1. Prove the Frisch–Waugh–Lovell theorem

Given the model:

$$y = D\beta_1 + W\beta_2 + \mu \tag{1}$$

where y is an $n \times 1$ vector, D is an $n \times k_1$ matrix, β_1 is a $k_1 \times 1$ vector, W is an $n \times k_2$ matrix, β_2 is a $k_2 \times 1$ vector, and μ is an $n \times 1$ vector of error terms.

We can construct the following equation:

$$\epsilon_y = \epsilon_D \phi + \xi \tag{2}$$

where ϵ_y is the vector of residuals from regressing y on W, ϵ_D is the matrix of residuals from regressing each column of D on W, and e is an $n \times 1$ vector of error terms.

Running y on W we get:

$$y = W\hat{\alpha}_1 + \epsilon_y \iff \epsilon_y = y - W\hat{\alpha}_1 \tag{3}$$

Similarly, running D on W gives us:

$$D = W\hat{\alpha}_2 + \epsilon_D \iff \epsilon_D = D - W\hat{\alpha}_2 \tag{4}$$

Running ϵ_y on ϵ_D:

$$\begin{aligned}
y - W\hat{\alpha}_1 &= (D - W\hat{\alpha}_2)\phi + \xi \\
y &= W\hat{\alpha}_1 + (D - W\hat{\alpha}_2)\phi + \xi \\
y &= W\hat{\alpha}_1 + D\phi - W\hat{\alpha}_2\phi + \xi \\
y &= D\phi + W(\hat{\alpha}_1 - \hat{\alpha}_2\phi) + \xi \qquad (5)
\end{aligned}$$

Comparing (1) with (5) we can see that:

$$\beta_1 = \phi \tag{6}$$

$$\beta_2 = \hat{\alpha}_1 - \hat{\alpha}_2\phi \tag{7}$$

$$\mu = \xi \tag{8}$$

2. Show that the Conditional Expectation Function minimizes the expected squared error.

Given the model:

$$Y = m(X) + e \tag{9}$$

Where $m(X)$ represents the conditional expectation of Y on X. Lets define the arbitrary model:

$$Y = g(X) + w \tag{10}$$

Where $g(X)$ represents any function of X.

Working with the expected squared error from model (10):

$$\begin{aligned}
E[(Y - g(X))^2] =& E[(Y - m(X) + m(X) - g(X))^2] \\
=& E[(Y - m(X))^2 + 2(Y - m(X))(m(X) - g(X)) + (m(X) - g(X))^2] \\
=& E[(Y - m(X))^2] + 2E[(Y - m(X))(m(X) - g(X))] + E[(m(X) - g(X))^2] \\
=& E[(e^2)] + 2E[(Y - m(X))(m(X) - g(X))] + E[(m(X) - g(X))^2]
\end{aligned}$$

Using the law of iterated expectations:

$$E[(Y - g(X))^2] = E[(e^2)] + 2E[E[(Y - m(X))(m(X) - g(X))|X]] + E[(m(X) - g(X))^2]$$

Since $m(X)$ and $g(X)$ are functions of X, the term $(m(X) - g(X))$ can be thought as constant when conditioning on X. Thus, we get:

$$E[(Y - g(X))^2] = E[(e^2)] + 2E[E[Y - m(X)|X](m(X) - g(X))] + E[(m(X) - g(X))^2]$$

It is important to note that $E[Y - m(X)|X] = 0$ by definition of $m(X)$, so we get:

$$E[(Y - g(X))^2] = E[(e^2)] + E[(m(X) - g(X))^2]$$

Because the second term in the equation is always non-negative, it is clear that the function is minimized when $g(X)$ equals $m(X)$. In which case:

$$E[(Y - g(X))^2] = E[(e^2)] \tag{11}$$