Example of Log-Linear Functional Form

The data set wage.dta contains hourly pay data for a sample of 379 men alongside their years of work experience. The idea is to test the relationship between pay and experience and decide whether a linear or log-linear specification works best.

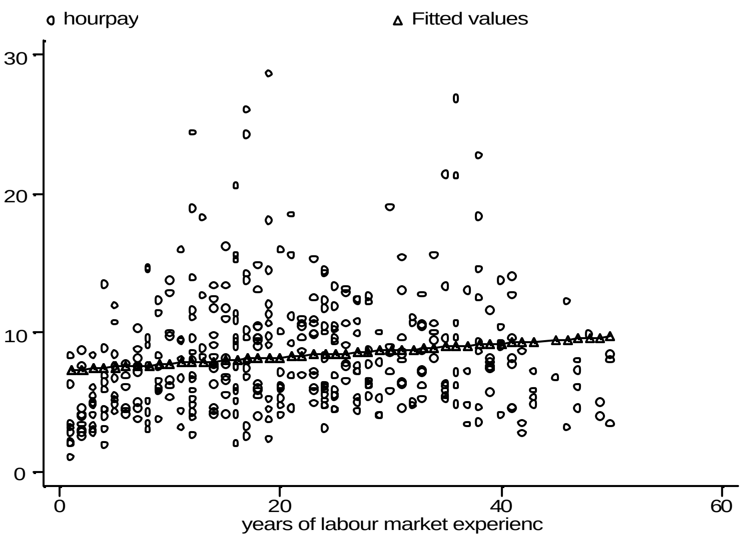
For the linear regression $\text{Hourpay} = b_0 + b_1\text{Experience} + u$

```
. reg hourpay xper

  Source |       SS       df       MS              Number of obs =     379
---------+------------------------------           F(  1,   377) =    7.53
   Model |  136.061219     1  136.061219           Prob > F      =  0.0064
Residual |  6815.41926   377  18.0780352           R-squared     =  0.0196
---------+------------------------------           Adj R-squared =  0.0170
   Total |  6951.48048   378    18.39016           Root MSE      =  4.2518

------------------------------------------------------------------------------
 hourpay |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    xper |   .0487259    .017761      2.743   0.006       .0138028     .083649
   _cons |    7.26455   .4333534     16.764   0.000       6.412457    8.116642
------------------------------------------------------------------------------

. predict reshat, resid       /* stata command "predict" followed by a name of
your choice followed by comma and command stata "resid"           */

. predict what
```

Relationship is statistically significant (1 extra year of experience adds 4.9 pence to the hourly wage)

Graph of fitted and actual values suggests may be over predicting at low levels of experience – would non-linear model fit better?

```
. gra hourpay what xper, xlab ylab c(.m)
```



Try log-linear model Log(Hourpay) = $b_0 + b_1$Log(Experience) + v

(Remember if b_1 is between 0 and 1 this suggests pay and experience positively related but that pay rises at a diminishing rate as experience increases – see lecture notes)

```
. g lpay=log(hourpay)
. g lexp=log(xper)

. reg lpay lexp

  Source |       SS       df       MS              Number of obs =     379
---------+------------------------------           F(  1,   377) =   46.03
   Model |  10.4791822     1  10.4791822           Prob > F      =  0.0000
Residual |  85.8187352   377  .227635902           R-squared     =  0.1088
---------+------------------------------           Adj R-squared =  0.1065
   Total |  96.2979174   378  .254756395           Root MSE      =  .47711

------------------------------------------------------------------------------
    lpay |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    lexp |   .1956242   .0288323      6.785   0.000       .1389319    .2523165
   _cons |   1.446405   .0840216     17.215   0.000       1.281195    1.611614
------------------------------------------------------------------------------
```

Coefficients are now **elasticities**
since dLnPay/dLnExp = b_1 = % change in pay wrt % change in price

So model suggests a 1% rise in experience increases hourly pay by 0.196%

```
. predict wlhat
. predict reslhat, resid
```

How big is 1% ? - look at sample means

```
. su hourpay xper
    Variable |     Obs        Mean   Std. Dev.       Min        Max
     hourpay |     379    8.291393   4.288375          1   28.63077
        xper |     379    21.07388   12.31293          1         50
```

(1% of experience is around 0.2 years 1% of pay is around 8 pence)

ie overall effect does not appear to be large

To compare predictions from 2 models (make units of measurement the same) take exponent of predicted value – converts logs into levels.

```
. g lhat=exp(wlhat)                /* stata command "exp" for exponent */

. gra hourpay what lhat xper, c(.mm) xlab ylab
```

Can see log specification runs closer to mass of data at lower levels of experience. Suggests log-linear may be better fit.