The Frisch-Waugh-Lovell Theorem (FWL Theorem)

The FWL Theorem shows how to decompose a regression of **y** on a set of variables **X** into two pieces. If we divide **X** into two sets of variables, (call them **X1** and **X2**) and regress **y** on all of the variables in **X1** and **X2**, you get the same coefficient estimates on **X2** and the same residuals if you regress **y** on all the variables in **X1** and **X2** or if you carry out the following steps:

1) Regress all the variables in **X2** on **X1** and take the residuals
2) Regress **y** on **X1** and take the residuals
3) Regress the residuals from step 2 (**y**) on the residuals from step 1 (**X2**)

As an illustration of this process, I will use the built-in Stata **auto** data set, but in order to make the graphs (shown later) show the process more clearly, I have added 8 mpg to the mileage figures of all the foreign cars. With this change, we get the following result:

```
. sysuse auto
. replace mpg = mpg+8 if foreign
(22 real changes made)

. regress mpg foreign weight

      Source |       SS       df       MS              Number of obs =      74
-------------+------------------------------           F(  2,    71) =  165.06
       Model |  3832.04445     2  1916.02223           Prob > F      =  0.0000
    Residual |  824.171761    71   11.608053           R-squared     =  0.8230
-------------+------------------------------           Adj R-squared =  0.8180
       Total |  4656.21622    73  63.7837838           Root MSE      =  3.4071

------------------------------------------------------------------------------
         mpg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     foreign |   6.349971   1.075994     5.90   0.000       4.2045    8.495442
      weight |  -.0065879   .0006371   -10.34   0.000    -.0078583   -.0053175
       _cons |    41.6797   2.165547    19.25   0.000     37.36172    45.99768
------------------------------------------------------------------------------
```

(In the real **auto** data set, the coefficient on foreign is **-1.650029**.)

I will use **foreign** as my **X1** variable and **weight** as the **X2** variable. (**X1** and **X2** can include multiple variables, but the illustration will be clearer with just one each.)

Step 1) Regress all the variables in **X2** on **X1** and take the residuals

```
. regress weight foreign

      Source |       SS       df       MS              Number of obs =      74
-------------+------------------------------           F(  1,    72) =   39.02
       Model |  15496779.3     1  15496779.3           Prob > F      =  0.0000
    Residual |  28597399.1    72  397186.099           R-squared     =  0.3514
-------------+------------------------------           Adj R-squared =  0.3424
       Total |  44094178.4    73  604029.841           Root MSE      =  630.23

------------------------------------------------------------------------------
      weight |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
```

```
     foreign |  -1001.206   160.2876    -6.25   0.000    -1320.734   -681.6788
       _cons |   3317.115   87.39676    37.95   0.000     3142.893    3491.338
```

```
. predict weightResidF, residual
```

Step 2) Regress **y** on **X1** and take the residuals

```
. reg mpg foreign
```

```
      Source |       SS       df       MS              Number of obs =      74
-------------+------------------------------           F(  1,    72) =   90.32
       Model |  2590.91027     1  2590.91027           Prob > F      =  0.0000
    Residual |  2065.30594    72  28.6848048           R-squared     =  0.5564
-------------+------------------------------           Adj R-squared =  0.5503
       Total |  4656.21622    73  63.7837838           Root MSE      =  5.3558

------------------------------------------------------------------------------
         mpg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     foreign |    12.9458   1.362162     9.50   0.000     10.23038    15.66122
       _cons |   19.82692   .7427186    26.70   0.000     18.34634    21.30751
------------------------------------------------------------------------------
```

```
. predict mpgResidF, residual
```

Step 3) Regress the residuals from **#2** on the residuals from **#1**

```
. reg mpgResidF weightResidF
```

```
      Source |       SS       df       MS              Number of obs =      74
-------------+------------------------------           F(  1,    72) =  108.43
       Model |  1241.13415     1  1241.13415           Prob > F      =  0.0000
    Residual |  824.171747    72  11.4468298           R-squared     =  0.6009
-------------+------------------------------           Adj R-squared =  0.5954
       Total |   2065.3059    73  28.2918616           Root MSE      =  3.3833

------------------------------------------------------------------------------
   mpgResidF |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
weightResidF |  -.0065879   .0006327   -10.41   0.000    -.0078491   -.0053267
       _cons |   4.73e-08   .3933025     0.00   1.000    -.7840342    .7840343
------------------------------------------------------------------------------
```

Notice that in the Step 3 regression, the coefficient on **weightResidF** is the same as the coefficient on **weight** in the original regression, and the SS Residual is the same. However, the MS Residual, standard error, and *t* statistic are all slightly different. The reason is that we have miscounted the df Residual. The original regression had 74 – 3 = 71 degrees of freedom, accounting for the 2 explanatory variables plus the intercept. The new regression has 72 degrees of freedom. But it should really have 71. We have "used up" a degree of freedom when we took residuals in order to get rid of **foreign** from the model. Stata doesn't know we did that, so it can't adjust the degrees of freedom. If we adjust the degrees of freedom, the standard error and *t* statistic will be an exact match.

