1. Interactions

Source: Wooldridge HPRICE1: Collected from the real estate pages of the *Boston Globe* during 1990. These are homes selling in the Boston, MA area.

```
price                    house price, $1000s
bdrms                    number of bedrooms
sqrft                    size of house in square feet
lotsize                  size of lot in square feet
colonial                 =1 if home is colonial style
```

```
. sum price bdrms lotsize sqrft colonial

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       price |        88     293.546    102.7134        111        725
       bdrms |        88    3.568182    .8413926          2          7
     lotsize |        88    9019.864    10174.15       1000      92681
       sqrft |        88    2013.693    577.1916       1171       3880
    colonial |        88    .6931818    .4638161          0          1

. reg price bdrms sqrft lotsize

      Source |       SS       df       MS              Number of obs =      88
-------------+------------------------------           F(  3,    84) =   57.46
       Model |  617130.701     3  205710.234           Prob > F      =  0.0000
    Residual |  300723.805    84   3580.0453           R-squared     =  0.6724
-------------+------------------------------           Adj R-squared =  0.6607
       Total |  917854.506    87  10550.0518           Root MSE      =  59.833

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       bdrms |   13.85252   9.010145     1.54   0.128    -4.065141    31.77018
       sqrft |   .1227782   .0132374     9.28   0.000     .0964541    .1491022
     lotsize |   .0020677   .0006421     3.22   0.002     .0007908    .0033446
       _cons |  -21.77031   29.47504    -0.74   0.462    -80.38466    36.84405
------------------------------------------------------------------------------


. gen bdrms_sqrft=bdrms*sqrft;
. reg price bdrms sqrft bdrms_sqrft lotsize

      Source |       SS       df       MS              Number of obs =      88
-------------+------------------------------           F(  4,    83) =   46.48
       Model |  634546.045     4  158636.511           Prob > F      =  0.0000
    Residual |  283308.461    83  3413.35495           R-squared     =  0.6913
-------------+------------------------------           Adj R-squared =  0.6765
       Total |  917854.506    87  10550.0518           Root MSE      =  58.424

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       bdrms |  -33.71534   22.82291    -1.48   0.143    -79.10919    11.67852
       sqrft |   .0337926   .0414616     0.82   0.417    -.0486728    .1162579
 bdrms_sqrft |   .0218268   .0096631     2.26   0.027     .0026074    .0410462
     lotsize |   .0019927   .0006279     3.17   0.002     .0007439    .0032416
       _cons |   165.4265   87.73015     1.89   0.063    -9.065246    339.9182
------------------------------------------------------------------------------
```

2. Confidence interval on predictions

```
. g bdrms0=bdrms-3
. g sqrft0=sqrft-1500
. gen lotsize0=lotsize-9000

. reg price bdrms0 sqrft0 lotsize0

      Source |       SS       df       MS              Number of obs =      88
-------------+------------------------------           F(  3,    84) =   57.46
       Model |  617130.701     3  205710.234           Prob > F      =  0.0000
    Residual |  300723.805    84   3580.0453           R-squared     =  0.6724
-------------+------------------------------           Adj R-squared =  0.6607
       Total |  917854.506    87  10550.0518           Root MSE      =  59.833

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      bdrms0 |   13.85252   9.010145     1.54   0.128    -4.065141    31.77018
      sqrft0 |   .1227782   .0132374     9.28   0.000     .0964541    .1491022
    lotsize0 |   .0020677   .0006421     3.22   0.002     .0007908    .0033446
       _cons |   222.5639   8.768905    25.38   0.000      205.126    240.0018
------------------------------------------------------------------------------
```

Prediction for the mean value of all houses with bdrms=3, sqrft=1500, and lotsize=9000:

$\hat{E} = 222.56$ with $se(\hat{E}) = 8.77$

To build the CI for the true value of E, recall that $\frac{\hat{E}-E}{se(\hat{E})}$ follow a t-distribution with 84 degrees of freedom. The critical value for the 95% confidence interval is 2 (between 1.987 and 2.000 to be exact).

