Data files in econometrics\metrics\lectures

Dummy variables.

e.g., in my own research, I study the investment decisions of telephone companies. There are over 1,000 telephone companies, but 4 of them, qwest, sbc, Bellsouth, and verizon, control 90% of the telephones in the united states. I am interested in researching qualitative issues like does Vz's behavior differ from smaller companies? Do companies who qualify for subsidized loans make different investments than firms that don't quality for the subsidy.

We use the dummy variables to classify data into mutually exclusive categories.

Now consider models where the parameters may vary from one observation to another (similar to chow stability test)

Let us consider looking at the relationship between U.S. consumption and income during the period 1929-70.

Yi = real per capita consumption year I
Xi = real per capital disposable income year I

$Yi = B_0 + B_1Xi$ for I = 1929…1970

But during this period there was a war and therefore the relationship between income and consumption was altered during these years (e.g., rationing)

One way to capture this is through dummy variable--usually binary, 0, 1.

D = 1 if characteristic is present
D = 0 if characteristic is not present

Could use other values, e.g., 1 and 2. But it would be more difficult to interpret the predicted value if used 1,2 rather than 0,1 (with 0, the coefficient is multiplied by 0 and falls out when the condition does not apply).

D = 1 1941, 1942…1946
D = 0 otherwise

Write out vector
1929 0
1930 0
…
1941 1
1942 1

…
1946 1
1947 0
…
1970 0

Now we want to decide how the qualitative factor of war affected the relationship between income and consumption. Let's assume through autonomous consumption B1.

$Yi = B_{0I} + B_1Xi$

Where have added subscribt to $B_{0I.}$ Expect B_{0I} to be different in war years than non-war years.
$B_{0I} = B_0 + \delta_0$ δ = delta
Assuming consumption is less during the way years, the parameter δ is the reduciton of autonomous consumption during the war years.

$Yi = B_0 + \delta_0 D_0 + B_1Xi$

$Yi = B_0 + \delta_0 D_{I0} + B_1Xi$ when Di = 1
$Yi = B_0 +$ B2Xi   when Di = 0

Draw income consumption function with 2 lines, same slope, two intercepts, $B_0 + \delta_0 D_{0\ w}$hen Di = 1 and $B_0$when D_0= 0

δ_0 is the coefficient for dummy variable

Now allow for random nature of consumption
$Yi = B_0 + \delta D_I + B2Xi$ **+ Ui**

The parameter δ is treated like the other parameters in the model--we can construct an interval estimate for it or we can test its signficance. If $\delta = 0$, than the war years had no effect on the amount of autonomous consumption.

Using 1958 constant dollars (all output for C=f(income) appears at page five).

^
Ci = 101.36 - 204.95Di + .86 Xi   R^2 = .99
       (3.98)   (-10.91)   (58.73)   t-stats in parenthesis
n=42
at alpha = .05,
Ho: δ_0=0
Ha: δ_0< 0
Why one tailed test?
critical t is approximately -1.68

Since -10.91 < -1.68, we reject Ho at alpha five percent

Had structural shift in economy during the war years. Draw income consumption line again, where now the Y intercept is -103.59 (101.36 - 204.95)

B_0at of 101.36 is the non-war years benchmark.

Instead of assuming that the effect of a war was to change the level of autonomous consumption, it could be t hat MPC changed.
$Yi = B_0+ B_1 Xi$

Where have added subscript I to the coefficient on income, indicating that it is a time varying parameter

$B_{1i}= B1+ \delta_1 Di$ Where δ_1 is the difference in the slope in the war years.
$Yi = B_0+ B_1 Xi + \delta_1 Di\ Xi$

Di = 0 non-war years
Di = 1 war years

$Yi = B_0+ B_1 Xi + \delta_1 Di\ Xi$ when Di = 1
$Yi = B_0\ \ B_1 Xi$ when Di = 0

Draw 2 SRF with same intercept but flatter slope during the war years
Slope for war years is B2+ δ_1

If assumed that the war affected both autonomous consumption and the marginal propensity to consume, both effects may be incorporated in the model.

