A study guide produced by Christopher Dougherty to accompany the module "EC2020 Elements of Econometrics" offered as part of the University of London International Programmes in Economics, Management, Finance, and the Social Sciences.

Chapter 3
Multiple regression analysis

3.1 Overview

This chapter introduces regression models with more than one explanatory variable. Specific topics are treated with reference to a model with just two explanatory variables, but most of the concepts and results apply straightforwardly to more general models. The chapter begins by showing how the least squares principle is employed to derive the expressions for the regression coefficients and how the coefficients should be interpreted. It continues with a discussion of the precision of the regression coefficients and tests of hypotheses relating to them. Next comes multicollinearity, the problem of discriminating between the effects of individual explanatory variables when they are closely related. The chapter concludes with a discussion of F tests of the joint explanatory power of the explanatory variables or subsets of them, and shows how a t test can be thought of as a marginal F test.

3.2 Learning outcomes

After working through the corresponding chapter in the text, studying the corresponding slideshows, and doing the starred exercises in the text and the additional exercises in this subject guide, you should be able to explain what is meant by:

- the principles behind the derivation of multiple regression coefficients (but you are not expected to learn the expressions for them or to be able to reproduce the mathematical proofs)
- how to interpret the regression coefficients
- the Frisch–Waugh–Lovell graphical representation of the relationship between the dependent variable and one explanatory variable, controlling for the influence of the other explanatory variables
- the properties of the multiple regression coefficients
- what factors determine the population variance of the regression coefficients
- what is meant by multicollinearity
- what measures may be appropriate for alleviating multicollinearity
- what is meant by a linear restriction
- the F test of the joint explanatory power of the explanatory variables

© Christopher Dougherty, 2016. All rights reserved.
Published on the Online Resource Centre to accompany Dougherty: Introduction to Econometrics, 5th edition, by Oxford University Press.

A study guide produced by Christopher Dougherty to accompany the module "EC2020 Elements of Econometrics" offered as part of the University of London International Programmes in Economics, Management, Finance, and the Social Sciences.

- the F test of the explanatory power of a group of explanatory variables
- why t tests on the slope coefficients are equivalent to marginal F tests.

You should know the expression for the population variance of a slope coefficient in a multiple regression model with two explanatory variables.

3.3 Additional exercises

A3.1 The output shows the result of regressing *FDHO*, expenditure on food consumed at home, on *EXP*, total household expenditure, and *SIZE*, number of persons in the household, using the *CES* data set. Provide an interpretation of the regression coefficients and perform appropriate tests.

```
. reg FDHO EXP SIZE if FDHO>0

      Source |       SS       df       MS              Number of obs =    6334
-------------+------------------------------           F(  2,  6331) = 2257.59
       Model |  1.1521e+09     2   576056293           Prob > F      =  0.0000
    Residual |  1.6154e+09  6331  255164.645           R-squared     =  0.4163
-------------+------------------------------           Adj R-squared =  0.4161
       Total |  2.7676e+09  6333   437006.15           Root MSE      =  505.14

------------------------------------------------------------------------------
        FDHO |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         EXP |    .056366   .0010435    54.02   0.000     .0543204    .0584116
        SIZE |   115.1636   4.341912    26.52   0.000      106.652    123.6752
       _cons |   130.5997   13.53959     9.65   0.000     104.0575    157.1419
------------------------------------------------------------------------------
```

A3.2 Perform a regression parallel to that in Exercise A3.1 for your *CES* category of expenditure, provide an interpretation of the regression coefficients and perform appropriate tests. Delete observations where expenditure on your category is zero.