Why does it work?

I will not give a formal proof of the FWL Theorem. However, the following numeric and graphical argument may be convincing.

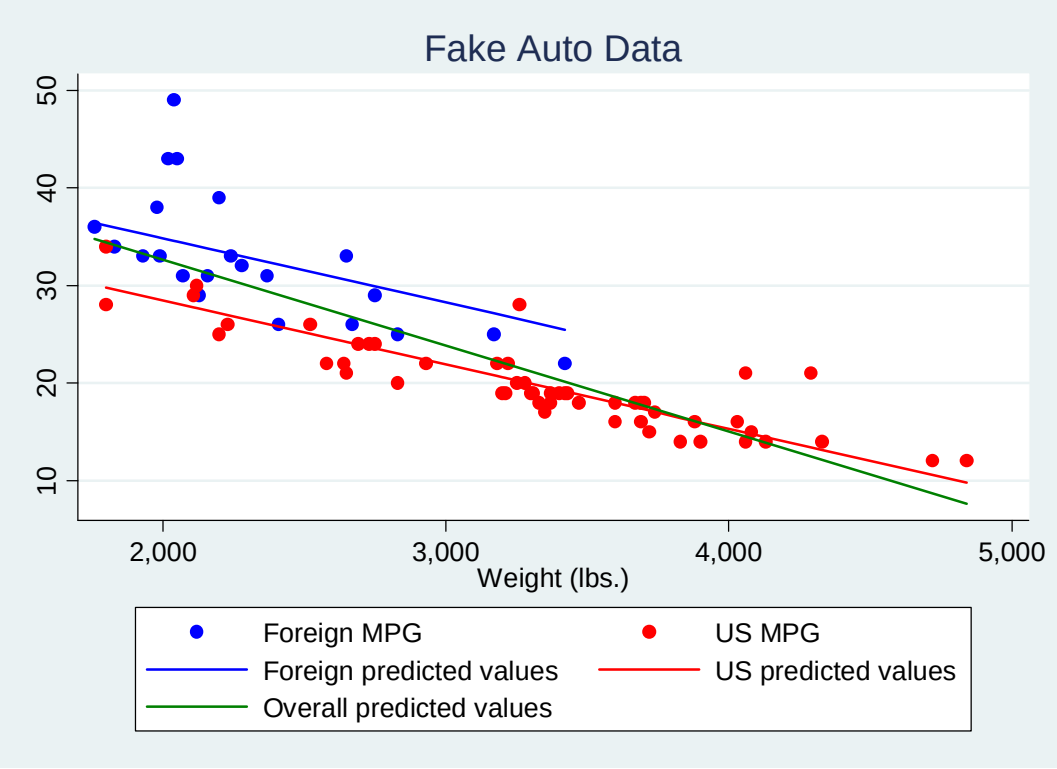
First, notice that in general, omitting a regressor changes the coefficients on other regressors.

In the regression of **mpg** on **foreign** and **weight**, the coefficient on **weight** was about -.0066. If we leave out **foreign** from the model, the coefficient is about -.0088.

```
. reg mpg weight

      Source |       SS       df       MS              Number of obs =      74
-------------+------------------------------           F(  1,    72) =  200.90
       Model |  3427.76371     1  3427.76371            Prob > F      =  0.0000
    Residual |  1228.45251    72  17.0618403            R-squared     =  0.7362
-------------+------------------------------           Adj R-squared =  0.7325
       Total |  4656.21622    73  63.7837838            Root MSE      =  4.1306

------------------------------------------------------------------------------
         mpg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      weight |  -.0088169    .000622   -14.17   0.000    -.0100569   -.0075769
       _cons |   50.29788   1.938648    25.94   0.000     46.43326    54.16251
------------------------------------------------------------------------------
```

The reason is the **foreign** is correlated with **weight**. Foreign cars tend to be lighter than US cars, and in our fake data, foreign cars have better gas mileage, so it makes it look like the impact of **weight** is larger (more negative) if we leave out **foreign**.