$Yi = B_0 + \delta_0 D_I + B_1 Xi + \delta_1 Di\ Xi + Ui$

Draw two SRFs
Non-war years has intercept and slope B1, B2
War years have intercept B1 + δ , slope B2 + α

Could do F test for both dummy variables:
$Yi = B1 + \delta D_I + B2\ Xi + \alpha Di\ Xi + Ui$
$Yi = B1 \qquad + B2\ Xi \qquad + Ui$

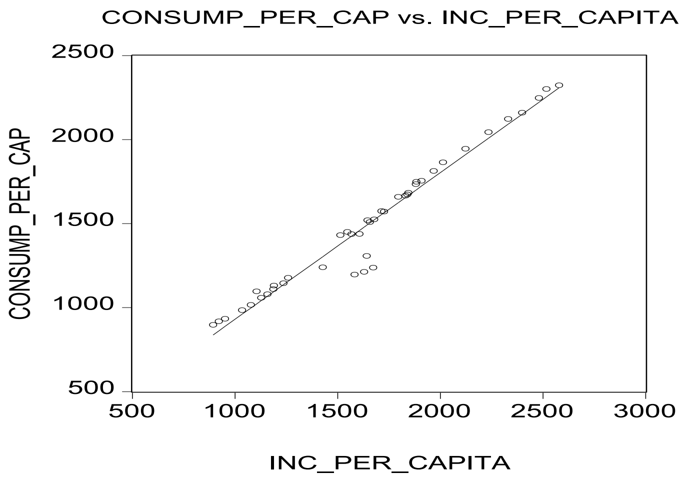
As pointed out by G. at page 309 in his discussion of savings/income relationship, you can get the same results from Chow stability test and dummy variables. The advantage of the dummy variable approach is that you run one regression versus the 3 regresssions doing the Chow stability test ((a) full model; (b) period 1970-81; (c) period 82-90). Furthermore, unlike with the Chow stability test, we can see why the periods are different (slope or intercept, or both, by looking at t statistics of coefficients associated with the dummy variables.).

We can have interactions between dummy variables. Suppose we had observations on the wages earned by individuals.
Yi = Wage = B0 + B2*gender + B3* race + B4*education
Gender 0 F, 1 M
Race 0 white, 1 non-white

B1 = mean wage WF. This is what G refers to at p. 302 as the base or benchmark category.
B0 + B2 = WM
B0+B2 + B3 = Non-white male
B0 + B3 nonwhite female

We can have interaction
Yi = Wage = B0 + B2*gender + B3* race + B4*education + B5gender * education