A3.3 The output shows the result of regressing *FDHOPC*, expenditure on food consumed at home per capita, on *EXPPC*, total household expenditure per capita, and *SIZE*, number of persons in the household, using the *CES* data set. Provide an interpretation of the regression coefficients and perform appropriate tests.

```
. reg FDHOPC EXPPC SIZE if FDHO>0

      Source |       SS       df       MS              Number of obs =    6334
-------------+------------------------------           F(  2,  6331) = 1572.95
       Model |   202590496     2   101295248           Prob > F      =  0.0000
    Residual |   407705728  6331  64398.3143           R-squared     =  0.3320
-------------+------------------------------           Adj R-squared =  0.3317
       Total |   610296223  6333  96367.6336           Root MSE      =  253.77
```


© Christopher Dougherty, 2016. All rights reserved.
Published on the Online Resource Centre to accompany Dougherty: Introduction to Econometrics, 5th edition, by Oxford University Press.

A study guide produced by Christopher Dougherty to accompany the module "EC2020 Elements of Econometrics" offered as part of the University of London International Programmes in Economics, Management, Finance, and the Social Sciences.


```
------------------------------------------------------------------------------
      FDHOPC |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       EXPPC |   .0480294   .0010064    47.72   0.000     .0460564    .0500023
        SIZE |  -26.45917   2.253999   -11.74   0.000    -30.87777   -22.04057
       _cons |   283.2498   8.412603    33.67   0.000     266.7582    299.7413
------------------------------------------------------------------------------
```

A3.4 Perform a regression parallel to that in Exercise A3.3 for your *CES* category of expenditure. Provide an interpretation of the regression coefficients and perform appropriate tests.

A3.5 The output shows the result of regressing *FDHOPC*, expenditure on food consumed at home per capita, on *EXPPC*, total household expenditure per capita, and *SIZEAM*, *SIZEAF*, *SIZEJM*, *SIZEJF*, and *SIZEIN*, numbers of adult males, adult females, junior males, junior females, and infants, respectively, in the household, using the *CES* data set. Provide an interpretation of the regression coefficients and perform appropriate tests.

```
. reg FDHOPC EXPPC SIZEAM SIZEAF SIZEJM SIZEJF SIZEIN if FDHO>0

      Source |       SS       df       MS              Number of obs =    6334
-------------+------------------------------           F(  6,  6327) =  524.59
       Model |  202746894     6    33791149            Prob > F      =  0.0000
    Residual |  407549329  6327  64414.3084            R-squared     =  0.3322
-------------+------------------------------           Adj R-squared =  0.3316
       Total |  610296223  6333  96367.6336            Root MSE      =   253.8

------------------------------------------------------------------------------
      FDHOPC |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       EXPPC |   .0479717   .0010087    47.56   0.000     .0459943    .0499491
      SIZEAM |  -25.77747   4.757056    -5.42   0.000    -35.10291   -16.45203
      SIZEAF |  -32.38649   5.065782    -6.39   0.000    -42.31714   -22.45584
      SIZEJM |  -20.24693   5.731645    -3.53   0.000     -31.4829   -9.010967
      SIZEJF |  -26.66374   6.122262    -4.36   0.000    -38.66544   -14.66203
      SIZEIN |   -28.6047   11.75666    -2.43   0.015    -51.65174   -5.557656
       _cons |   287.5695   9.280372    30.99   0.000     269.3769    305.7622
------------------------------------------------------------------------------
```

A3.6 Perform a regression parallel to that in Exercise A3.5 for your *CES* category of expenditure. Provide an interpretation of the regression coefficients and perform appropriate tests.

A3.7 A researcher hypothesises that, for a typical enterprise, V, the logarithm of value added per worker, is related to K, the logarithm of capital per worker, and S, the logarithm of the average years of schooling of the workers, the relationship being:

$$V = \beta_1 + \beta_2 K + \beta_3 S + u$$

where u is a disturbance term that satisfies the usual regression model assumptions. She fits the relationship (1) for a sample of 25 manufacturing enterprises, and (2) for a sample of 100 services enterprises. The table provides some data on the samples.

© Christopher Dougherty, 2016. All rights reserved.
Published on the Online Resource Centre to accompany Dougherty: Introduction to Econometrics, 5th edition, by Oxford University Press.

A study guide produced by Christopher Dougherty to accompany the module "EC2020 Elements of Econometrics" offered as part of the University of London International Programmes in Economics, Management, Finance, and the Social Sciences.

