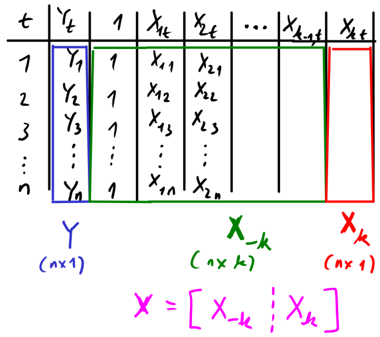
Proof of Frisch-Waugh-Lovell (FWL) Theorem (Based on Freedman (2005)) Assume that $\mathbf{X}$ has rank $k+1 \le n$ (rules out perfect multicollinearity). Let $\widehat{\beta}$ be a minimizer of the least squares objective function. Recall our setup:



In the proof, we will be using a couple of facts of the least squares fit:

Orthogonality $\mathbf{X}'\mathbf{e} = \mathbf{0}$, i.e. the least squares residual vector is orthogonal to the columns of $\mathbf{X}$.

Uniqueness If $\widetilde{\beta}$ is another $(k+1) \times 1$ vector with a residual vector $\mathbf{Y} - \mathbf{X}\widetilde{\beta}$ orthogonal to the columns of $\mathbf{X}$, then $\widehat{\beta} = \widetilde{\beta}$.

Partition $\mathbf{X}$ as $\mathbf{X} = \begin{pmatrix} \mathbf{X}_{-k} & \vdots & \mathbf{X}_k \end{pmatrix}$. Similarly, partition $\widehat{\beta} = \begin{pmatrix} \widehat{\beta}_{-k} \\ \widehat{\beta}_k \end{pmatrix}$, where $\widehat{\beta}_{-k}$ is a $k \times 1$ vector while $\widehat{\beta}_k$ is a scalar.

1. Apply least squares and obtain the linear regression of (a) $\mathbf{Y}$ on $\mathbf{X}$, (b) $\mathbf{Y}$ on $\mathbf{X}_{-k}$, and (c) $\mathbf{X}_k$ on $\mathbf{X}_{-k}$.

2. As a result, you can split the actual values into fitted values plus residuals.

   (a) Least squares coefficient vector $\widehat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y})$ and $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$

   (b) Least squares coefficient vector $\widehat{\gamma} = \left(\mathbf{X}'_{-k}\mathbf{X}_{-k}\right)^{-1}\left(\mathbf{X}'_{-k}\mathbf{Y}\right)$ and $\mathbf{Y} = \mathbf{X}_{-k}\widehat{\gamma} + \widehat{\boldsymbol{\eta}}$

   (c) Least squares coefficient vector $\widehat{\delta} = \left(\mathbf{X}'_{-k}\mathbf{X}_{-k}\right)^{-1}\left(\mathbf{X}'_{-k}\mathbf{X}_k\right)$ and $\mathbf{X}_k = \mathbf{X}_{-k}\widehat{\delta} + \widehat{\boldsymbol{v}}$

3. Now, apply least squares again to obtain the linear regression of $\widehat{\boldsymbol{\eta}}$ on $\widehat{\boldsymbol{v}}$. You will obtain a least squares coefficient (now, a scalar)

$$\widehat{\theta} = (\widehat{\boldsymbol{v}}'\widehat{\boldsymbol{v}})^{-1}\widehat{\boldsymbol{v}}'\widehat{\boldsymbol{\eta}}.$$

   Our task is to prove that $\widehat{\beta}_k = \widehat{\theta}$.

4. I have to show that $\widehat{\boldsymbol{\eta}} - \widehat{\boldsymbol{v}}\widehat{\theta}$ is orthogonal to the columns of $\mathbf{X}$. To see this, I start with the first k columns of $\mathbf{X}$.

$$\mathbf{X}_{-\boldsymbol{k}}{}'\left(\widehat{\boldsymbol{\eta}} - \widehat{\boldsymbol{v}}\widehat{\theta}\right) = \mathbf{X}_{-\boldsymbol{k}}{}'\widehat{\boldsymbol{\eta}} - \mathbf{X}'_{-k}\widehat{\boldsymbol{v}}\widehat{\theta} = \mathbf{0} - \mathbf{0} \cdot \widehat{\theta} = \mathbf{0}.$$

   The second to the last equality follows from the fact that $\widehat{\boldsymbol{\eta}}$ and $\widehat{\boldsymbol{v}}$ are both orthogonal to the first k columns of $\mathbf{X}_{-k}$.