B5 is picking up earnings for males with a given level of education is different than for females



```
Dependent Variable: CON_PER_CAPITA
Method: Least Squares
Date: 05/03/03   Time: 08:32
Sample: 1929 1970
Included observations: 42

        Variable          Coefficient    Std. Error    t-Statistic      Prob.

           C               57.89570      49.94908       1.159094       0.2533
    INC_PER_CAPITA         0.872335      0.029190       29.88482       0.0000

R-squared                  0.957132   Mean dependent var            1498.786
Adjusted R-squared         0.956061   S.D. dependent var            403.3989
S.E. of regression         84.55942   Akaike info criterion         11.75923
Sum squared resid          286011.8   Schwarz criterion             11.84198
Log likelihood            -244.9439   F-statistic                   893.1023
Durbin-Watson stat         0.337628   Prob(F-statistic)             0.000000
```

```
Dependent Variable: CON_PER_CAPITA
Method: Least Squares
Date: 05/03/03   Time: 08:17
Sample: 1929 1970
Included observations: 42

        Variable          Coefficient    Std. Error    t-Statistic      Prob.

           C               101.3574      25.44646       3.983162       0.0003
    INC_PER_CAPITA         0.863749      0.014708       58.72483       0.0000
           D1             -204.9532      18.78852      -10.90842       0.0000

R-squared                  0.989418   Mean dependent var            1498.786
Adjusted R-squared         0.988876   S.D. dependent var            403.3989
S.E. of regression         42.54731   Akaike info criterion         10.40786
Sum squared resid          70600.67   Schwarz criterion             10.53198
Log likelihood            -215.5650   F-statistic                   1823.304
Durbin-Watson stat         1.634669   Prob(F-statistic)             0.000000
```

```
Dependent Variable: CON_PER_CAPITA
Method: Least Squares
Date: 05/03/03   Time: 08:13
Sample: 1929 1970
Included observations: 42

        Variable          Coefficient    Std. Error    t-Statistic      Prob.

           C               100.9623      24.23278       4.166353       0.0002
    INC_PER_CAPITA         0.864224      0.014012       61.67629       0.0000
   D1*INC_PER_CAPITA      -0.130357      0.011219      -11.61963       0.0000
```

```
R-squared                 0.990393   Mean dependent var         1498.786
Adjusted R-squared        0.989900   S.D. dependent var         403.3989
S.E. of regression        40.54129   Akaike info criterion      10.31127
Sum squared resid         64100.27   Schwarz criterion          10.43539
Log likelihood           -213.5366   F-statistic                2010.182
Durbin-Watson stat        1.606587   Prob(F-statistic)          0.000000
=========================================================================
```

```
Dependent Variable: CON_PER_CAPITA
Method: Least Squares
Date: 05/03/03   Time: 08:18
Sample: 1929 1970
Included observations: 42
=========================================================================
        Variable            Coefficient   Std. Error   t-Statistic   Prob.
=========================================================================
           C                 95.86297      21.85102     4.387116    0.0001
           D1                957.1416      299.3662     3.197226    0.0028
     INC_PER_CAPITA          0.867056      0.012632     68.63733    0.0000
   D1*INC_PER_CAPITA        -0.729283      0.187598    -3.887476    0.0004
=========================================================================
R-squared                 0.992429   Mean dependent var         1498.786
Adjusted R-squared        0.991831   S.D. dependent var         403.3989
S.E. of regression        36.45912   Akaike info criterion      10.12065
Sum squared resid         50512.16   Schwarz criterion          10.28615
Log likelihood           -208.5337   F-statistic                1660.425
Durbin-Watson stat        1.267800   Prob(F-statistic)          0.000000
```

Note that the intercept during the war years is now = 95.86297 + 957.1416 = 1053. The slope during the way years is .867056 -.729283 = .1377. This same result can be obtained by just running a regression for the war years:

```
Dependent Variable: CON_PER_CAPITA
Method: Least Squares
Date: 05/03/03   Time: 12:32
Sample: 1941 1946
Included observations: 6
==================================================================
    Variable       Coefficie  Std. Error  t-Statistic     Prob.
                          nt
==================================================================
        C          1053.005   816.3406    1.289908      0.2666
 INC_PER_CAPIT     0.137773   0.511764    0.269212      0.8011
        A
==================================================================
R-squared          0.017796     Mean dependent           1272.50
                           var                                 0
Adjusted R-               -     S.D. dependent           89.9661
squared            0.227755 var                                0
S.E. of regression 99.68612     Akaike info              12.3031
                           criterion                           3
Sum squared        39749.29     Schwarz criterion        12.2337
resid                                                          2
Log likelihood            -     F-statistic              0.07247
                   34.90939                                    5
Durbin-Watson      0.719312     Prob(F-statistic)        0.80108
stat                                                           3
==================================================================
```

Note the negative R^2. Also, the t-statistics are low illustrating the value of putting all of the data into one SRF.

Now, returning to the unrestricted model:
Ho: B1= B3 = 0

Ha: At least one of the Bs does not equal zero
Alpha = .05
F = ((.992429 - 0.957132) /2) / ((1 - .992429) / (42-4)) = 88.508
F(2,38) at alpha = 5% = approximately 3.24. Since 88.508 > 3.24 we do not accept Ho at alpha

Omitted Variable Bias

Economic theory provides a basis for choosing the variables to include as RHS variables. However, economic theory often provides only a general guideline. In most instances there is uncertainty with regards to the variables that should be included or excluded from a regression. Economic theory may suggest a primary set of variables that should be included, but there may be a secondary, possibly extraneous, set of variables that may or may not have a systematic affect on Y. E.g., suppose you are modeling the demand for a good. Clearly you include the price of the product. But how many competing and complementary products' prices should be included? E.g., modeling the demand for Coke? Do you include the price of milk? Beer? Vodka? Orange Juice?