The CI for E is thus $222.56 \ \pm 2\ (8.77) \ \approx [205{,}240\}$. We predict the mean price to be between \$205,000 and \$240,000.

Prediction for the value of a specific house with bdrms=3, sqrft=1500, and lotsize=9000:

$\hat{p} = 222.56$ with a variance $var(\hat{p}) = var(\hat{E}) + var(u)$ which we can estimate by $se^2(\hat{p}) = 8.77^2 + 3580$

This gives $se(\hat{p}) = 60.5$

To build the CI for the true value of p, we also use a t-distribution with 84 degrees of freedom.

The CI for p is thus $222.56 \ \pm 2\ (60.5) \ \approx [101{,}343\}$. The price of a house with these characteristics is between \$101,000 and \$343,000.

3. Regression with endogenous variable in log

```
. reg lprice bdrms sqrft lotsize

      Source |       SS       df       MS              Number of obs =      88
-------------+------------------------------           F(  3,    84) =   46.13
       Model |  4.98917377     3  1.66305792           Prob > F      =  0.0000
    Residual |  3.02842975    84  .036052735           R-squared     =  0.6223
-------------+------------------------------           Adj R-squared =  0.6088
       Total |  8.01760352    87  .092156362           Root MSE      =  .18988

------------------------------------------------------------------------------
      lprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       bdrms |   .0252388   .0285928     0.88   0.380    -.0316211    .0820987
       sqrft |   .0003641    .000042     8.67   0.000     .0002806    .0004477
     lotsize |   5.60e-06   2.04e-06     2.75   0.007     1.55e-06    9.65e-06
       _cons |   4.759375   .0935361    50.88   0.000     4.573369    4.945382
------------------------------------------------------------------------------
```

3.1. Prediction of average price for different values of bdrms sqrft lotsize
Assume residuals are normal

$\log y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u$ gives $E(\log y \mid x) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$

and $E(y \mid x) = e^{\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k} e^{\sigma_u^2/2}$

For prediction, we use: $\hat{y} = e^{\hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_k x_k} e^{\hat{\sigma}_u^2/2}$

```
. reg lprice bdrms0 sqrft0 lotsize0

      Source |       SS           df       MS      Number of obs   =        88
-------------+----------------------------------   F(3, 84)        =     46.13
       Model |  4.98917377         3  1.66305792   Prob > F        =    0.0000
    Residual |  3.02842975        84  .036052735   R-squared       =    0.6223
-------------+----------------------------------   Adj R-squared   =    0.6088
       Total |  8.01760352        87  .092156362   Root MSE        =    .18988

------------------------------------------------------------------------------
      lprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      bdrms0 |   .0252388   .0285928     0.88   0.380    -.0316211    .0820987
      sqrft0 |   .0003641    .000042     8.67   0.000     .0002806    .0004477
    lotsize0 |   5.60e-06   2.04e-06     2.75   0.007     1.55e-06    9.65e-06
       _cons |   5.431684   .0278272   195.19   0.000     5.376346    5.487022
------------------------------------------------------------------------------
```

3.2. Choosing between regression on price or on log(price) : Use $\left[\text{cor}(y,\hat{y})\right]^2$

1) First do the regression on lprice (above) and get the prediction for all prices based on the previous method

```
. qui reg lprice bdrms sqrft lotsize
. predict lpricehat
. g pricehat=exp(lpricehat)*exp(.036052735/2)

. sum price pricehat

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       price |        88     293.546    102.7134        111        725
    pricehat |        88    293.8179    84.32516   199.3371   608.2882


. correl price pricehat
(obs=88)

             |    price pricehat
-------------+------------------
       price |   1.0000
    pricehat |   0.8372   1.0000
```

R2 = $(0.8372)^2$ = .70090384

2) Now compare with the R2 obtained in the linear regression in section 1: 0.6724

Recall that R2 in the linear model is also equal to $\left[\text{cor}(y,\hat{y})\right]^2$

What do you conclude?