Solution Manual for Econometrics by Example 2nd Edition by Gujarati ISBN 1137375019 9781137375018 Full link download Solution Manual:

CHAPTER 2 Functional Forms of Regression Models

2.1. Consider the following production function, known in the literature as the transcendental production function (TPF).

$$Q_i = B_1 L_i^{B_2} K_i^{B_3} e^{B_4 L_i + B_5 K_i}$$

where *Q*, *L* and *K* represent output, labor and capital, respectively.

(*a*) How would you linearize this function? (Hint: logarithms.)

Taking the natural log of both sides, the transcendental production function above can be written in linear form as:

$$\ln Q_i = \ln B_1 + B_2 \ln L_i + B_3 \ln K_i + B_4 L_i + B_5 K_i + u_i$$

(*b*) What is the interpretation of the various coefficients in the TPF?

The coefficients may be interpreted as follows:

ln B_1 is the y-intercept, which may not have any viable economic interpretation, although B_1 may be interpreted as a technology constant in the Cobb-Douglas production function.

The elasticity of output with respect to labor may be interpreted as (B_2 + B_4*L). This is because

$\frac{\partial \ln Q_i}{\partial \ln L_i} = B_2 + B_4 \frac{1}{\frac{1}{L}} = B_2 + B_4 L$. Recall that $\frac{\partial \ln Q_i}{\partial \ln L_i} = \frac{\partial \ln Q_i}{(\frac{1}{L})\partial L_i}$.

Similarly, the elasticity of output with respect to capital can be expressed as (B_3 + B_5*K).

(*c*) Given the data in Table 2.1, estimate the parameters of the TPF.

The parameters of the transcendental production function are given in the following Stata output:

```
. reg lnoutput lnlabor lncapital labor capital

      Source |       SS       df       MS              Number of obs =      51
-------------+------------------------------           F(  4,     46) =  312.65
       Model |    91.95773     4  22.9894325           Prob > F      =  0.0000
    Residual |  3.38240102    46  .073530457           R-squared     =  0.9645
-------------+------------------------------           Adj R-squared =  0.9614
       Total |   95.340131    50  1.90680262           Root MSE      =  .27116

------------------------------------------------------------------------------
    lnoutput |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lnlabor |   .5208141   .1347469     3.87   0.000      .2495826    .7920456
   lncapital |   .4717828   .1231899     3.83   0.000      .2238144    .7197511
       labor |  -2.52e-07   4.20e-07    -0.60   0.552     -1.10e-06    5.94e-07
     capital |   3.55e-08   5.30e-08     0.67   0.506     -7.11e-08    1.42e-07
       _cons |   3.949841   .5660371     6.98   0.000      2.810468    5.089215
------------------------------------------------------------------------------
```

$B_1 = e^{3.949841} = 51.9271$.

B_2 = 0.5208141

B_3 = 0.4717828

B_4 = -2.52e-07

B_5 = 3.55e-08

Evaluated at the mean value of labor (373,914.5), the elasticity of output with respect to labor is 0.4266.

Evaluated at the mean value of capital (2,516,181), the elasticity of output with respect to capital is 0.5612.

(*d*) Suppose you want to test the hypothesis that $B_4 = B_5 = 0$. How would you test these hypotheses? Show the necessary calculations. (Hint: restricted least squares.)

I would conduct an F test for the coefficients on labor and capital. The output in Stata for this test gives the following:

```
. test   labor capital

( 1)   labor = 0
( 2)   capital = 0

F(  2,     46) =     0.23
Prob > F =    0.7992
```

This shows that the null hypothesis of $B_4 = B_5 = 0$ cannot be rejected in favor of the alternative hypothesis of $B_4 \neq B_5 \neq 0$. We may thus question the choice of using a transcendental production function over a standard Cobb-Douglas production function.

We can also use restricted least squares and perform this calculation "by hand" by conducting an *F* test as follows:

$$F = \frac{(RSS_R - RSS_{UR})/(n-k+2-n+k)}{RSS_{UR}/(n-k)} \sim F_{2,46}$$

The restricted regression is:

$$\ln Q_i = \ln B_1 + B_2 \ln L_i + B_3 \ln K_i + u_i,$$

which gives the following Stata output:

```
. reg lnoutput lnlabor lncapital;

      Source |       SS       df       MS              Number of obs =      51
-------------+------------------------------           F(  2,     48) =  645.93
       Model |  91.9246133     2  45.9623067           Prob > F      =  0.0000
    Residual |  3.41551772    48  .071156619           R-squared     =  0.9642
-------------+------------------------------           Adj R-squared =  0.9627
       Total |   95.340131    50  1.90680262           Root MSE      =  .26675

------------------------------------------------------------------------------
    lnoutput |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lnlabor |   .4683318   .0989259     4.73   0.000       .269428   .6672357
   lncapital |   .5212795    .096887     5.38   0.000       .326475   .7160839
       _cons |   3.887599   .3962281     9.81   0.000      3.090929   4.684269
------------------------------------------------------------------------------
```