In demand theory, we have talked about how the demand for tuna is also influenced by the price of other brands and temperature. This suggests the need for more than one right hand side explanatory variable.

When we turn to an economic model with more than one explanatory variable into its corresponding linear statistical model, where the outcome is a function of more than one explanatory variable, it becomes a general linear statistical model or the multiple regression model.

We are now going to learn how to use leaste squares rules to estimate Bs for multiple regression model.

Let’s try example where McDonald’s must decide how much money should be spent advertising their products and what special (lower prices) should be introduced for that week. Of particular interest is how total receipts changes as the level of advertising expenditures changes. Does an increase in advertising expenditures lead to an increase in total receipts? Is this a profitable undertaking? Also, does special increase total sales?

Imagine you are an economic consultant hired by McDonald’s to solve this problem. You may posit a linear model like
TR = B_0 + B_1*P + B_2*A
TR = total receipts (measured in thousands)
P = price
A = advertising expenditures (in thousands)

P is an average price for all products.

Interpreting Parameters
B_1 = the change in TR when p is increased by one unit (one dollar) and advertising expenditures is held constant (partial TR with respect to P).

A partial derivate describes how one variable changes (TR in this case) when another variable changes (P in this case) and all other variables are held constant.

B_1 can be considered a response coefficient.
If an increase in B is positive, the demand for McDonald's hamburgers is price ineleastic. If $B_1 < 0$, demand elastic.

Would we ever observe $B_1 > 0$? No, top half of demand curve is elastic region. P down, TR down, but costs up, hence profits down. Don't expect elasticity to be less than 1 in absolute value terms.

Response parameter B_2.

B_2 = change in TR (in thousands of dollars) when a is increased by one unit (one thousand dollars) and p is held constant. Partial TR w.r.t. A.

Expect $B_2 > 0$. if $B_2 < 1$, imply don't advertise; increase in expenditures is greater than increase in receipts; also have cost of additional sales.

$B_2 > 1$, may advertise.

B_1. Mathematically if P=A=0, value of TR since we will never have data where p and a are zero, it is unrealistic to think our model will be a good approximation to reality in that region, and so the strict mathematical interpretation is not realistic. Thus we are not interested in B1 for its own sake, but it is important ingredient for estimation of the equation as a whole.

What have we done so far—we have turned management's general questions about advertising and pricing policies into questions that require us to use sample data to estimate and draw inferences about the unknown Bs. Now have to convert economic model into statistical model.

Like earlier models, we recognize observations can't fall on straight line, have to allow for random error term. Adding Ui making it a statistical model.

$TR = B_0 + B_1 *P + B_2 *A + Ui$

TR is now a random variable that depends on P and A.

More generally,

Yi = B0 + B1Xi2 + B2Xi3+…BkXik + Ui

Where Xi2, Xi3, … Xik are called regressor variables, or more simply regressors. B1, B2, …Bk are response coefficient. B0 is the intercept parameter.

The true relationship among economic variables is never known with certainty and therefore may accept Ho: Bk=0 when not true.

Suppose true error is
Yi = B0 + B1Xi1 +B2Xi2 + Ui
But we estimate
Yi = B0 + B1Xi1 + Vi

Consequences of omitting X2 can be as following

1. If X2 is correlated X1, B0hat and B1 hat are biased (expectations don't equal population parameters). It can be shown that E(B1hat) = using Wooldridge 3.23, p. 79 and 3.45 p. 90, E(B1~) = B1 + B2$\bar{\delta}_1$ where $\bar{\delta}_1$ = slope from reg X2 (the omitted variable) on X1 (the included variable). Unless b21=0, B1hat is biased. If B2 and b21 are positive B1 hat will have an upward bias. If B2 > 0, and b12 < 0, B1hat < B1.