	(1) Manufacturing sample	(2) Services sample
Number of enterprises	25	100
Estimate of variance of u	0.16	0.64
Mean square deviation of K	4.00	16.00
Correlation between K and S	0.60	0.60

The mean square deviation of K is defined as $\frac{1}{n}\sum\left(K_i - \overline{K}\right)^2$, where n is the number of enterprises in the sample and $\overline{K}$ is the average value of K in the sample.

The researcher finds that the standard error of the coefficient of K is 0.050 for the manufacturing sample and 0.025 for the services sample. Explain the difference quantitatively, given the data in the table.

A3.8 A researcher is fitting earnings functions using a sample of data relating to individuals born in the same week in 1958. He decides to relate Y, gross hourly earnings in 2001, to S, years of schooling, and PWE, potential work experience, using the semilogarithmic specification:

$$\log Y = \beta_1 + \beta_2 S + \beta_3 PWE + u$$

where u is a disturbance term assumed to satisfy the regression model assumptions. PWE is defined as age – years of schooling – 5. Since the respondents were all aged 43 in 2001, this becomes:

$$PWE = 43 - S - 5 = 38 - S.$$

The researcher finds that it is impossible to fit the model as specified. Stata output for his regression is reproduced below:

```
. reg LGY S PWE

      Source |       SS       df       MS              Number of obs =    5660
-------------+------------------------------           F(  1,  5658) = 1232.62
       Model |  237.170265     1  237.170265           Prob > F      =  0.0000
    Residual |  1088.66373  5658  .192411405           R-squared     =  0.1789
-------------+------------------------------           Adj R-squared =  0.1787
       Total |    1325.834  5659  .234287682           Root MSE      =  .43865

------------------------------------------------------------------------------
         LGY |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           S |   .1038011   .0029566    35.11   0.000     .0980051    .1095971
         PWE |  (dropped)
       _cons |   .5000033   .0373785    13.38   0.000     .4267271    .5732795
------------------------------------------------------------------------------
```

Explain why the researcher was unable to fit his specification.

Explain how the coefficient of S might be interpreted.

© Christopher Dougherty, 2016. All rights reserved.
Published on the Online Resource Centre to accompany Dougherty: Introduction to Econometrics, 5th edition, by Oxford University Press.

A study guide produced by Christopher Dougherty to accompany the module "EC2020 Elements of Econometrics" offered as part of the University of London International Programmes in Economics, Management, Finance, and the Social Sciences.

3.4 Answers to the starred exercises in the textbook

3.5 Explain why the intercept in the regression of $EEARN$ on ES is equal to zero.

Answer:

The intercept is calculated as $\overline{EEARN} - \widehat{\beta}_2\overline{ES}$. However, since the mean of the residuals from an OLS regression is zero, both $\overline{EEARN}$ and $\overline{ES}$ are zero, and hence the intercept is zero.

3.6 Show that, in the general case, the mean of the residuals from a fitted OLS multiple regression is equal to zero, provided that an intercept is included in the specification. Note: This is an extension of one of the useful results in Section 1.5.

Answer:

If the model is:

$$\begin{aligned} Y &= \beta_1 + \beta_2 X_2 + \cdots + \beta_k X_k + u \\ \widehat{\beta}_1 &= \overline{Y} - \widehat{\beta}_2\overline{X}_2 - \cdots - \widehat{\beta}_k\overline{X}_k. \end{aligned}$$

For observation i we have:

$$\widehat{u}_i = Y_i - \widehat{Y}_i = Y_i - \widehat{\beta}_1 - \widehat{\beta}_2 X_{2i} - \cdots - \widehat{\beta}_k X_{ki}.$$

Hence:

$$\begin{aligned} \overline{\widehat{u}} &= \overline{Y} - \widehat{\beta}_1 - \widehat{\beta}_2\overline{X}_2 - \cdots - \widehat{\beta}_k\overline{X}_k \\ &= \overline{Y} - \left[\overline{Y} - \widehat{\beta}_2\overline{X}_2 - \cdots - \widehat{\beta}_k\overline{X}_k\right] - \widehat{\beta}_2\overline{X}_2 - \cdots - \widehat{\beta}_k\overline{X}_k = 0. \end{aligned}$$