The unrestricted regression is the original one shown in 2(c). This gives the following:

$$F = \frac{(3.4155177 - 3.382401)/(51-5+2-51+5)}{3.382401/(51-5)} = 0.22519 \sim F_{2,46}$$

Since 0.225 is less than the critical *F* value of 3.23 for 2 degrees of freedom in the numerator and 40 degrees in the denominator (rounded using statistical tables), we cannot reject the null hypothesis of $B_4 = B_5 = 0$ at the 5% level.

(*e*) How would you compute the output-labor and output capital elasticities for this model? Are they constant or variable?

See answers to 2(b) and 2(c) above. Since the values of L and K are used in computing the elasticities, they are *variable*.

2.2. How would you compute the output-labor and output-capital elasticities for the linear production function given in Table 2.3?

The Stata output for the linear production function given in Table 2.3 is:

```
. reg output labor capital

      Source |       SS       df       MS              Number of obs =      51
-------------+------------------------------           F(  2,    48) = 1243.51
       Model | 9.8732e+16     2 4.9366e+16             Prob > F      =  0.0000
    Residual | 1.9055e+15    48 3.9699e+13             R-squared     =  0.9811
-------------+------------------------------           Adj R-squared =  0.9803
       Total | 1.0064e+17    50 2.0127e+15             Root MSE      =  6.3e+06

------------------------------------------------------------------------------
      output |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       labor |   47.98736   7.058245     6.80   0.000      33.7958    62.17891
     capital |   9.951891   .9781165    10.17   0.000     7.985256    11.91853
       _cons |   233621.6    1250364     0.19   0.853     -2280404     2747648
------------------------------------------------------------------------------
```

The elasticity of output with respect to labor is: $\frac{\partial Q_i / Q_i}{\partial L_i / L_i} = B_2 \frac{L}{Q}$.

It is often useful to compute this value at the mean. Therefore, evaluated at the mean values of labor and output, the output-labor elasticity is: $B_2 \frac{\bar{L}}{\bar{Q}} = 47.98736 \frac{373914.5}{4.32e+07} = 0.41535$.

Similarly, the elasticity of output with respect to capital is: $\frac{\partial Q_i / Q_i}{\partial K_i / K_i} = B_3 \frac{K}{Q}$.

Evaluated at the mean, the output-capital elasticity is: $B_3 \frac{\bar{K}}{\bar{Q}} = 9.951891 \frac{2516181}{4.32e+07} = 0.57965$.

2.3. For the food expenditure data given in Table 2.8, see if the following model fits the data well:

$$\mathbf{SFDHO_i = B_1 + B_2\ Expend_i + B_3\ Expend_i^2}$$

and compare your results with those discussed in the text.

The Stata output for this model gives the following:

```
. reg sfdho expend expend2

      Source |       SS       df       MS              Number of obs =     869
-------------+------------------------------           F(  2,   866) =  204.68
       Model |  2.02638253     2  1.01319127           Prob > F      =  0.0000
```

Residual	4.28671335	866	.004950015		R-squared	= 0.3210
					Adj R-squared	= 0.3194
Total	6.31309589	868	.007273152		Root MSE	= .07036

sfdho	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
expend	-5.10e-06	3.36e-07	-15.19	0.000	-5.76e-06	-4.44e-06
expend2	3.23e-11	3.49e-12	9.25	0.000	2.54e-11	3.91e-11
_cons	.2563351	.0065842	38.93	0.000	.2434123	.2692579

Similarly to the results in the text (shown in Tables 2.9 and 2.10), these results show a strong nonlinear relationship between share of food expenditure and total expenditure. Both total expenditure and its square are highly significant. The negative sign on the coefficient on "expend" combined with the positive sign on the coefficient on "expend2" implies that the share of food expenditure in total expenditure is *decreasing* at an *increasing* rate, which is precisely what the plotted data in Figure 2.3 show.

The R^2 value of 0.3210 is only slightly lower than the R^2 values of 0.3509 and 0.3332 for the lin-log and reciprocal models, respectively. (As noted in the text, we are able to compare R^2 values across these models since the dependent variable is the same.)