To illustrate, consider the hamburger example. For this data set, we saw if estimate Y = B0 + B1*price = 122.038 - .8569*price
S_{12}= .13893 = sample cov(price,advertising expenditures).get this in stata correlate price, adv, covariance
S_{11} = sample variance price = .267712^2 = .07167
B2hat = B2 + B3*(S_{23}/S_{22}) = -6.64193 + 2.9843*(0.13893/.07167) = -.8569

```
. reg receipts price

      Source |       SS           df       MS      Number of obs   =        52
-------------+----------------------------------   F(1, 50)        =      0.01
       Model |   2.6838586         1   2.6838586   Prob > F        =    0.9212
    Residual |  13578.6685        50  271.573369   R-squared       =    0.0002
-------------+----------------------------------   Adj R-squared   =   -0.0198
       Total |  13581.3523        51  266.301026   Root MSE        =    16.479

------------------------------------------------------------------------------
    receipts |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       price |  -.8568949   8.619684    -0.10   0.921    -18.17004    16.45625
       _cons |   122.0383   17.40497     7.01   0.000     87.07944    156.9973
------------------------------------------------------------------------------

. reg receipts price advertising

      Source |       SS           df       MS      Number of obs   =        52
-------------+----------------------------------   F(2, 49)        =    159.83
       Model |  11776.1841         2  5888.09204   Prob > F        =    0.0000
    Residual |  1805.16823        49   36.840168   R-squared       =    0.8671
-------------+----------------------------------   Adj R-squared   =    0.8617
       Total |  13581.3523        51  266.301026   Root MSE        =    6.0696
```

```
------------------------------------------------------------------------------
    receipts |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       price |  -6.641929   3.191193    -2.08   0.043    -13.05487   -.2289874
 advertising |   2.984299   .1669361    17.88   0.000     2.648828     3.31977
       _cons |   104.7855   6.482719    16.16   0.000       91.758     117.813
------------------------------------------------------------------------------

. sum

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
    receipts |        52    120.3231    16.31873         89      153.5
   intercept |        52           1           0          1          1
       price |        52    2.001731    .2677116        1.5       2.47
 advertising |        52    9.661538    5.117642        1.8       19.6
    quantity |        52    61.20974    12.07142   44.61538         95
-------------+--------------------------------------------------------
     lnprice |        52       .6851    .1354609   .4054651   .9042181
       lnadv |        52    2.098994    .6319994   .5877867    2.97553
         lnq |        52    4.096047     .191564   3.798079   4.553877

. corr advertising price
(obs=52)

             | advert~g    price
-------------+------------------
 advertising |   1.0000
       price |   0.1014   1.0000
```

The error *u* arises because of factors, or variables, that influence *Y* but are not included in the regression function. There are always omitted variables.

The bias in the OLS estimator that occurs as a result of an omitted factor, or variable, is called ***omitted variable*** bias. For omitted variable bias to occur, the omitted variable "*Z*" must satisfy two conditions:

<u>The two conditions for omitted variable bias</u>

1. *Z* is a determinant of *Y* (i.e. *Z* is part of *u*); **and**
2. *Z* is correlated with the regressor *X* (*i.e.* corr(*Z*,*X*) ≠ 0)

***Both** conditions must hold for the omission of Z to result in omitted variable bias*.

A formula for omitted variable bias: recall the equation,

$$\hat{\beta}_1 - \beta_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})u_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$= \frac{\frac{1}{n}\sum_{i=1}^{n} v_i}{\left(\frac{n-1}{n}\right) s_X^2}$$

where $v_i = (X_i - \bar{X})u_i \approx (X_i - \mu_X)u_i$. Under Least Squares Assumption #1, $E[(X_i - \mu_X)u_i] = \mathrm{cov}(X_i,u_i) = 0$.

