BS2280 – Econometrics I
Homework 11: Multi-collinearity - Solution

1

What is multi-collinearity and how does its presence affect your regression results?

Multi-collinearity arises when a correlation exists between X variables that move together, but that correlation is not perfect. There are two main problems:

1. Ceteris paribus will not hold anymore, making our marginal effect interpretations misleading.
2. The estimators have a larger variance, making precise estimation difficult. A larger variance means a low t statistic and therefore statistically insignificant coefficients.

2

Work experience is generally found to be an important determinant of earnings. If a direct measure is lacking in a data set, it is standard practice to use potential work experience, PWE, defined by

$$PWE = AGE–S–5$$

as a proxy. This is the maximum number of years since the completion of full-time education, assuming that an individual enters first grade at the age of six. We first regress $EARNINGS$ on S and PWE, and then fit the regression a second time adding AGE as well. Comment on the regression results.

```
lm(formula = EARNINGS ~ S + PWE, data = EAWE22)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -12.2978     8.8682  -1.387    0.166
S             1.8510     0.3935   4.704  3.3e-06 ***
PWE           0.4227     0.3616   1.169    0.243
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.37 on 497 degrees of freedom
Multiple R-squared:  0.1126,    Adjusted R-squared:  0.109
F-statistic: 31.52 on 2 and 497 DF,  p-value: 1.292e-13
```

```
lm(formula = EARNINGS ~ S + PWE + AGE, data = EAWE22)

Coefficients: (1 not defined because of singularities)
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -12.2978     8.8682  -1.387    0.166
S             1.8510     0.3935   4.704  3.3e-06 ***
PWE           0.4227     0.3616   1.169    0.243
AGE               NA         NA      NA       NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.37 on 497 degrees of freedom
Multiple R-squared:  0.1126,    Adjusted R-squared:  0.109
F-statistic: 31.52 on 2 and 497 DF,  p-value: 1.292e-13
```

When AGE is added to the specification, one has an exact linear relationship among the variables (PWE and S) and the regression cannot be performed.
In this situation, R drops one of the variables. Exact multicollinearity has occurred despite the fact that the variables are not perfectly correlated.

3

We regress S on SM, SF, $ASVABAR$ (arithmetic reasoning), $ASVABWK$ (word knowledge), and $ASVABPC$ (paragraph comprehension) , the three components of the $ASVABC$ composite score. Compare the coefficients and their standard errors with those of $ASVABC$ in a regression of S on SM, SF, and $ASVABC$. Making also reference to the correlation table, is multi-collinearity present?

```
lm(formula = S ~ SM + SF + ASVABAR + ASVABWK + ASVABPC, data = EAWE22)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.579182   0.605989  17.458  < 2e-16 ***
SM           0.179259   0.048293   3.712 0.000229 ***
SF           0.100219   0.041628   2.408 0.016428 *
ASVABAR      0.509701   0.171431   2.973 0.003091 **
ASVABWK      0.009398   0.173879   0.054 0.956920
ASVABPC      0.843375   0.184772   4.564 6.33e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.253 on 494 degrees of freedom
Multiple R-squared:  0.3439,    Adjusted R-squared:  0.3373
F-statistic:  51.8 on 5 and 494 DF,  p-value: < 2.2e-16
```

```
lm(formula = S ~ SM + SF + ASVABC, data = EAWE22)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.64842    0.60201  17.688  < 2e-16 ***
SM           0.18212    0.04834   3.768 0.000185 ***
SF           0.09049    0.04164   2.173 0.030254 *
ASVABC       1.26116    0.11458  11.006  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.267 on 496 degrees of freedom
Multiple R-squared:  0.3335,    Adjusted R-squared:  0.3295
F-statistic: 82.74 on 3 and 496 DF,  p-value: < 2.2e-16
```

Correlation table

	EAWE22.ASVABAR	EAWE22.ASVABWK	EAWE22.ASVABPC
EAWE22.ASVABAR	1.0000000	0.7004902	0.7634174
EAWE22.ASVABWK	0.7004902	1.0000000	0.7652013
EAWE22.ASVABPC	0.7634174	0.7652013	1.0000000

ASVABAR and *ASVABPC* have highly significant effects, despite the high correlations (see correlation table) between them and *ASVABWK*, but *ASVABWK* (word knowledge) does not. As expected, the standard errors are greater than that for ASVABC in previous regressions, indicating that multi-collinearity is present.