Introduction to Econometrics II: FWL

Seung-hun Lee*

January 24th, 2022

1 Frisch-Waugh-Lovell Theorem

Consider a multivariate regression, where the model is

$$y = X_1\beta_1 + X_2\beta_2 + e \tag{Long}$$

where X_1 and X_2 each contains k_1 and k_2 regressors. We are interested in estimating β_2, or the effect of X_2 while holding X_1 fixed. One simple way of doing this is to control for X_1, obtain the β estimates for both X_1 and X_2 (denoted as $\hat{\beta}$) and get $\hat{\beta}_2$ components.

Another way of doing this is to partial out X_1 from y and X_2 by projecting both onto the space of X_1 and using the 'residualized' (with respect to X_1) version of the two variables. Since this nets out the influence of X_1 from both variables, we are also holding X_1 fixed by doing this. To achieve this, I define a projection P_1 and a residual maker matrix M_1.

$$P_1 = X_1(X_1'X_1)^{-1}X_1', \;\; M_1 = I - X_1(X_1'X_1)^{-1}X_1' = I - P_1$$

Then, multiply the long data generating process by premultiplying M_1

$$\begin{aligned} M_1 y &= M_1X_1\beta_1 + M_1X_2\beta_2 + M_1 e \\ &= M_1X_2\beta_2 + M_1 e \\ \rightarrow \widetilde{y} &= \widetilde{X_2}\beta_2 + \widetilde{e} \end{aligned} \tag{Short}$$

*Contact me at sl4436@columbia.edu if you spot any errors or have suggestions on improving this note.

Where $M_1X_1 = X_1 - X_1(X_1'X_1)^{-1}X_1' = 0$.

From this setup, we can prove two features of the Frisch-Wald-Lovell theorem.

Theorem 1.1 (Frisch-Waugh-Lovell theorem)**.** *From the long and short form regressions above,*

- *The OLS estimates of β_2 from the two regressions are identical*
- *The residuals from the OLS estimate of the two regressions are identical*

To proceed, it should be made clear that both projection and the residual maker matrix are idempotent and symmetric

Property 1.1 (Properties of Projection Matrix $P_X = X(X'X)^{-1}X'$)**.** *Note that*

- *Symmetric:* $P_X' = (X(X'X)^{-1}X')' = X(X'X)^{-1}X' = P_X$
- *Idempotent:* $P_X^2 = X(X'X)^{-1}X'X(X'X)^{-1}X' = X(X'X)^{-1}X' = P_X$

Property 1.2 (Properties of Residual Matrix $M_X = I - X(X'X)^{-1}X'$)**.** *Note that*

- *Symmetric:* $M_X' = I' - (X(X'X)^{-1}X')' = I - X(X'X)^{-1}X' = I - P_X = M_X$
- *Idempotent:* $M_X^2 = (I - P_X)(I - P_X) = I - P_X - P_X + P_X'P_X = I - P_X = M_X$

Furthermore, we must show the following

Property 1.3 (Projection and residual matrices of the subspace of X)**.** *Let $X = [X_1\ X_2]$ and P, M be a projection and residual maker matrix for X. Then we can show*

- $P_1P = PP_1 = P_1$
- $M_1M = MM_1 = M$

Proof. Rewrite $PP_1 = X(X'X)^{-1}X'X_1(X_1'X_1)^{-1}X_1'$. The part on $X(X'X)^{-1}X'X_1$ is a projection of X_1 onto space of $[X_1\ X_2]$, a wider space in which X_1 is included. So we have $X(X'X)^{-1}X'X_1 = X_1$. Thus, $PP_1 = P_1$.

As for P_1P, Note that $(PP_1)' = P_1'P' = P_1P$ by symmetry of projection matrix. So $P_1P = P_1' = P_1$.