5. Next, I show that $\widehat{\boldsymbol{\eta}} - \widehat{\boldsymbol{v}}\widehat{\theta}$ is orthogonal to the last or $(k+1)$th column of $\mathbf{X}$. Observe that
$$\mathbf{X}_k'\widehat{\boldsymbol{v}} = \left(\mathbf{X}_{-k}\widehat{\delta} + \widehat{\boldsymbol{v}}\right)'\widehat{\boldsymbol{v}} = \widehat{\delta}'\boldsymbol{X'}_{-k}\widehat{\boldsymbol{v}} + \widehat{\boldsymbol{v}}'\widehat{\boldsymbol{v}} = \widehat{\boldsymbol{v}}'\widehat{\boldsymbol{v}}.$$
The last equality follows from the fact that $\widehat{\boldsymbol{v}}$ is orthogonal to the first $k-1$ columns of $\mathbf{X}_{-k}$. Therefore,
$$\begin{aligned}
\mathbf{X}_k'\left(\widehat{\boldsymbol{\eta}} - \widehat{\boldsymbol{v}}\widehat{\theta}\right) &= \mathbf{X}_k'\widehat{\boldsymbol{\eta}} - \mathbf{X}_k'\widehat{\boldsymbol{v}}\widehat{\theta} \\
&= \mathbf{X}_k'\widehat{\boldsymbol{\eta}} - \widehat{\boldsymbol{v}}'\widehat{\boldsymbol{v}}\widehat{\theta} \\
&= \mathbf{X}_k'\widehat{\boldsymbol{\eta}} - \widehat{\boldsymbol{v}}'\widehat{\boldsymbol{\eta}} \\
&= \left(\mathbf{X}_k' - \widehat{\boldsymbol{v}}'\right)\widehat{\boldsymbol{\eta}} \\
&= \left(\mathbf{X}_{-k}\widehat{\delta}\right)'\widehat{\boldsymbol{\eta}} \\
&= \widehat{\delta}'\mathbf{X}'_{-k}\widehat{\boldsymbol{\eta}} \\
&= 0
\end{aligned}$$
The second equality uses the immediately preceding result. The third equality follows from
$$\widehat{\theta} = \left(\widehat{\boldsymbol{v}}'\widehat{\boldsymbol{v}}\right)^{-1}\widehat{\boldsymbol{v}}'\widehat{\boldsymbol{\eta}} \Rightarrow \widehat{\boldsymbol{v}}'\widehat{\boldsymbol{v}}\widehat{\theta} = \widehat{\boldsymbol{v}}'\widehat{\boldsymbol{\eta}}.$$
The fourth equality follows from factoring out a common term. The fifth equality follows from the least squares fit $\mathbf{X}_k = \mathbf{X}_{-k}\widehat{\delta} + \widehat{\boldsymbol{v}}$. The last equality follows from orthogonality.

6. As a result,
$$\begin{aligned}
\mathbf{Y} &= \mathbf{X}_{-k}\widehat{\gamma} + \widehat{\boldsymbol{\eta}} \\
&= \mathbf{X}_{-k}\widehat{\gamma} + \widehat{\boldsymbol{v}}\widehat{\theta} + \left(\widehat{\boldsymbol{\eta}} - \widehat{\boldsymbol{v}}\widehat{\theta}\right) \\
&= \mathbf{X}_{-k}\widehat{\gamma} + \left(\mathbf{X}_k - \mathbf{X}_{-k}\widehat{\delta}\right)\widehat{\theta} + \left(\widehat{\boldsymbol{\eta}} - \widehat{\boldsymbol{v}}\widehat{\theta}\right) \\
&= \mathbf{X}_{-k}\left(\widehat{\gamma} - \widehat{\delta}\widehat{\theta}\right) + \mathbf{X}_k\widehat{\theta} + \left(\widehat{\boldsymbol{\eta}} - \widehat{\boldsymbol{v}}\widehat{\theta}\right) \\
&= \underbrace{\left(\begin{array}{c:c} \mathbf{X}_{-k} & \mathbf{X}_k \end{array}\right)}_{\mathbf{X}} \underbrace{\left(\begin{array}{c} \widehat{\gamma} - \widehat{\delta}\widehat{\theta} \\ \widehat{\theta} \end{array}\right)}_{\widetilde{\beta}} + \underbrace{\left(\widehat{\boldsymbol{\eta}} - \widehat{\boldsymbol{v}}\widehat{\theta}\right)}_{\text{residual}}.
\end{aligned}$$

7. So, we have found another $\widetilde{\beta}$, which just like $\widehat{\beta}$, has residuals $\mathbf{Y} - \mathbf{X}\widetilde{\beta} = \widehat{\boldsymbol{\eta}} - \widehat{\boldsymbol{v}}\widehat{\theta}$ which are also orthogonal to the columns of $\mathbf{X}$. The uniqueness of the least squares minimizer implies that $\widehat{\beta}$ has to be equal to $\widetilde{\beta}$:
$$\widehat{\beta} = \left(\begin{array}{c} \widehat{\beta}_{-k} \\ \widehat{\beta}_k \end{array}\right) = \left(\begin{array}{c} \widehat{\gamma} - \widehat{\delta}\widehat{\theta} \\ \widehat{\theta} \end{array}\right).$$
Therefore, $\widehat{\beta}_k = \widehat{\theta}$.