But what if $E[(X_i - \mu_X)u_i] = \mathrm{cov}(X_i,u_i) = \sigma_{Xu} \neq 0$?

Multicollearity

Variance of the OLS Estimators

In economics, unlike natural sciences, we generally do not conduct experiments. Other fields are able to hold everything constant, change one variable, and measure impact of change. Don't/Can't do this in economics --have variables moving at the same time.

This cause a problem because in the model:
Y = B0 + B1X1 + B2X2 + Ui
Var (B1 hat) = sigma squared/sum[Xi1-Xbar1)(1-r_{12}^2)

As r_{12} goes to 1, var(B2 hat) goes to infinity.

Or Wooldridge 3.51

σ^2

________'

$SST_j(1-R_j^2)$

Where $SST_j = \sum (X_{ij} - \bar{X}_j)^2$

R_j is the R^2 from regression X_j on all other independent variables.

For example, if want to estimate demand for airline tickets. Delta lowrs price as a promotional fare & everyone matches => in demand equation, high r_{12} .

Macro example--consider an aggregate consumption function in which we express consumption in a given period as a function of income. May also believe that income in past period or periods influence your consumption habits. Reflect notion of business cycle where next quarter affected by this quarter and past quarter. Would expect the R.H.S. variables to be highly correlated.

Or if estimating output fucntion over time as a function of amount of various inputs would expect to see K&L used inr elatively fixed proportions--problem of collinearity.

Klein-Goldberger example estimate consumption as a function of
W - wage income
P - non-wage income
A - farm income

Expect components of income to move together over time due to linkages in the economy

Y= Consumption = $B_0 + B_1w+B_2p + B_3a + Ui$

Run regression using data 1928-1950 with 1942 to 1944 excluded.

	B0	B1	B2	B3
Coefficient	8.133	1.059	.452	.121
Standard error	8.92	0.17	0.66	1.09
t-value	.91	6.1	0.69	0.11
C.I.E. 95%	(-10.78,27.04)	(.69,1.43)	(-.94,1.84)	(-2.18,2.43)

R^2.95; n=20; F107.37; F3,16=5.29 => model doing a good job but doesn't show up in t values. Also note that B2>1

The interrelationships among the different components of income mask separate contributions of each component toward the explanation of spending behavior. Since the variables are highly correlated there is no way of disentangling the separate influences of the variables.

Statistical consequences of collinearity

1. If an exact linear relationship exists between variables then can't estimate OLS--Bhats are not defined.

When the data on RHS exhibit nearly exact linear dependencies, or collinearity

2. Sampling variance large for Bs=> interval C.I.E. wide and the information provided by the sample data about the unknown parameters is relatively imprecise.
3. With high variance of Bs, t tests will likely lead to accept Ho: B=0 despite high R^2 or F. the problem is that collinear variables do not provide enough information to estimate their separate effects even though economic theory, and their total effect, may indicate their importance in the relationship.
4. Parameter estimates may be very sensitive to the selection or deletion of a few observations or the deletion of an apparently insignificant variable.
5. Despite the difficulties in isolating the effect of an individual variable from such a sample, accurate forecasts may still be possible if the nature of the collinear relationships remain the same within the new (future) observations.
6. Parameter estimates are still unbiased.

When are we likely to have a problem?

If correlation coefficient between two explanatory variables is greater than 0.8 or 0.9. In the Klein-Goldber mdoel

	W	P	A
W	1		
P	.718	1	
A	.915	.630	1

Correlation coefficients work when problem is with 2 variables. G. table 10.1 at page 351 nicely illustrates how the correlation between X2 and X3 increases the variance of the coefficients. Also see Wooldridge figure 3.1.

But doesn't work if have linear relationship with 3 or more variables.

For 3 or more variables, could run auxiliary regression (G. p.361).
First, before we consider the 3 variable case, recall that in the two variable model,

Y = B0 + B2X1 + B3X2
In this linear model, the correlation between X2,X3= r. r^2 = R^2 (p.85 G).