3.16 A researcher investigating the determinants of the demand for public transport in a certain city has the following data for 100 residents for the previous calendar year: expenditure on public transport, E, measured in dollars; number of days worked, W; and number of days not worked, NW. By definition NW is equal to $365 - W$. He attempts to fit the following model:

$$E = \beta_1 + \beta_2 W + \beta_3 NW + u.$$

Explain why he is unable to fit this equation. (Give both intuitive and technical explanations.) How might he resolve the problem?

Answer:

There is exact multicollinearity since there is an exact linear relationship between W, NW and the constant term. As a consequence it is not possible to tell whether variations in E are attributable to variations in W or variations in NW, or both. Noting that $NW_i - \overline{NW} = -(W_i - \overline{W})$, we have:

$$\begin{aligned} \widehat{\beta}_2 &= \frac{\sum\left(E_i-\overline{E}\right)\left(W_i-\overline{W}\right)\sum\left(NW_i-\overline{NW}\right)^2 - \sum\left(E_i-\overline{E}\right)\left(NW_i-\overline{NW}\right)\sum\left(W_i-\overline{W}\right)\left(NW_i-\overline{NW}\right)}{\sum\left(W_i-\overline{W}\right)^2\sum\left(NW_i-\overline{NW}\right)^2 - \left(\sum\left(W_i-\overline{W}\right)\left(NW_i-\overline{NW}\right)\right)^2} \\ &= \frac{\sum\left(E_i-\overline{E}\right)\left(W_i-\overline{W}\right)\sum\left(W_i-\overline{W}\right)^2 - \sum\left(E_i-\overline{E}\right)\left(-W_i+\overline{W}\right)\sum\left(W_i-\overline{W}\right)\left(-W_i+\overline{W}\right)}{\sum\left(W_i-\overline{W}\right)^2\sum\left(W_i-\overline{W}\right)^2 - \left(\sum\left(W_i-\overline{W}\right)\left(-W_i+\overline{W}\right)\right)^2} \\ &= \frac{0}{0}. \end{aligned}$$

© Christopher Dougherty, 2016. All rights reserved.
Published on the Online Resource Centre to accompany Dougherty: Introduction to Econometrics, 5th edition, by Oxford University Press.

A study guide produced by Christopher Dougherty to accompany the module "EC2020 Elements of Econometrics" offered as part of the University of London International Programmes in Economics, Management, Finance, and the Social Sciences.

One way of dealing with the problem would be to drop NW from the regression. The interpretation of $\widehat{\beta}_2$ now is that it is an estimate of the extra expenditure on transport per day worked, compared with expenditure per day not worked.

3.21 The researcher in Exercise 3.16 decides to divide the number of days not worked into the number of days not worked because of illness, I, and the number of days not worked for other reasons, O. The mean value of I in the sample is 2.1 and the mean value of O is 120.2. He fits the regression (standard errors in parentheses):

$$\begin{aligned} \widehat{E} &= \underset{(8.3)}{-9.6} + \underset{(1.98)}{2.10}W + \underset{(1.77)}{0.45}O \qquad R^2 = 0.72 \end{aligned}$$

Perform t tests on the regression coefficients and an F test on the goodness of fit of the equation. Explain why the t tests and F test have different outcomes.

Answer:

Although there is not an exact linear relationship between W and O, they must have a very high negative correlation because the mean value of I is so small. Hence one would expect the regression to be subject to multicollinearity, and this is confirmed by the results. The t statistics for the coefficients of W and O are only 1.06 and 0.25, respectively, but the F statistic:

$$F(2,97) = \frac{0.72/2}{(1-0.72)/97} = 124.7$$

is greater than the critical value of F at the 0.1 per cent level, 7.41.