Here's a graph of predicted values from the original model and from the model with only weight as the predictor.



The graph shows fitted values for regression of mpg on weight and foreign, overlaid with a fitted line for regression of mpg on weight only. The green line is the regression of mpg on weight only. It is tilted slightly relative to the red and blue lines.

Is there any way that we can eliminate the correlation between **weight** and **foreign** without destroying the essential characteristics of the original regression? Suppose that we arbitrarily shifted the weight of foreign cars by some constant. Algebraically, that means adding *k****foreign** to **weight**, which will affect foreign cars but not US cars.

Original model: $\hat{\mathbf{y}} = \hat{\beta}_0 + \hat{\beta}_1 \mathbf{weight} + \hat{\beta}_2 \mathbf{foreign}$

New model: $\hat{\mathbf{y}} = \hat{\gamma}_0 + \hat{\gamma}_1 (\mathbf{weight} + k * \mathbf{foreign}) + \hat{\gamma}_2 \mathbf{foreign}$

We see that $\hat{\beta}_1 = \hat{\gamma}_1, \quad \hat{\beta}_0 = \hat{\gamma}_0, \qquad \hat{\beta}_2 = \hat{\gamma}_2 + k\hat{\gamma}_1$

So we expect the coefficient on **foreign** to change, but not the other coefficients. This is confirmed below.

```
. gen newweight = weight + 1000*foreign

. reg mpg newweight foreign

      Source |       SS       df       MS              Number of obs =      74
-------------+------------------------------           F(  2,    71) =  165.06
       Model |  3832.04445     2  1916.02223           Prob > F      =  0.0000
    Residual |  824.171761    71   11.608053           R-squared     =  0.8230
-------------+------------------------------           Adj R-squared =  0.8180
       Total |  4656.21622    73  63.7837838           Root MSE      =  3.4071

------------------------------------------------------------------------------
         mpg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   newweight |  -.0065879   .0006371   -10.34   0.000    -.0078583   -.0053175
     foreign |   12.93786   .8665282    14.93   0.000     11.21005    14.66567
       _cons |    41.6797   2.165547    19.25   0.000     37.36172    45.99768
------------------------------------------------------------------------------
```

There is one particular value of *k* that gives a correlation of zero between **foreign** and **weight** + *k****foreign.** If we run a regression of **weight** on **foreign**, we are calculating $\widehat{\mathbf{weight}} = k * \mathbf{foreign} + \hat{\mathbf{u}}$, and by the method of moments, we know that the residuals are uncorrelated with **foreign**. That is the reason for Step 1.

The result of the regression is:

```
. regress mpg foreign weightResidF

      Source |       SS       df       MS              Number of obs =      74
-------------+------------------------------           F(  2,    71) =  165.06
       Model |  3832.04445     2  1916.02222           Prob > F      =  0.0000
    Residual |  824.171767    71  11.6080531           R-squared     =  0.8230
-------------+------------------------------           Adj R-squared =  0.8180
       Total |  4656.21622    73  63.7837838           Root MSE      =  3.4071

------------------------------------------------------------------------------
         mpg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     foreign |    12.9458   .8665279    14.94   0.000       11.218    14.67361
weightResidF |  -.0065879   .0006371   -10.34   0.000    -.0078583   -.0053175
       _cons |   19.82692   .4724741    41.96   0.000     18.88484    20.76901
------------------------------------------------------------------------------
```

The weights have been shifted in each group, but within groups, the relationship between **mpg** and **weight** is unchanged, so the coefficient on **weightResidF** is the same as the former coefficient on **weight**.