Once this is shown, we can easily work with $M_1 M$ and MM_1 using the above facts and show that

$$\begin{aligned} M_1 M &= (I - P_1)(I - P) \\ &= I - P - P_1 + P_1 P = I - P - P_1 + P_1 = I - P = M \\ MM_1 &= (I - P)(I - P_1) \\ &= I - P_1 - P + PP_1 = I - P_1 - P + P_1 = I - P = M \end{aligned}$$

□

Another useful property is to show that the OLS residual $\hat{e}$ can be written using M

Property 1.4 (Residual of the OLS)**.** *Let $\hat{e}$ be the OLS residual from the regression of y onto X, whose original GDP is $y = X\beta + e$. Then we can write $\hat{e} = My = Me$*

Proof. Rewrite $\hat{e}$ into the following and use the fact that $\hat{\beta} = (X'X)^{-1}X'y$ to get

$$\begin{aligned} \hat{e} &= y - X\hat{\beta} \\ &= [I - X(X'X)^{-1}X']y = My \\ &= M(X\beta + e) \\ &= X\beta - X(X'X)^{-1}X'X\beta + Me = Me \end{aligned}$$

Thus $\hat{e} = My = Me$ □

Now we have what it takes to prove the theorem. Keep in mind that any vector y can be expressed as the sum of its projection onto some space X and the residual from that projection. So we have

$$y = Py + My$$

Note that $Py = X(X'X)^{-1}X'y = X\hat{\beta}$. We can split this into $X_1\hat{\beta}_1 + X_2\hat{\beta}_2$. Then we can write

$$y = X_1\hat{\beta}_1 + X_2\hat{\beta}_2 + My$$

Then, we premultiply $X_2'M_1$ on both sides to get

$$\begin{aligned} X_2'M_1 y &= X_2'M_1X_1\hat{\beta}_1 + X_2'M_1X_2\hat{\beta}_2 + X_2'M_1My \\ &= X_2'M_1X_2\hat{\beta}_2 + X_2'My \end{aligned}$$

We can further show that $X_2' M y = 0$, since

$$X_2' M y = X_2' y - X_2' P y = 0$$

where $X_2' P = X_2'$ since this is a transpose of a projection of X_2 onto a wider space that includes itself. Thus $P X_2 = X_2$ and take transpose to both sides.

Take all the conditions and we can derive that

$$X_2' M_1 y = X_2' M_1 X_2 \hat{\beta}_2 \iff \hat{\beta}_2 = (X_2' M_1 X_2)^{-1} (X_2' M_1 y)$$

Now we need to verify that $(X_2' M_1 X_2)^{-1} (X_2' M_1 y)$ is the expression for the OLS estimation on the short version of the data generating process. Apply OLS on $\widetilde{y} = \widetilde{X_2} \beta_2 + \widetilde{e}$ to get

$$\begin{aligned}
\widehat{\widetilde{\beta}}_2 &= (\widetilde{X_2}' \widetilde{X_2})^{-1} (\widetilde{X_2}' \widetilde{y}) \\
&= (X_2' M_1' M_1 X_2)^{-1} (X_2' M_1' M_1 y) \\
&= (X_2' M_1 X_2)^{-1} (X_2' M_1 y)
\end{aligned}$$

by symmetry and idempotency of M_1. Thus, the first part of the Frisch-Waugh-Lovell theorem is verified.

Now we need to show the equivalence of the residuals. Start with $y = X_1 \hat{\beta}_1 + X_2 \hat{\beta}_2 + \hat{e}$ and premultiply M_1 to get

$$M_1 y = M_1 X_2 \hat{\beta}_2 + M_1 \hat{e} \rightarrow \widetilde{y} = \widetilde{X_2} \hat{\beta}_2 + M_1 \hat{e}$$

By the equivalence of the OLS estimates on β_2, $M_1 \hat{e}$ is the residual of regressing $\widetilde{y}$ onto $\widetilde{X_2}$ using OLS, $\widehat{\widetilde{e}}$. We can further write

$$\widehat{\widetilde{e}} = M_1 \hat{e} = M_1 M e = M e = \hat{e}$$

using the fact that $M_1 M = M$ we have shown above. Thus, the residuals are also equal.

Key takeaway is that the long and short DGP derives $\hat{\beta}_2$ by fixing X_1. Long DGP does this by including X_1 as a control variable and deriving $\hat{\beta}_2$ using partial derivatives. The short DGP is doing the same thing, but by netting out the influence of X_1 from y and X_2 by residualizing both variables after projecting them onto the space of X_1. If you think of it this way, it is natural that the OLS estimators and the residuals are identical.