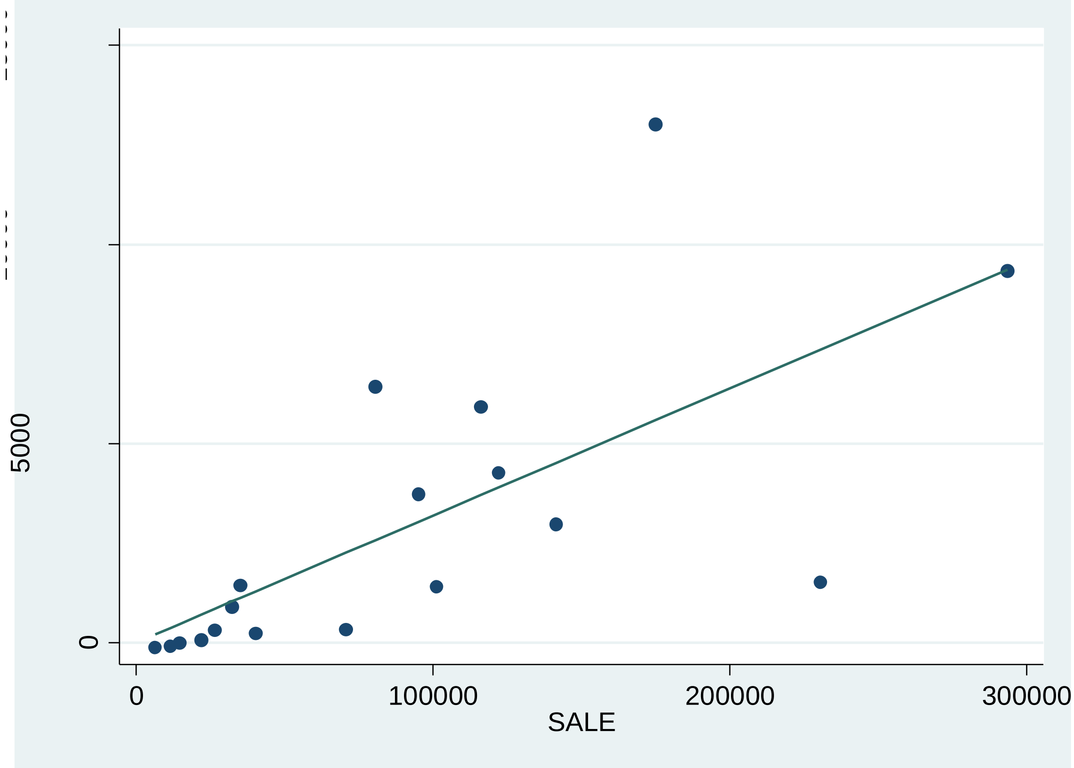
```
. rvfplot, yline(0)
```



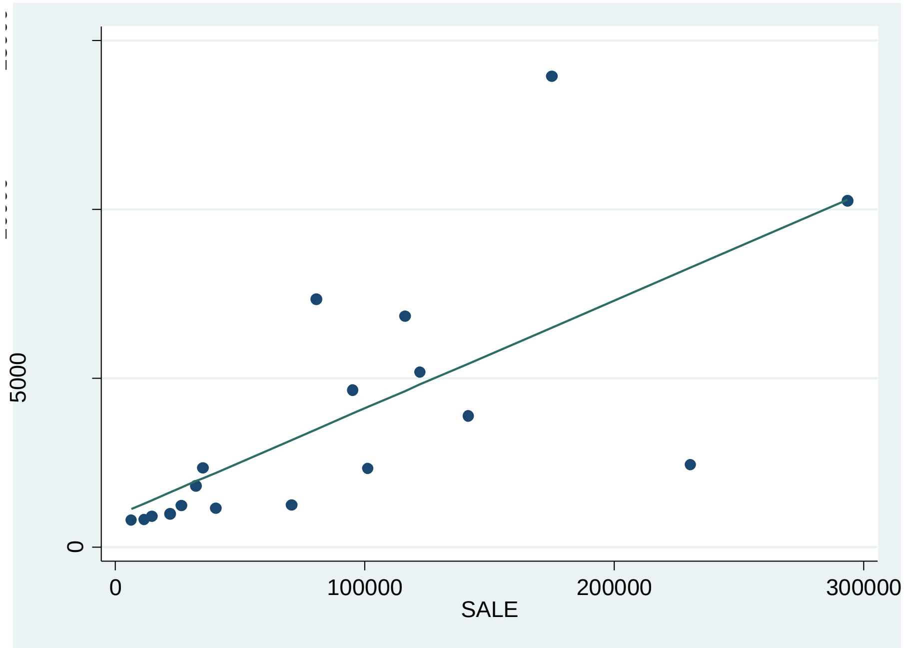
```
. avplot sale, recast(scatter)
```



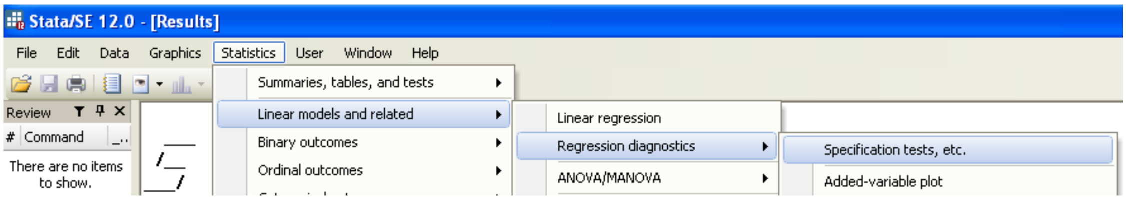

```
. cprplot sale, recast(scatter)
```

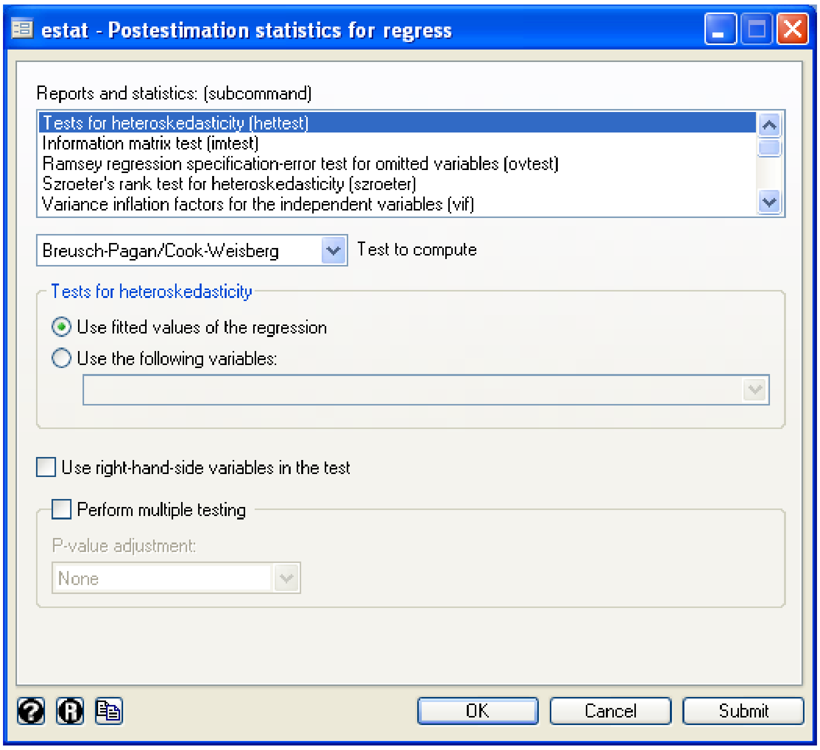


```
. acprplot sale, recast(scatter)
```




Formal Test:



Breusch-Pagan Test:

```
. estat hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of rd

         chi2(1)      =     8.91
         Prob > chi2  =   0.0028
```

White's Test:

```
. whitetst

White's general test statistic :  5.212492  Chi-sq( 2)  P-value =  .0738
```

Information Matrix Test (White's test):

```
. estat imtest

Cameron & Trivedi's decomposition of IM-test

---------------------------------------------------
              Source |       chi2     df      p
---------------------+-----------------------------
  Heteroskedasticity |       5.21      2    0.0738
            Skewness |       0.94      1    0.3315
            Kurtosis |       3.59      1    0.0582
---------------------+-----------------------------
               Total |       9.74      4    0.0450
---------------------------------------------------
```

Szroteter's Test:

```
. estat szroeter sale

Szroeter's test for homoskedasticity

    Ho: variance constant
    Ha: variance monotonic in sale

         chi2(1)      =     9.56
         Prob > chi2  =   0.0020
```