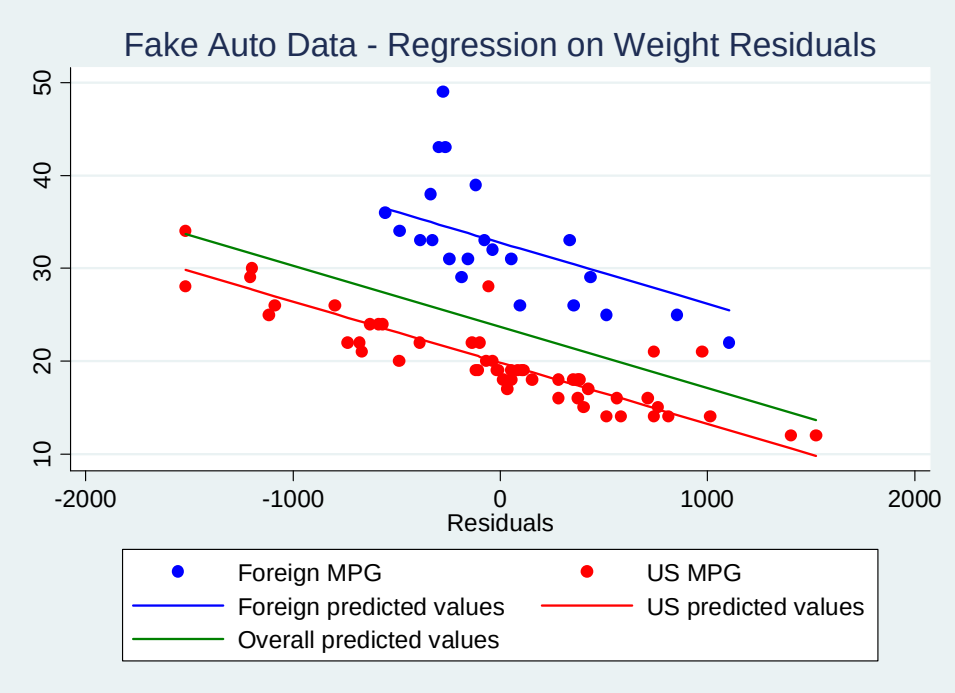
We can confirm directly that the residuals are uncorrelated with **foreign**.

```
. corr foreign weightResidF
(obs=74)

             |  foreign weight~F
-------------+------------------
     foreign |   1.0000
weightResidF |  -0.0000   1.0000
```

Because now there is no correlation between **weightResidF** and **foreign**, including or leaving out **foreign** from the model cannot change the coefficient on **weightResidF**.

```
. reg mpg weightresidf

      Source |       SS       df       MS              Number of obs =      74
-------------+------------------------------           F(  1,    72) =   26.17
       Model |  1241.13421     1  1241.13421           Prob > F      =  0.0000
    Residual |    3415.082    72  47.4316945           R-squared     =  0.2666
-------------+------------------------------           Adj R-squared =  0.2564
       Total |  4656.21622    73  63.7837838           Root MSE      =  6.8871

------------------------------------------------------------------------------
         mpg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
weightResidF |  -.0065879   .0012879    -5.12   0.000    -.0091552   -.0040206
       _cons |   23.67568   .8006053    29.57   0.000      22.0797    25.27165
------------------------------------------------------------------------------
```

The graph shows these two regressions.



The green line is the predicted values from a regression of **mpg** on **weightResidF** without **foreign**. The red and blue points and lines are from the model including **foreign**. You can see that the green line is parallel to the red and blue lines, corresponding to the unchanged coefficient on **weightResidF.**

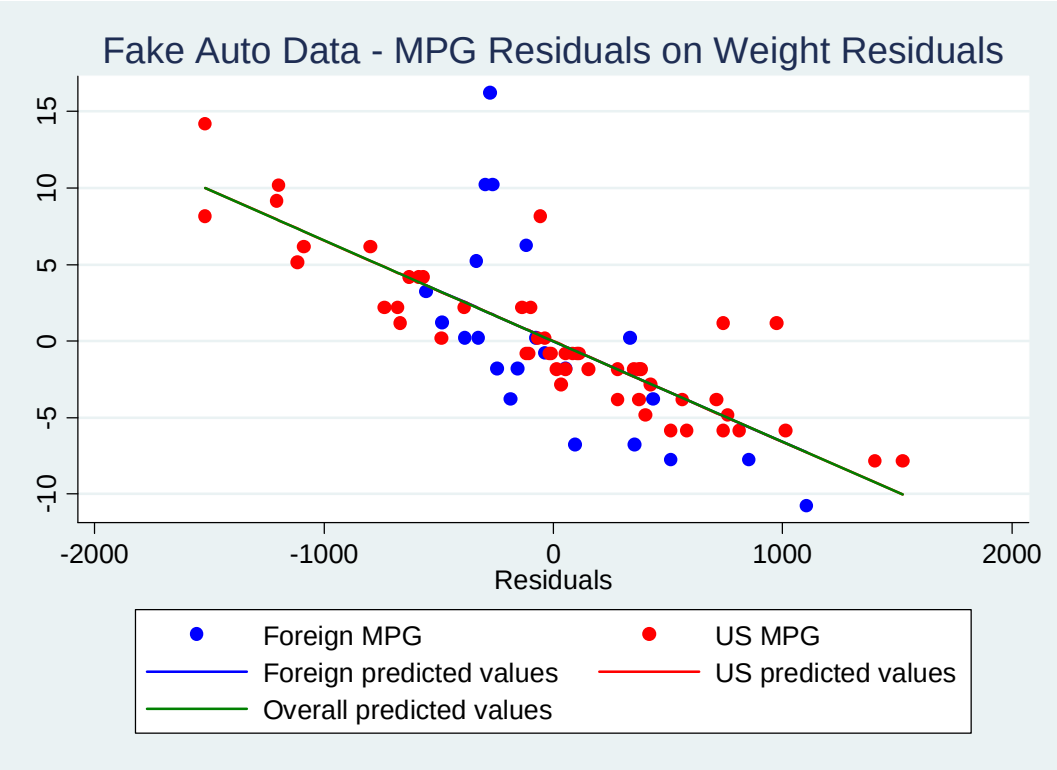
Notice that the relationships between the red and blue points and their corresponding lines have not changed, even though the blue points and line have been moved relatively to the right. This means that in the regression of **mpg** on **weightResidF** and **foreign**, the residuals are the same as in the regression of **mpg** on **weight** and **foreign**. However, the distance from the green line to the actual data points is not the same, so we don't get the same residuals in the regression of **mpg** on **weightResidF** without **foreign**.