2.4 Would it make sense to standardize variables in the log-linear Cobb-Douglas production function and estimate the regression using standardized variables? Why or why not? Show the necessary calculations.

This would mean standardizing the natural logs of *Y*, *K*, and *L*. Since the coefficients in a log-linear (or double-log) production function already represent unit-free changes, this may not be necessary. Moreover, it is easier to interpret a coefficient in a log linear model as an elasticity. If we were to standardize, the coefficients would represent percentage changes in the standard deviation units. Standardizing would reveal, however, whether capital or labor contributes more to output.

2.5. Show that the coefficient of determination, R^2, can also be obtained as the squared correlation between actual *Y* values and the *Y* values estimated from the regression model (= $\hat{Y}_i$), where *Y* is the dependent variable. Note that the coefficient of correlation between variables *Y* and *X* is defined as:

$$r = \frac{\sum y_i x_i}{\sqrt{\sum x_i^2 \sum y_i^2}}$$

where $y_i = Y_i - \bar{Y}$; $x_i = X_i - \bar{X}$. Also note that the mean values of Y_i and $\hat{Y}$ are the same, namely, $\bar{Y}$.

The estimated Y values from the regression can be rewritten as: $\hat{Y}_i = B_1 + B_2 X_i$.

Taking deviations from the mean, we have: $\hat{y}_i = B_2 x_i$.

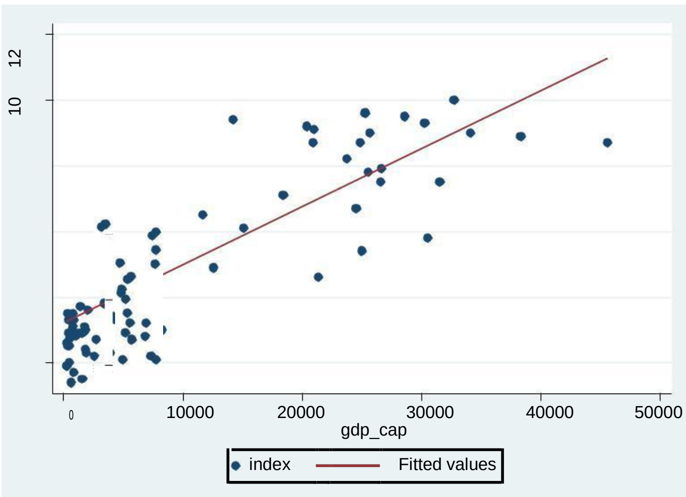
Therefore, the squared correlation between actual Y values and the Y values estimated from the regression model is represented by:

$$r = \frac{\sum y_i \hat{y}_i}{\sqrt{\sum y_i^2 \sum \hat{y}_i^2}} = \frac{\sum y_i (B_2 x_i)}{\sqrt{\sum y_i^2 \sum (B_2 x_i)^2}} = \frac{B_2 \sum y_i x_i}{B_2 \sqrt{\sum y_i^2 \sum x_i^2}} = \frac{\sum y_i x_i}{\sqrt{\sum y_i^2 \sum x_i^2}},$$

which is the coefficient of correlation. If this is squared, we obtain the coefficient of determination, or R^2.

2.6. Table 2.18 gives cross-country data for 83 countries on per worker GDP and Corruption Index for 1998.

(*a*) Plot the index of corruption against per worker GDP.



(*b*) Based on this plot what might be an appropriate model relating corruption index to per worker GDP?

A slightly nonlinear relationship may be appropriate, as it looks as though corruption may increase at a decreasing rate with increasing GDP per capita.

(*c*) Present the results of your analysis.

Results are as follows:

```
. reg  index gdp_cap  gdp_cap2

      Source |       SS       df       MS              Number of obs =      83
-------------+------------------------------           F(  2,    80)  = 126.61
       Model |    365.6695     2   182.83475           Prob > F       = 0.0000
    Residual |  115.528569    80  1.44410711           R-squared      = 0.7599
-------------+------------------------------           Adj R-squared =  0.7539
       Total |  481.198069    82  5.86826913           Root MSE       = 1.2017

------------------------------------------------------------------------------
       index  |      Coef.  Std. Err.       t    P>|t|     [95% Conf. Interval]
-------------------------------------------------------------------------------
     gdp_cap |   .0003182  .0000393      8.09   0.000      .0002399    .0003964
    gdp_cap2 |  -4.33e-09  1.15e-09     -3.76   0.000     -6.61e-09   -2.04e-09
       _cons |   2.845553  .1983219     14.35   0.000      2.450879    3.240226
------------------------------------------------------------------------------
```

(*d*) If you find a positive relationship between corruption and per capita GDP, how would you rationalize this outcome?