3.5 Answers to the additional exercises

A3.1 The regression indicates that 5.6 cents out of the marginal expenditure dollar is spent on food consumed at home, and that expenditure on this category increases by $115 for each individual in the household, keeping total expenditure constant. Both of these effects are very highly significant. Just over 40 per cent of the variance in *FDHO* is explained by *EXP* and *SIZE*. The intercept has no plausible interpretation.

A3.2 With the exception of *LOCT*, all of the categories have positive coefficients for *EXP*, with high significance levels, but the *SIZE* effect varies:

- Positive, significant at the 1 per cent level: *FDHO*, *TELE*, *CLOT*, *FOOT*, *GASO*.
- Positive, significant at the 5 per cent level: *LOCT*.
- Negative, significant at the 1 per cent level: *TEXT*, *FEES*, *READ*.
- Negative, significant at the 5 per cent level: *SHEL*, *EDUC*.
- Not significant: *FDAW*, *DOM*, *FURN*, *MAPP*, *SAPP*, *TRIP*, *HEAL*, *ENT*, *TOYS*, *TOB*.

At first sight it may seem surprising that *SIZE* has a significant negative effect for some categories. The reason for this is that an increase in *SIZE* means a reduction

© Christopher Dougherty, 2016. All rights reserved.
Published on the Online Resource Centre to accompany Dougherty: Introduction to Econometrics, 5th edition, by Oxford University Press.

A study guide produced by Christopher Dougherty to accompany the module "EC2020 Elements of Econometrics" offered as part of the University of London International Programmes in Economics, Management, Finance, and the Social Sciences.

in expenditure per capita, if total household expenditure is kept constant, and thus *SIZE* has a (negative) income effect in addition to any direct effect. Effectively poorer, the larger household has to spend more on basics and less on luxuries. To determine the true direct effect, we need to eliminate the income effect, and that is the point of the re-specification of the model in the next exercise.

		EXP		*SIZE*			
	n	$\widehat{\beta}_2$	s.e.$(\widehat{\beta}_2)$	$\widehat{\beta}_3$	s.e.$(\widehat{\beta}_3)$	R^2	F
ADM	2,815	0.0238	0.0008	−8.09	4.19	0.230	418.7
CLOT	4,500	0.0309	0.0010	16.39	4.50	0.178	488.2
DOM	1,661	0.0388	0.0026	52.34	14.06	0.141	136.2
EDUC	561	0.1252	0.0090	−179.23	48.92	0.258	97.2
ELEC	5,828	0.0121	0.0004	18.92	1.57	0.199	725.5
FDAW	5,102	0.0538	0.0010	−20.72	4.47	0.357	1,413.7
FDHO	6,334	0.0564	0.0010	115.16	4.34	0.416	2,257.6
FOOT	1,827	0.0056	0.0005	3.24	2.05	0.083	83.0
FURN	487	0.0541	0.0071	−61.87	35.92	0.108	29.3
GASO	5,710	0.0347	0.0008	50.29	3.40	0.305	1,250.9
HEAL	4,802	0.0580	0.0019	−9.96	8.60	0.175	507.4
HOUS	6,223	0.1997	0.0027	−38.78	11.41	0.470	2,760.4
LIFE	1,253	0.0198	0.0017	−9.01	8.99	0.102	70.9
LOCT	692	0.0062	0.0011	14.61	4.72	0.072	26.8
MAPP	399	0.0309	0.0050	44.48	23.94	0.110	24.4
PERS	3,817	0.0070	0.0002	−2.17	1.03	0.214	519.4
READ	2,287	0.0049	0.0003	−1.06	1.58	0.104	132.7
SAPP	1,037	0.0046	0.0008	−3.12	3.99	0.035	18.5
TELE	5,788	0.0150	0.0004	17.92	1.47	0.287	1,161.2
TEXT	992	0.0041	0.0006	−0.71	2.90	0.051	26.8
TOB	1,155	0.0161	0.0016	6.79	6.24	0.089	56.4
TOYS	2,504	0.0140	0.0010	12.19	4.88	0.078	106.2
TRIP	516	0.0450	0.0045	37.48	31.21	0.188	59.5