## Solving Heteroscedasticity Problem

From OLS Estimated Results:

```
. reg rd sale

      Source |       SS       df       MS              Number of obs =      18
-------------+------------------------------           F(  1,    16) =   14.67
       Model |  111675212     1   111675212            Prob > F      =  0.0015
    Residual |  121806834    16  7612927.12            R-squared     =  0.4783
-------------+------------------------------           Adj R-squared =  0.4457
       Total |  233482046    17    13734238            Root MSE      =  2759.2

------------------------------------------------------------------------------
          rd |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        sale |   .0319003    .008329     3.83   0.001     .0142436     .049557
       _cons |   192.9932   990.9858     0.19   0.848    -1907.803    2293.789
------------------------------------------------------------------------------
```

Determine Variance equation

Functional form: $\hat{u}_i^2 \approx \sigma_i^2 = \alpha_1 + \alpha_2 SALE_i + v_i$

```
. predict uhat, resid
. g uhat2=uhat^2
. reg uhat2 sale
      Source |       SS       df       MS              Number of obs =      18
-------------+------------------------------           F(  1,    16) =    4.56
       Model |  8.1603e+14     1  8.1603e+14           Prob > F      =  0.0484
    Residual |  2.8605e+15    16  1.7878e+14           R-squared     =  0.2220
-------------+------------------------------           Adj R-squared =  0.1733
       Total |  3.6765e+15    17  2.1627e+14           Root MSE      =  1.3e+07

------------------------------------------------------------------------------
       uhat2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        sale |   86.23211   40.36254     2.14   0.048     .6673552    171.7969
       _cons |  -974468.8    4802344    -0.20   0.842    -1.12e+07     9206045
------------------------------------------------------------------------------
```

Weighted Least Squares:

Remedy by using Weighted Least Squares

Functional form: $w \cdot RD = \beta_1 w + \beta_2 w \cdot SALE + w \cdot u$

where $w = \frac{1}{\hat{\sigma}_i}$

```
. predict sigma2, xb

. g sigmahat=sqrt(sigma2)

. g w=1/sigmahat

. g wrd=w*rd

. g wsale=w*sale

. reg wrd wsale w, noconst
      Source |       SS       df       MS              Number of obs =      17
-------------+------------------------------           F(  2,    15) =   17.24
       Model |  24.9762465     2  12.4881233           Prob > F      =  0.0001
    Residual |   10.868054    15  .724536932           R-squared     =  0.6968
-------------+------------------------------           Adj R-squared =  0.6564
       Total |  35.8443005    17  2.10848827           Root MSE      =   .8512

------------------------------------------------------------------------------
         wrd |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       wsale |   .0378137   .0068275     5.54   0.000     .0232613    .0523661
           w |   -347.254   168.4375    -2.06   0.057    -706.2701    11.76205
------------------------------------------------------------------------------
```

Using command Weighted Least Squares

```
. vwls rd sale, sd(sigmahat)

Variance-weighted least-squares regression           Number of obs   =      17
Goodness-of-fit chi2(15)   =   10.87                 Model chi2(1)   =   22.22
Prob > chi2                =  0.7619                 Prob > chi2     =  0.0000
------------------------------------------------------------------------------
          rd |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        sale |   .0378137    .008021     4.71   0.000     .0220928    .0535346
       _cons |   -347.254    197.883    -1.75   0.079    -735.0977    40.58962
------------------------------------------------------------------------------
```

## Relieve the Problem: White's Robust Standard Error

From OLS Estimated Results:

```
. reg rd sale

      Source |       SS       df       MS              Number of obs =      18
-------------+------------------------------           F(  1,    16) =   14.67
       Model |   111675212     1   111675212           Prob > F      =  0.0015
    Residual |   121806834    16  7612927.12           R-squared     =  0.4783
-------------+------------------------------           Adj R-squared =  0.4457
       Total |   233482046    17    13734238           Root MSE      =  2759.2

------------------------------------------------------------------------------
          rd |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        sale |   .0319003    .008329     3.83   0.001     .0142436     .049557
       _cons |   192.9932   990.9858     0.19   0.848    -1907.803    2293.789
------------------------------------------------------------------------------
```

OLS with White's Robust Standard Error Estimated Results:

```
. reg rd sale, vce(robust)

Linear regression                                      Number of obs =      18
                                                       F(  1,    16) =    9.88
                                                       Prob > F      =  0.0063
                                                       R-squared     =  0.4783
                                                       Root MSE      =  2759.2

------------------------------------------------------------------------------
             |               Robust
          rd |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        sale |   .0319003    .010147     3.14   0.006     .0103896     .053411
       _cons |   192.9932   533.9317     0.36   0.722    -938.8916    1324.878
------------------------------------------------------------------------------
```