

## Multicollinearity Problem

### The Model

True relationship  $Y_t = 1.3 + 0.7X_{1t} + 1.3X_{2t} + 1.1X_{3t} + 0.6X_{4t} + u_t$

where  $u_t$  is random normally distributed error term and  $u_t = 0.9u_{t-1} + \varepsilon_t$

```
*Generate Data
set obs 200
g t=_n
tsset t
g u=0 in 1
replace u=0.9*u+rnormal(0,25) if t>1
g x1=rnormal(10,29)
mat C=(1, 0.8153\ 0.8153,1)
corr2data x2 x3, mean(0 0) sds(5 9) corr(C)
corr x2 x3
g x4=rnormal(-5,18)
g y=1.3+0.7*x1+1.3*x2+1.1*x3+0.6*x4+u
g x5=rnormal(7,21)
```

In this case,  $x_2$  and  $x_3$  are highly correlated while  $x_5$  has no relationship with  $y$ .

### Incorrect Specified Model

If we estimate incorrect specified model by including irrelevant variable,  $x_5$ , into the model, the estimated result will involve specification error problem.

The model  $Y_t = \beta_0 + \beta_1X_{1t} + \beta_2X_{2t} + \beta_3X_{3t} + \beta_4X_{4t} + \beta_5X_{5t} + u_t$

```
. reg y x1 x2 x3 x4 x5
```

Source	SS	df	MS	Number of obs	=	200
Model	138260.997	5	27652.1994	F(5, 194)	=	17.34
Residual	309316.866	194	1594.41684	Prob > F	=	0.0000
Total	447577.863	199	2249.13499	R-squared	=	0.3089
				Adj R-squared	=	0.2911
				Root MSE	=	39.93

  

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x1	.6337296	.0975038	6.50	0.000	.441426 .8260333
x2	1.363036	.9800132	1.39	0.166	-.5698118 3.295884
x3	1.108566	.5443478	2.04	0.043	.0349661 2.182165
x4	.5370413	.1640534	3.27	0.001	.2134841 .8605985
x5	.0352846	.1353227	0.26	0.795	-.231608 .3021772
_cons	-16.05227	3.183279	-5.04	0.000	-22.33055 -9.773993

Signal of the problem – High F-test, but low t-test.

### Detecting Serious Multicollinearity Problem

#### Checking Partial Correlation:

```
. pcorr y x1 x2 x3 x4 x5
(obs=200)
```

Partial and semipartial correlations of y with

Variable	Partial Corr.	Semipartial Corr.	Partial Corr.^2	Semipartial Corr.^2	Significance Value
x1	0.4229	0.3879	0.1788	0.1505	0.0000
x2	0.0994	0.0830	0.0099	0.0069	0.1659
x3	0.1447	0.1215	0.0209	0.0148	0.0431
x4	0.2288	0.1954	0.0523	0.0382	0.0013
x5	0.0187	0.0156	0.0004	0.0002	0.7946

#### Checking Pairwise Correlation:

```
. corr y x1 x2 x3 x4 x5
(obs=200)
```

	y	x1	x2	x3	x4	x5
y	1.0000					
x1	0.3880	1.0000				
x2	<b>0.3361</b>	0.0494	1.0000			
x3	<b>0.3450</b>	0.0235	<b>0.8153</b>	1.0000		
x4	0.1846	-0.0659	0.0096	0.0425	1.0000	
x5	<b>0.0139</b>	<b>-0.0641</b>	<b>-0.0196</b>	-0.0033	<b>0.1347</b>	1.0000

#### Checking Variance Inflation Factor (VIF):

```
. estat vif
```

Variable	VIF	1/VIF
x2	3.00	0.333691
x3	3.00	0.333819
x4	1.03	0.975278
x5	1.02	0.978315
x1	1.01	0.989578
Mean VIF	1.81	