We find a perhaps unexpected positive relationship because of the way corruption is defined. As the Transparency International website states, “Since 1995 Transparency International has published each year the CPI, ranking countries on a scale from 0 (perceived to be highly corrupt) to 10 (perceived to have low levels of corruption).” This means that *higher* values for the corruption index indicate *less* corruption. Therefore, countries with higher GDP per capita have lower levels of corruption.

2.7 Table 2.19 gives fertility and other related data for 64 countries. Develop suitable model(s) to explain child mortality, considering the various function forms and the measures of goodness of fit discussed in the chapter.

The following is a linear model explaining child mortality as a function of the female literacy rate, per capita GNP, and the total fertility rate:

```
. reg  cm flr pgnp tfr

      Source |       SS       df       MS              Number of obs =      64
-------------+------------------------------           F(  3,    60) =   59.17
       Model |  271802.616     3  90600.8721           Prob > F      = 0.0000
    Residual |  91875.3836    60  1531.25639           R-squared     = 0.7474
-------------+------------------------------           Adj R-squared =  0.7347
       Total |      363678    63  5772.66667           Root MSE      = 39.131

------------------------------------------------------------------------------
          cm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         flr |  -1.768029  .2480169    -7.13   0.000     -2.264137   -1.271921
        pgnp |  -.0055112  .0018782    -2.93   0.005     -.0092682   -.0017542
         tfr |   12.86864  4.190533     3.07   0.003      4.486323    21.25095
       _cons |   168.3067  32.89166     5.12   0.000      102.5136    234.0998
------------------------------------------------------------------------------
```

The results suggest that higher rates of female literacy and per capita GNP reduce child mortality, which one would expect. Moreover, as the fertility rate goes up, one might expect child mortality to go up, which we see. All results are statistically significant at the 1% level, and the value of r-squared is quite high at 0.7474.

2.8: Verify Equations (2.35), (2.36) and (2.37). Hint: Minimize:

$$\sum u_i^2 = \sum (Y_i - B_2 X)^2$$

$$R_i - r_f = \beta_i (R_m - r_f) + u_i \qquad \textbf{(2.35)}$$

$$Y_i = B_2 X_i + u_i \qquad \textbf{(2.36)}$$

$$b_2 = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2} \qquad \textbf{(2.37)}$$

$$\text{var}(b_2) = \frac{\sigma^2}{\sum_{i=1} X_i^2} \qquad \textbf{(2.38)}$$

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-1} \qquad \textbf{(2.39)}$$

We move from equation 2.35 to 2.36 by definition. (We have definied Y as $R - r_f$ and X as $R_m - r_f$.) There is no intercept in this model. Because of that, we can see that, in minimizing the sum of u_i^2 with respect to B_2 and setting the equation equal to zero, we obtain equation 2.37: (In this case, there is only one equation and one unknown.)

$$\frac{d\sum u_i^2}{dB_2} = -\sum X_i(Y - B_2X) = 0$$

$$\sum XY - B_2\sum X^2 = 0$$

$$\sum XY = B_2\sum X^2$$

$$B_2 = \frac{\sum XY}{\sum X^2}$$

2.9: Consider the following model without any regressors.

$$Y_i = B_1 + u_i$$

How would you obtain an estimate of B_1? What is the meaning of the estimated value? Does it make any sense?

If you have a model without regressors, B_1 simply gives you the average value of Y. We can see this by using the data in Table 2.19 (from Exercise 2.7) and running a regression of with only a "dependent" variable, child mortality:

```
. reg cm

      Source |       SS       df       MS              Number of obs =      64
-------------+------------------------------           F(  0,     63) =    0.00
       Model |           0     0           .           Prob > F      =       .
    Residual |      363678    63  5772.66667           R-squared     =  0.0000
-------------+------------------------------           Adj R-squared =  0.0000
       Total |      363678    63  5772.66667           Root MSE      =  75.978

------------------------------------------------------------------------------
          cm |      Coef.   Std. Err.       t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |      141.5   9.497258     14.90   0.000     122.5212    160.4788
------------------------------------------------------------------------------
```

This is clearly not very useful and does not make much sense. B_1, the intercept, gives you the mean value of child mortality. Summarizing this variable would give us the same value:

```
. su cm

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
          cm |        64       141.5    75.97807         12         31
```