A3.3 Another surprise, perhaps. The purpose of this specification is to test whether household size has an effect on expenditure per capita on food consumed at home, controlling for the income effect of variations in household size mentioned in the answer to Exercise A3.2. Expenditure per capita on food consumed at home increases by 4.8 cents out of the marginal dollar of total household expenditure per capita. Now *SIZE* has a very significant negative effect. Expenditure per capita on *FDHO* decreases by \$26 per year for each extra person in the household, suggesting that larger households are more efficient than smaller ones with regard to expenditure on this category, the effect being highly significant. R^2 is lower than in Exercise A3.1, but a comparison is invalidated by the fact that the dependent variable is different.

A3.4 Nearly all of the categories have negative *SIZE* effects, the majority highly significant. One explanation of the negative effects could be economies of scale, but

© Christopher Dougherty, 2016. All rights reserved.
Published on the Online Resource Centre to accompany Dougherty: Introduction to Econometrics, 5th edition, by Oxford University Press.

A study guide produced by Christopher Dougherty to accompany the module "EC2020 Elements of Econometrics" offered as part of the University of London International Programmes in Economics, Management, Finance, and the Social Sciences.

this is not plausible in the case of some. Another might be family composition – larger families having more children. In the case of *DOM*, *SIZE* has a positive effect, significant at the 5 per cent level. Again, this might be attributable to larger families having more children and needing greater expenditure on childcare.

		EXP		*SIZE*			
	n	$\widehat{\beta}_2$	s.e.($\widehat{\beta}_2$)	$\widehat{\beta}_3$	s.e.($\widehat{\beta}_3$)	R^2	F
ADM	2,815	0.0244	0.0008	2.56	2.26	0.251	470.4
CLOT	4,500	0.0324	0.0012	−1.07	2.91	0.151	400.8
DOM	1,661	0.0311	0.0025	18.54	7.35	0.086	78.1
EDUC	561	0.1391	0.0108	−31.92	27.57	0.290	113.7
ELEC	5,828	0.0117	0.0004	−17.53	0.89	0.247	953.9
FDAW	5,102	0.0528	0.0011	−13.51	2.53	0.375	1,526.3
FDHO	6,334	0.0480	0.0010	−26.46	2.25	0.332	1,573.0
FOOT	1,827	0.0068	0.0005	−8.13	1.11	0.194	219.5
FURN	487	0.0935	0.0091	3.40	26.82	0.216	66.6
GASO	5,710	0.0308	0.0008	−12.43	1.80	0.255	976.5
HEAL	4,802	0.0597	0.0020	−34.16	4.99	0.197	588.5
HOUS	6,223	0.2127	0.0030	−48.86	6.67	0.501	3,123.3
LIFE	1,253	0.0205	0.0017	−10.33	4.65	0.131	94.4
LOCT	692	0.0062	0.0010	−9.06	2.54	0.098	37.4
MAPP	399	0.0384	0.0051	−15.52	12.32	0.171	41.0
PERS	3,817	0.0071	0.0003	−3.96	0.63	0.228	564.0
READ	2,287	0.0052	0.0003	−3.60	0.84	0.154	208.1
SAPP	1,037	0.0076	0.0010	−6.71	2.61	0.090	51.1
TELE	5,788	0.0139	0.0003	−9.77	0.75	0.307	1,282.6
TEXT	992	0.0041	0.0005	−8.96	1.45	0.138	79.2
TOB	1,155	0.0220	0.0019	−22.68	3.55	0.187	132.1
TOYS	2,504	0.0216	0.0012	−8.86	2.92	0.141	205.7
TRIP	516	0.0361	0.0043	−16.33	16.32	0.150	45.2

A3.5 The coefficients of the *SIZE* variables are fairly similar, suggesting that household composition is not important for this category of expenditure.