#### Auxiliary Regression:

```
. reg x1 x2 x3 x4 x5
```

Source	SS	df	MS	Number of obs	=	200
Model	1766.29781	4	441.574453	F(4, 195)	=	0.51
Residual	167709.888	195	860.050708	Prob > F	=	0.7259
Total	169476.186	199	851.639125	R-squared	=	0.0104
				Adj R-squared	=	-0.0099
				Root MSE	=	29.327

x1	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]

```

      x2 |   .4854557   .7189287    0.68   0.500   -.9324185    1.90333
      x3 |  -.1364981   .3996757   -0.34   0.733   -.9247402    .6517439
      x4 |  -.0960229   .1202923   -0.80   0.426   -.3332639    .1412181
      x5 |  -.0757044   .0992395   -0.76   0.446   -.271425    .1200162
    _cons |   9.181639   2.243592    4.09   0.000    4.756818   13.60646
-----+-----

```

```

. sca r2_1=e(r2)
. sca toll1=1-r2_1
. sca vif1=1/toll1
. reg x2 x1 x3 x4 x5

```

```

      Source |           SS          df           MS      Number of obs   =       200
-----+-----+-----+-----+-----+-----
      Model |   3314.88542            4   828.721354      F(4, 195)       =       97.34
      Residual |   1660.11457          195   8.51340804      Prob > F         =       0.0000
-----+-----+-----+-----+-----
      Total |   4974.99998          199  24.9999999      R-squared        =       0.6663
                                           Adj R-squared    =       0.6595
                                           Root MSE        =       2.9178

```

```

-----+-----
      x2 |           Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----+-----+-----+-----+-----+-----
      x1 |   .0048054   .0071165     0.68   0.500   -.0092298   .0188405
      x3 |   .4530663   .0230112    19.69   0.000    .4076836   .498449
      x4 |  -.0061777   .0119795    -0.52   0.607   -.0298038   .0174484
      x5 |  -.0028744   .0098862    -0.29   0.772   -.0223719   .0166231
    _cons |  -.0598491   .232569    -0.26   0.797   -.5185226   .3988243
-----+-----

```

```

. sca r2_2=e(r2)
. sca tol2=1-r2_2
. sca vif2=1/tol2
. reg x3 x1 x2 x4 x5

```

```

      Source |           SS          df           MS      Number of obs   =       200
-----+-----+-----+-----+-----+-----
      Model |   10738.1748            4  2684.5437      F(4, 195)       =       97.29
      Residual |   5380.82501          195  27.5939744      Prob > F         =       0.0000
-----+-----+-----+-----+-----
      Total |  16118.9998          199   80.9999999      R-squared        =       0.6662
                                           Adj R-squared    =       0.6593
                                           Root MSE        =       5.253

```

```

-----+-----
      x3 |           Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----+-----+-----+-----+-----+-----
      x1 |  -.0043794   .0128232    -0.34   0.733   -.0296695   .0209106
      x2 |   1.468495   .0745846    19.69   0.000    1.321399   1.615591
      x4 |   .0168718   .0215481     0.78   0.435   -.0256256   .0593691
      x5 |   .0031197   .0178009     0.18   0.861   -.0319873   .0382268
    _cons |   .108027   .4187037     0.26   0.797   -.7177421   .9337961
-----+-----

```

```

. sca r2_3=e(r2)
. sca tol3=1-r2_3
. sca vif3=1/tol3

```

```
. reg x4 x1 x2 x3 x5
```

Source	SS	df	MS	Number of obs	=	200
Model	1501.72141	4	375.430353	F(4, 195)	=	1.24
Residual	59242.237	195	303.806344	Prob > F	=	0.2970
				R-squared	=	0.0247
				Adj R-squared	=	0.0047
Total	60743.9584	199	305.246022	Root MSE	=	17.43