What would happen if we now subtracted the groups means (foreign and US) from every MPG? We would shift the red and blue points and lines down, but we wouldn't change the relationship between **mpg** and **weightResidF**, because that is a slope that doesn't depend on the overall mean of the data, and it also doesn't depend on the individual group means. The group means can be subtracted by regressing on **foreign** and taking the residuals.

```
. reg mpg foreign
… Output omitted
. predict mpgResidF, residual

. reg mpgResidF weightResidF foreign

      Source |       SS       df       MS              Number of obs =      74
-------------+------------------------------           F(  2,    71) =   53.46
       Model |  1241.13415     2  620.567076            Prob > F      =  0.0000
    Residual |  824.171747    71  11.6080528            R-squared     =  0.6009
-------------+------------------------------           Adj R-squared =  0.5897
       Total |   2065.3059    73  28.2918616            Root MSE      =  3.4071

------------------------------------------------------------------------------
   mpgResidF |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
weightResidF |  -.0065879   .0006371   -10.34   0.000    -.0078583   -.0053175
     foreign |  -1.36e-07   .8665279    -0.00   1.000    -1.727808    1.727807
       _cons |   8.78e-08   .4724741     0.00   1.000    -.9420865    .9420867
------------------------------------------------------------------------------
```

Because **mpgResidF** is uncorrelated with **foreign**, the coefficient on **foreign** is zero (to within numerical precision) in the regression of **mpgResidF** on **weightResidF** and **foreign**. So we can drop **foreign** from the model without changing anything.

```
. reg mpgResidF weightResidF

      Source |       SS       df       MS              Number of obs =      74
-------------+------------------------------           F(  1,    72) =  108.43
       Model |  1241.13415     1  1241.13415            Prob > F      =  0.0000
    Residual |  824.171747    72  11.4468298            R-squared     =  0.6009
-------------+------------------------------           Adj R-squared =  0.5954
       Total |   2065.3059    73  28.2918616            Root MSE      =  3.3833

------------------------------------------------------------------------------
   mpgResidF |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
weightResidF |  -.0065879   .0006327   -10.41   0.000    -.0078491   -.0053267
       _cons |   4.73e-08   .3933025     0.00   1.000    -.7840342    .7840343
```



The green line now perfectly matches the original red and blue lines.

The coefficient on **weightResidF** is the same as the original coefficient on **weight**, and the residuals are the same.

As noted above, we need to account for the fact that we dropped one predictor out of the model. It's coefficient was essentially 0, so the coefficients and sums of squares are unchanged, but the degrees of freedom no longer accounts for that predictor. The predictor is still effectively in the model – we used it and took residuals, but Stata doesn't know that, so we have fooled Stata into using the wrong degrees of freedom, not accounting for the predictor **foreign** that we have "residualized" out of the regression.

If we want to get the right standard error, we need to adjust the residual degrees of freedom down by 1. In general, if we residualize *k* predictors, we need to adjust the residual degrees of freedom down by *k*.

Reference

A proof of the FWL theorem and additional insight using regression geometry can be found in Section 2.4 of Russell Davidson and James MacKinnon (2004), *Econometric Theory and Methods*, Oxford University Press.