A3.6 The regression results for this specification are summarised in the table below. In the case of *SHEL*, the regression indicates that the *SIZE* effect is attributable to *SIZEAM*. To investigate this further, the regression was repeated: (1) restricting the sample to households with at least one adult male, and (2) restricting the sample to households with either no adult male or just 1 adult male. The first regression produces a negative effect for *SIZEAM*, but it is smaller than with the whole sample and not significant. In the second regression the coefficient of *SIZEAM* jumps dramatically, from −\$424 to −\$793, suggesting very strong economies of scale for this particular comparison.

As might be expected, the *SIZE* composition variables on the whole do not appear to have significant effects if the *SIZE* variable does not in Exercise A3.4. The

© Christopher Dougherty, 2016. All rights reserved.
Published on the Online Resource Centre to accompany Dougherty: Introduction to Econometrics, 5th edition, by Oxford University Press.

A study guide produced by Christopher Dougherty to accompany the module "EC2020 Elements of Econometrics" offered as part of the University of London International Programmes in Economics, Management, Finance, and the Social Sciences.

results for *TOB* are puzzling, in that the apparent economies of scale do not appear to be related to household composition.

Category	*ADM*	*CLOT*	*DOM*	*EDUC*	*ELEC*	*FDAW*	*FDHO*	*FOOT*
EXP	0.0245	0.0309	0.0422	0.1191	0.0120	0.0531	0.0561	0.0056
	(0.0008)	(0.0011)	(0.0026)	(0.0092)	(0.0004)	(0.0010)	(0.0011)	(0.0005)
SIZEAM	−37.17	12.84	−141.47	120.11	23.40	29.36	129.69	2.65
	(9.22)	(10.33)	(32.71)	(107.51)	(3.44)	(9.88)	(9.64)	(4.71)
SIZEAF	−40.47	12.26	−67.26	−58.21	35.73	−45.07	105.17	9.40
	(9.52)	(10.95)	(34.79)	(107.96)	(3.60)	(10.17)	(9.96)	(5.25)
SIZEJM	1.33	17.11	114.68	−413.28	12.53	−24.45	126.94	1.23
	(9.86)	(11.41)	(31.91)	(107.79)	(4.06)	(11.53)	(11.35)	(4.99)
SIZEJF	48.55	29.98	93.82	−287.35	8.93	−26.03	105.01	6.32
	(10.54)	(12.15)	(33.66)	(103.15)	(4.31)	(12.05)	(12.07)	(5.01)
SIZEIN	−34.51	−2.08	441.46	−123.20	−4.05	−61.38	95.90	−16.33
	(22.79)	(22.20)	(59.10)	(289.63)	(8.36)	(23.77)	(23.34)	(11.07)
R^2	0.243	0.179	0.184	0.278	0.204	0.361	0.417	0.086
F	150.1	163.0	62.1	35.6	249.2	480.2	753.6	28.5
n	2,815	4,500	1,661	561	5,828	5,102	6,334	1,827

Category	*FURN*	*GASO*	*HEAL*	*HOUS*	*LIFE*	*LOCT*	*MAPP*	*PERS*
EXP	0.0547	0.0341	0.0579	0.2022	0.0195	0.0061	0.0321	0.0071
	(0.0072)	(0.0008)	(0.0019)	(0.0027)	(0.0017)	(0.0011)	(0.0051)	(0.0002)
SIZEAM	−119.30	90.70	3.01	−175.23	10.54	12.02	2.41	−13.99
	(81.65)	(7.47)	(18.25)	(25.24)	(19.50)	(9.90)	(54.58)	(2.23)
SIZEAF	−55.42	52.23	89.64	−111.39	25.43	19.16	0.75	12.33
	(93.37)	(7.79)	(19.10)	(26.12)	(20.83)	(10.61)	(63.11)	(2.34)
SIZEJM	−27.44	30.83	−62.83	52.32	−23.28	−6.41	131.15	−3.33
	(87.24)	(8.72)	(22.56)	(29.65)	(21.17)	(12.81)	(61.75)	(2.59)
SIZEJF	−15.06	46.24	−57.94	34.65	−15.65	32.97	24.87	−2.10
	(89.23)	(9.27)	(23.96)	(31.58)	(22.98)	(15.85)	(64.61)	(2.71)
SIZEIN	−146.90	−8.90	−109.08	119.91	−116.37	33.48	26.25	−11.30
	(160.29)	(18.02)	(46.46)	(61.40)	(46.00)	(25.82)	(139.98)	(5.32)
R^2	0.110	0.310	0.181	0.475	0.109	0.077	0.116	0.228
F	9.9	427.6	177.0	937.6	25.3	9.6	8.6	187.4
n	487	5,710	4,802	6,223	1,253	692	399	3,817