  

x4	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x1	-.0339194	.0424923	-0.80	0.426	-.1177229 .0498842
x2	-.2204566	.4274974	-0.52	0.607	-1.063569 .6226555
x3	.1857561	.2372424	0.78	0.435	-.2821344 .6536466
x5	.1075167	.0585662	1.84	0.068	-.0079879 .2230213
_cons	-5.262855	1.337457	-3.93	0.000	-7.900594 -2.625116

```
. sca r2_4=e(r2)
```

```
. sca tol4=1-r2_4
```

```
. sca vif4=1/tol4
```

```
. reg x5 x1 x2 x3 x4
```

Source	SS	df	MS	Number of obs	=	200
Model	1929.9697	4	482.492425	F(4, 195)	=	1.08
Residual	87068.375	195	446.504487	Prob > F	=	0.3673
				R-squared	=	0.0217
				Adj R-squared	=	0.0016
Total	88998.3447	199	447.227863	Root MSE	=	21.131

  

x5	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x1	-.0393028	.0515213	-0.76	0.446	-.1409132 .0623077
x2	-.1507554	.5185012	-0.29	0.772	-1.173346 .8718347
x3	.0504811	.288041	0.18	0.861	-.5175945 .6185566
x4	.1580174	.0860748	1.84	0.068	-.0117398 .3277746
_cons	6.542265	1.618101	4.04	0.000	3.35104 9.733491

```
. sca r2_5=e(r2)
```

```
. sca tol5=1-r2_5
```

```
. sca vif5=1/tol5
```

```
. sca list vif2 tol2 vif3 tol3 vif4 tol4 vif5 tol5 vif1 tol1
```

```
vif2 = 2.9967811
tol2 = .33369137
vif3 = 2.9956372
tol3 = .33381879
vif4 = 1.0253488
tol4 = .97527785
vif5 = 1.0221661
tol5 = .97831454
vif1 = 1.0105319
tol1 = .9895779
```

. estat vif		
Variable	VIF	1/VIF
x2	3.00	0.333691
x3	3.00	0.333819
x4	1.03	0.975278
x5	1.02	0.978315
x1	1.01	0.989578
Mean VIF	1.81	

### Determining Serious Multicollinearity Problem

```
. reg y x1 x2 x3 x4 x5
```

Source	SS	df	MS	Number of obs	=	200
Model	138260.997	5	27652.1994	F(5, 194)	=	17.34
Residual	309316.866	194	1594.41684	Prob > F	=	0.0000
				R-squared	=	0.3089
				Adj R-squared	=	0.2911
Total	447577.863	199	2249.13499	Root MSE	=	39.93

  

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x1	.6337296	.0975038	6.50	0.000	.441426 .8260333
x2	1.363036	.9800132	1.39	0.166	-.5698118 3.295884
x3	1.108566	.5443478	2.04	0.043	.0349661 2.182165
x4	.5370413	.1640534	3.27	0.001	.2134841 .8605985
x5	.0352846	.1353227	0.26	0.795	-.231608 .3021772
_cons	-16.05227	3.183279	-5.04	0.000	-22.33055 -9.773993

```
. estat vif
```

Variable	VIF	1/VIF
x2	3.00	0.333691
x3	3.00	0.333819
x4	1.03	0.975278
x5	1.02	0.978315
x1	1.01	0.989578
Mean VIF	1.81	

```
. corr y x1 x2 x3 x4 x5  
(obs=200)
```

	y	x1	x2	x3	x4	x5
y	1.0000					
x1	0.3880	1.0000				
x2	<b>0.3361</b>	0.0494	1.0000			
x3	<b>0.3450</b>	0.0235	<b>0.8153</b>	1.0000		
x4	0.1846	-0.0659	0.0096	0.0425	1.0000	
x5	<b>0.0139</b>	<b>-0.0641</b>	<b>-0.0196</b>	-0.0033	<b>0.1347</b>	1.0000

### Drop x3

```
. reg y x1 x2 x4 x5
```

Source	SS	df	MS	Number of obs	=	200
Model	131648.405	4	32912.1012	F(4, 195)	=	20.31
Residual	315929.458	195	1620.15107	Prob > F	=	0.0000
				R-squared	=	0.2941
				Adj R-squared	=	0.2797
Total	447577.863	199	2249.13499	Root MSE	=	40.251

