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
```

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 379</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>136.061219</td>
<td>1</td>
<td>136.061219</td>
<td>F( 1, 377) = 7.53</td>
</tr>
<tr>
<td>Residual</td>
<td>6815.41926</td>
<td>377</td>
<td>18.0780352</td>
<td>Prob &gt; F = 0.0064</td>
</tr>
<tr>
<td>Total</td>
<td>6951.48048</td>
<td>378</td>
<td>18.39016</td>
<td>Adj R-squared = 0.0170</td>
</tr>
</tbody>
</table>

| hourpay | Coef.  | Std. Err. | t     | P>|t|  | [95% Conf. Interval] |
|---------|--------|-----------|-------|-----|----------------------|
| xper    | 0.0487259 | 0.017761  | 2.743 | 0.006 | 0.0138028 – 0.083649 |
| _cons   | 7.26455  | 0.4333534 | 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” */

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 \[ \log(\text{Hourpay}) = b_0 + b_1 \log(\text{Experience}) + v \]

(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)

\[ g \text{lpay}=\log(\text{hourpay}) \]
\[ g \text{lexp}=\log(\text{xper}) \]

\[ . \text{reg lpay lexp} \]

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 379</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>10.479</td>
<td>1</td>
<td>10.479</td>
<td>F( 1, 377) = 46.03</td>
</tr>
<tr>
<td>Residual</td>
<td>85.819</td>
<td>377</td>
<td>.228</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>96.297</td>
<td>378</td>
<td>.255</td>
<td>R-squared = 0.108</td>
</tr>
</tbody>
</table>

| lpay   | Coef.  | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|--------|--------|-----------|-------|-----|----------------------|
| lexp   | .196   | .029      | 6.8   | 0.000 | .139                 |
| _cons | 1.446  | .084      | 17.2  | 0.000 | 1.281                 |

Coefficients are now elasticities since \(\frac{d\log(\text{Pay})}{d\log(\text{Exp})} = b_1\) = % change in pay wrt % change in price

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

\[ . \text{predict what} \]
\[ . \text{predict resihat, resid} \]

How big is 1%? - look at sample means

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>hourpay</td>
<td>379</td>
<td>8.29</td>
<td>4.29</td>
<td>1</td>
<td>28.63</td>
</tr>
<tr>
<td>xper</td>
<td>379</td>
<td>21.07</td>
<td>12.31</td>
<td>1</td>
<td>50</td>
</tr>
</tbody>
</table>

(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.