It is r^2 that appears in our denominator for the variance of B2hat $r_{23}{}^2$
What if we have more than 2 explanatory variables? G. talks of auxiliary regressions.
X1 = B0 + B2X1 + B3X2 + … BkXk

Can run this regression for each of the Xs. Klien's rule of thumb suggests that multicollinearity may be a troublesome problem if the Rsquared obtained from an auxiliary regression is greater than the overall R squared. P.361 Gujarati.

For the case where we have more than two explanatory variables, consider material at page 101 of Wooldridge: equation 3.51. This is a related way of detecting if we have a problem with multicolinearity.

$$\text{Var(Bk)} = \frac{\text{sigma squared}}{\text{sum(Xk-Xkbar)^2}} * \frac{1}{\text{(1 - Rj^2)}}$$

$$= \frac{\text{sigma squared}}{\text{sum(Xk-Xkbar)^2}} * \text{VIF (variance inflating factor)}$$

where Rj^2 = R^2 in the regression of Xj on the remaining (k-2) regressors

Large Rj^2 => larger var(Bk)

Intuitively, the precision of estimators of a single parameter is the ratio of the uncertainty of the model, sigma squared, to the amount of variability in the kth explanatory variable that is not explained by the other k-1 explanatory variables, (1 - Rj^2). . If VIF high, or the Rsqaured for the auxiliary is high, a collinear relationship exists.

Solutions for collinear Data

1. Data set doesn't have enough information about individual effects of Xs. **Obtain more data**. But this is not always possible—expensive to collect and sometimes especially time series have limited data. If do longer period of time the structure of the economy can change. Also the new data may also be collinear.
2. Introduce non-sample information in the form of restrictions of the parameters. For example, in the consumption function, we might assume the effect of income on consumption, B1, is greater than the effect of non-wage income, B2, and from farm income, B3. Perhaps from earlier research we might impose the restriction that B3 = .625B1.

Y= Consumption = B0 + B1w+B2p + B3a + Ui
= B0 + B1w+B2p + .625B1a + Ui
= B1 + B1(w + .625a) + b2p + Ui
run this regression. Once estimate B1, B3 = .625B1

Known as restricted least squares.

How do we know if the restriction is valid?
Stata: test wage = .625 * nonwageincome

Use F test

$$F = \frac{\text{(RSSr) – RSSur)/r}}{}$$

RSSu/(n-k)

$$= \frac{(R\text{^}2ur – R\text{^}2r)/r}{(1-R\text{^}2ur)/(n-k)}$$

=(.9527-.9521)/1 /(1-.9527)/16 = .21

For data set 1928 – 1950

	B0	B1	B2	
Coefficient restricted least squares	8.486	.989	.491	
s.e.	8.68	.08	.63	
T	.98	12.36	.78	
Old s.e.	8.92	.17	.66	
t-value (old)	.91	6.1	0.69	0.11

Note that B1 is now less than 1. And the t statistics have increased relative to prior to imposing the restriction.

How do we get B3 = B1*.625 = .989 * .625 = .618125

Var(ax) = a^2var(x) page 881 G. and ec249.
Std(ax) = sqrt(var(ax)) = sqrt(a^2var(x)) = astd(x) = .625(.08) = .05

So the t value for B3 =.618/.05 = 12.36, which is also the t value for B1.

For this problem $F^{.01}_{c}(1,16) = 4.49$

F value for restricted model = .21 => don’t reject Ho: B3=.625B1

We have shown that using theory or prior work (non-sample information) in the form of linear constraints on the parameter values reduces estimator-sampling variability. However, the restricted estimator is biased unless the restrictions are true. Thus it is important to use good non-sample information, so that the reduced sampling variability is not brought at a price of large estimator biases.

Other solutions:

1. collect more data
2. dropping a variable (but this raises the likelihood of specification bias—chapter 13).
3. transform the data. E.g. $Y = B_1 + B_2X2 + B_3X3 + Ui$

Divide each side by X3 and get $Y/x3 = B_1 (1/x3) + B_2(x2/x3) + B_3 + u/x3$

or do first differences