  

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x1	.6288748	.0982581	6.40	0.000	.4350897 .8226599
x2	2.99096	.5715049	5.23	0.000	1.863836 4.118084
x4	.5557448	.1651127	3.37	0.001	.2301089 .8813807
x5	.0387431	.1363997	0.28	0.777	-.2302649 .3077511
_cons	-15.93252	3.208318	-4.97	0.000	-22.25997 -9.605058

## Drop x2

```
. reg y x1 x3 x4 x5
```

Source	SS	df	MS	Number of obs	=	200
Model	135176.723	4	33794.1807	F(4, 195)	=	21.09
Residual	312401.14	195	1602.05713	Prob > F	=	0.0000
				R-squared	=	0.3020
				Adj R-squared	=	0.2877
Total	447577.863	199	2249.13499	Root MSE	=	40.026

  

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x1	.6402796	.0976231	6.56	0.000	.4477469 .8328122
x3	1.726111	.3156645	5.47	0.000	1.103557 2.348666
x4	.5286208	.164334	3.22	0.002	.2045207 .852721
x5	.0313667	.1356172	0.23	0.817	-.236098 .2988314
_cons	-16.13385	3.190355	-5.06	0.000	-22.42588 -9.841816

## Drop x4

```
. reg y x1 x2 x5
```

Source	SS	df	MS	Number of obs	=	200
Model	113293.781	3	37764.5938	F(3, 196)	=	22.14
Residual	334284.082	196	1705.53103	Prob > F	=	0.0000
				R-squared	=	0.2531
				Adj R-squared	=	0.2417
Total	447577.863	199	2249.13499	Root MSE	=	41.298

  

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x1	.6095115	.100641	6.06	0.000	.4110332 .8079898
x2	3.020131	.5863029	5.15	0.000	1.863859 4.176403
x5	.0990058	.1387367	0.71	0.476	-.1746025 .3726142
_cons	-18.85533	3.168902	-5.95	0.000	-25.10485 -12.60581

In this case, the test result indicates that  $x_5$  should be dropped.

```
. reg y x1 x2 x3 x4
```

Source	SS	df	MS	Number of obs	=	200
Model	138152.596	4	34538.149	F(4, 195)	=	21.77
Residual	309425.267	195	1586.79624	Prob > F	=	0.0000
				R-squared	=	0.3087
				Adj R-squared	=	0.2945
Total	447577.863	199	2249.13499	Root MSE	=	39.835

  

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x1	.6323429	.0971257	6.51	0.000	.4407912 .8238946
x2	1.357717	.9774565	1.39	0.166	-.5700267 3.285461
x3	1.110347	.5430026	2.04	0.042	.0394349 2.181259
x4	.5426169	.1622647	3.34	0.001	.2225979 .8626359
_cons	-15.82143	3.050376	-5.19	0.000	-21.83739 -9.805466

**Check by using Information Criteria:****Model with  $x_1 x_2 x_3 x_4 x_5$ :**

```
. estat ic
```

```
Akaike's information criterion and Bayesian information criterion
```

```
-----  
Model |      Obs  ll(null)  ll(model)    df      AIC      BIC  
-----+-----  
. |      200 -1055.117  -1018.168     6  2048.336  2068.126  
-----
```

```
Note: N=Obs used in calculating BIC; see [R] BIC note.
```

**Model with  $x_1 x_2 x_3 x_4$ :**

```
. estat ic
```

```
Akaike's information criterion and Bayesian information criterion
```

```
-----  
Model |      Obs  ll(null)  ll(model)    df      AIC      BIC  
-----+-----  
. |      200 -1055.117  -1018.203     5  2046.406  2062.898  
-----
```

```
Note: N=Obs used in calculating BIC; see [R] BIC note.
```