Frisch-Waugh-Lovell Theorem Derivation

Adapted from Greene, 2008, *Econometric Analysis*, page 27

The normal equations in matrix form are $\mathbf{X'Xb = X'y}$. If $\mathbf{X}$ is partitioned into two segments, the partitioned version is

$$\begin{bmatrix} \mathbf{X_1'X_1} & \mathbf{X_1'X_2} \\ \mathbf{X_2'X_1} & \mathbf{X_2'X_2} \end{bmatrix} \begin{bmatrix} \mathbf{b_1} \\ \mathbf{b_2} \end{bmatrix} = \begin{bmatrix} \mathbf{X_1'y} \\ \mathbf{X_2'y} \end{bmatrix}$$

Multiplication of partitioned matrices is done using the same basic rule as matrix multiplication, so this can be divided into two separate pieces corresponding to the two rows:

$$\begin{bmatrix} \mathbf{X_1'X_1} & \mathbf{X_1'X_2} \end{bmatrix} \begin{bmatrix} \mathbf{b_1} \\ \mathbf{b_2} \end{bmatrix} = \begin{bmatrix} \mathbf{X_1'y} \end{bmatrix} \qquad \text{(Equation 1)}$$

$$\begin{bmatrix} \mathbf{X_2'X_1} & \mathbf{X_2'X_2} \end{bmatrix} \begin{bmatrix} \mathbf{b_1} \\ \mathbf{b_2} \end{bmatrix} = \begin{bmatrix} \mathbf{X_2'y} \end{bmatrix} \qquad \text{(Equation 2)}$$

Our goal is to find an expression for $\mathbf{b_2}$ that does not involve $\mathbf{b_1}$. Although these two equations both involve matrices, we can solve them in a way similar to ordinary simultaneous equations in algebra, by solving one equation for $\mathbf{b_1}$ in terms of $\mathbf{b_2}$, then substituting that solution into the second equation.

The first equation can be multiplied out:

$$\mathbf{X_1'X_1b_1} + \mathbf{X_1'X_2b_2} = \mathbf{X_1'y}$$
$$\mathbf{X_1'X_1b_1} = \mathbf{X_1'y} - \mathbf{X_1'X_2b_2}$$
$$\mathbf{b_1} = \left(\mathbf{X_1'X_1}\right)^{-1}\mathbf{X_1'y} - \left(\mathbf{X_1'X_1}\right)^{-1}\mathbf{X_1'X_2b_2} = \left(\mathbf{X_1'X_1}\right)^{-1}\mathbf{X_1'}\left(\mathbf{y} - \mathbf{X_2b_2}\right)$$

Similarly, equation 2 can be multiplied out:

$$\mathbf{X_2'X_1b_1} + \mathbf{X_2'X_2b_2} = \mathbf{X_2'y}$$

Plugging in the solution for $\mathbf{b_1}$ gives

$$\mathbf{X_2'X_1}\left(\left(\mathbf{X_1'X_1}\right)^{-1}\mathbf{X_1'}\left(\mathbf{y} - \mathbf{X_2b_2}\right)\right) + \mathbf{X_2'X_2b_2} = \mathbf{X_2'y} \qquad \text{(Equation 3)}.$$

The middle part of the first term is $\mathbf{X_1}\left(\mathbf{X_1'X_1}\right)^{-1}\mathbf{X_1'}$, which is the projection matrix from a regression of $\mathbf{y}$ on $\mathbf{X_1}$, $\mathbf{P_{X_1}}$. Rewriting using this projection matrix and splitting up the terms involving $\mathbf{y}$ and $\mathbf{b_2}$ gives

$$\mathbf{X_2'P_{X_1}y} - \mathbf{X_2'P_{X_1}X_2b_2} + \mathbf{X_2'X_2b_2} = \mathbf{X_2'y} \qquad \text{(Equation 3)}$$

Recall that we can multiply by an identity matrix **I** without changing anything, so this can be rewritten as:

$$\mathbf{X_2'IX_2b_2} - \mathbf{X_2'P_{X_1}X_2b_2} + = \mathbf{X_2'Iy} - \mathbf{X_2'P_{X_1}y} \qquad \text{(Equation 3a)}$$

Combining terms gives:

$$\mathbf{X_2'}\left(\mathbf{I - P_{X_1}}\right)\mathbf{X_2b_2} = \mathbf{X_2'}\left(\mathbf{I - P_{X_1}}\right)\mathbf{y} \qquad \text{(Equation 4)}$$

Now $\left(\mathbf{I - P_{X_1}}\right)$ is the residualizer matrix $\mathbf{M_{X_1}}$, so equation 4 can be rewritten as

$$\mathbf{X_2'M_{X_1}X_2b_2} = \mathbf{X_2'M_{X_1}y} \qquad \text{(Equation 4a)}$$

which can be solved to give

$$\mathbf{b_2} = \left(\mathbf{X_2'M_{X_1}X_2}\right)^{-1}\mathbf{X_2'M_{X_1}y}$$

Recall that the residualizer matrix is symmetric and idempotent, so

$$\mathbf{M_{X_1}} = \mathbf{M_{X_1}M_{X_1}} = \mathbf{M_{X_1}^{\mathit{T}}M_{X_1}}$$

So the expression for $\mathbf{b_2}$ can be rewritten as

$$\mathbf{b_2} = \left(\mathbf{X_2'M_{X_1}^{\mathit{T}}M_{X_1}X_2}\right)^{-1}\mathbf{X_2'M_{X_1}^{\mathit{T}}M_{X_1}y} = \left(\left(\mathbf{M_{X_1}X_2}\right)^{T}\left(\mathbf{M_{X_1}X_2}\right)\right)^{-1}\left(\mathbf{M_{X_1}X_2}\right)^{T}\mathbf{M_{X_1}y}.$$

Looking closely at this formula, you should recognize the OLS regression solution, with $\mathbf{M_{X_1}X_2}$ in place of **X** and $\mathbf{M_{X_1}y}$ in place of **y**. But $\mathbf{M_{X_1}X_2}$ represents the residuals from a regression of $\mathbf{X_2}$ on $\mathbf{X_1}$, and similarly, $\mathbf{M_{X_1}y}$ represents the residuals from a regression of $\mathbf{X_2}$ on **y**. So the solution of the regression coefficients $\mathbf{b_2}$ in a regression that includes other regressors $\mathbf{X_1}$ is the same as first regressing all of $\mathbf{X_2}$ and **y** on $\mathbf{X_1}$, then regressing the residuals from the **y** regression on the residuals from the $\mathbf{X_2}$ regression. This is the FWL Theorem.