Category	*READ*	*SAPP*	*TELE*	*TEXT*	*TOB*	*TOYS*	*TRIP*
EXP	0.0049	0.0046	0.0148	0.0040	0.0151	0.0148	0.0448
	(0.0003)	(0.0008)	(0.0004)	(0.0006)	(0.0016)	(0.0010)	(0.0045)
SIZEAM	−6.37	−1.64	29.33	7.42	30.92	−39.66	64.35
	(3.46)	(8.26)	(3.25)	(5.98)	(13.49)	(11.19)	(59.55)
SIZEAF	1.69	8.95	35.59	2.58	22.09	1.30	4.87
	(3.80)	(9.65)	(3.38)	(6.77)	(13.68)	(12.49)	(71.23)
SIZEJM	0.63	−13.21	6.38	−15.90	17.42	42.46	81.61
	(3.93)	(9.73)	(3.78)	(7.51)	(16.52)	(11.30)	(79.96)
SIZEJF	4.73	1.17	12.74	−4.92	−45.12	19.34	102.45
	(4.26)	(10.88)	(4.06)	(7.50)	(16.82)	(11.71)	(91.86)
SIZEIN	−18.98	−19.58	−26.42	19.17	2.92	50.91	−294.14
	(8.56)	(18.58)	(7.82)	(14.13)	(32.83)	(22.49)	(157.82)
R^2	0.108	0.038	0.296	0.059	0.100	0.090	0.197
F	45.8	6.7	404.9	10.4	21.2	41.2	20.8
n	2,287	1,037	5,788	992	368	2,504	516

© Christopher Dougherty, 2016. All rights reserved.
Published on the Online Resource Centre to accompany Dougherty: Introduction to Econometrics, 5th edition, by Oxford University Press.

A study guide produced by Christopher Dougherty to accompany the module "EC2020 Elements of Econometrics" offered as part of the University of London International Programmes in Economics, Management, Finance, and the Social Sciences.

A3.7 The standard error is given by:

$$\text{s.e.}(\widehat{\beta}_2) = \widehat{\sigma}_u \times \frac{1}{\sqrt{n}} \times \frac{1}{\sqrt{\text{MSD}(K)}} \times \frac{1}{\sqrt{1 - r_{K,S}^2}}.$$

	Data		Factors	
	manufacturing sample	services sample	manufacturing sample	services sample
Number of enterprises	25	100	0.20	0.10
Estimate of variance of u	0.16	0.64	0.40	0.80
Mean square deviation of K	4	16	0.50	0.25
Correlation between K and S	0.6	0.6	1.25	1.25
Standard errors			0.050	0.025

The table shows the four factors for the two sectors. Other things being equal, the larger number of enterprises and the greater MSD of K would separately cause the standard error of $\widehat{\beta}_2$ for the services sample to be half that in the manufacturing sample. However, the larger estimate of the variance of u would, taken in isolation, causes it to be double. The net effect, therefore, is that it is half.

A3.8 Exact multicollinearity. An extra year of schooling implies one fewer year of potential work experience. Thus the coefficient of schooling estimates the proportional increase in earnings associated with an additional year of schooling, taking account of the loss of a year of potential work experience.

© Christopher Dougherty, 2016. All rights reserved.
Published on the Online Resource Centre to accompany Dougherty: Introduction to Econometrics, 5th edition, by Oxford University Press.