Geometric Intuition for the Frisch–Waugh–Lovell Theorem

Joe D. Campbell*

June 22, 2026

1 Motivation

The Frisch–Waugh–Lovell (FWL) theorem is a core part of any first-year graduate econometrics sequence. It is key to understanding exactly which variation is used for the identification of parameters in ordinary least squares estimation. However, most textbook proofs of the theorem neglect the underlying geometric intuition, relying instead on terse but opaque matrix algebra. This note is intended to bridge that gap by providing a clear geometric framework for the theorem.

2 Definitions

2.1 Ordinary Least Squares Estimation

Ordinary Least Squares (OLS) is motivated by the following model for unit i's outcome y_i

$$y_i = X_i\beta + u_i = X_{1i}\beta_1 + X_{2i}\beta_2 + \cdots + X_{Ki}\beta_K + u_i.$$

More concretely, in a finite sample of size n, we stack our observations into vectors and matrices. Let Y be an $n \times 1$ vector of outcomes, and let X be an $n \times K$ matrix of covariates. We partition X into our covariate of interest X_1 (an $n \times 1$ vector) and the remaining covariates X_{-1} (an $n \times (K-1)$ matrix).[1]

Then the stacked elements take the form

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} X_1 & X_{-1} \end{pmatrix} = \begin{pmatrix} X_{11} & X_{21} & X_{31} & \dots & X_{K1} \\ X_{12} & X_{22} & X_{32} & \dots & X_{K2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_{1n} & X_{2n} & X_{3n} & \dots & X_{Kn} \end{pmatrix}.$$

The OLS estimator $\hat{\beta}^{\text{OLS}} = (\hat{\beta}_1^{\text{OLS}\prime}, \hat{\beta}_{-1}^{\text{OLS}\prime})'$ solves the sum of squares minimization problem

$$\min_{\beta} \|Y - X\beta\|^2 = \min_{\beta} (Y - X\beta)'(Y - X\beta).$$

Provided that X has full column rank, this minimization problem necessarily has a unique solution because the objective function is strictly convex. We take the derivative of the objective function with

*joe.campbell.pro@gmail.com. Ayman Moazzam provided helpful comments and feedback.

[1] Without loss of generality, this partition applies to any single coefficient by simply shuffling the columns of X.

respect to β to obtain the first order condition

$$\left.\frac{\partial \|Y - X\beta\|^2}{\partial \beta}\right|_{\beta=\hat{\beta}^{\mathrm{OLS}}} = -2X'(Y - X\hat{\beta}^{\mathrm{OLS}}) = 0$$

which we can rearrange to obtain the closed-form solution

$$\hat{\beta}^{\mathrm{OLS}} = (X'X)^{-1}X'Y.$$

Let $\hat{U} = (Y - X\hat{\beta}^{\mathrm{OLS}})$ denote the residual vector. The first order condition guarantees that

$$X'\hat{U} = 0.$$

This means that the residual vector $\hat{U}$ is orthogonal to each X_i.

2.2 The Frisch–Waugh–Lovell Theorem

Let V_1 denote the residuals that remain after regressing X_1 on X_{-1}. The FWL theorem states that the coefficient $\hat{\beta}_1^{\mathrm{OLS}}$ obtained from the multivariate regression of Y on X_1 and X_{-1} is algebraically identical to the estimate $\hat{\beta}_1^{\mathrm{FWL}}$ obtained from regressing Y on V_1. Formally, we define:

$$\hat{\beta}_1^{\mathrm{OLS}} \equiv ((X'X)^{-1}X'Y)_1 \quad \text{and} \quad \hat{\beta}_1^{\mathrm{FWL}} \equiv (V_1'V_1)^{-1}V_1'Y$$

The theorem establishes that $\hat{\beta}_1^{\mathrm{OLS}} = \hat{\beta}_1^{\mathrm{FWL}}$. See figure 1 for a visual demonstration.

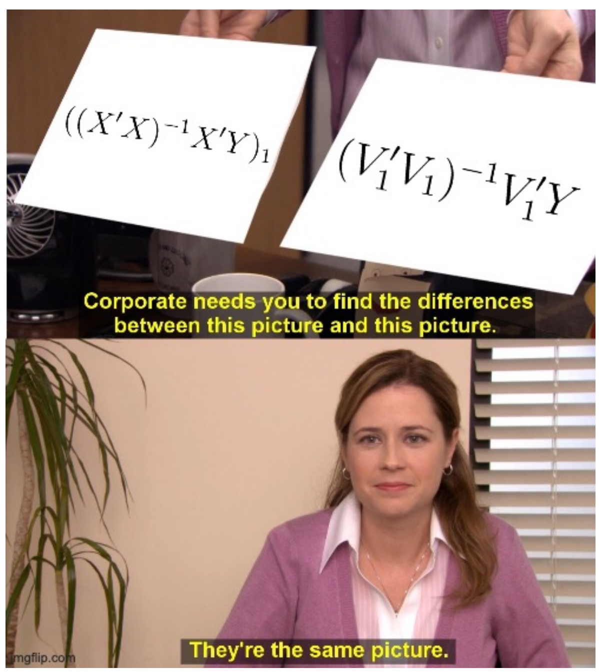

Figure 1: My meme

3 OLS as Projection

The above minimization problem has a geometric interpretation. The vector $\hat{Y} = X\hat{\beta}^{\text{OLS}}$ is the projection of Y onto the span of X, denoted $S(X)$. This is the linear subspace in $\mathbb{R}^n$ spanned by the columns of X. For a visual depiction see figure 2.

$$S(X) = \{X\beta \mid \beta \in \mathbb{R}^K\} = \{\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_K X_K\}.$$

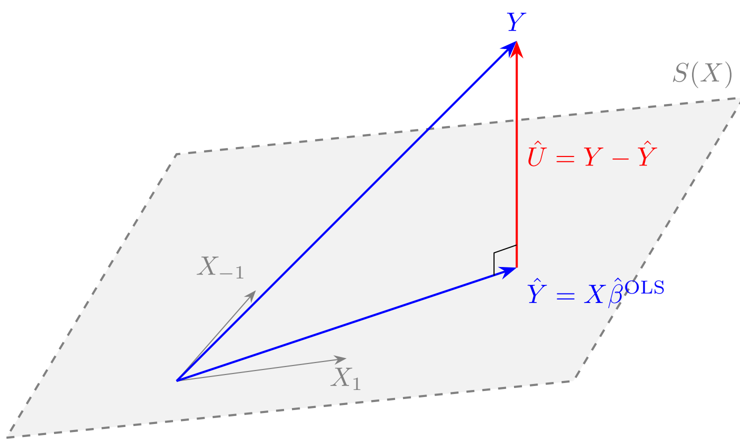

Figure 2: Geometry of projection

Moreover, the original minimization problem is equivalent to finding the weights such that $X\beta = \hat{Y}$. To see this, write

$$Y - X\hat{\beta}^{\text{OLS}} = (Y - \hat{Y}) + (\hat{Y} - X\hat{\beta}^{\text{OLS}}) = \hat{U} + (\hat{Y} - X\hat{\beta}^{\text{OLS}})$$

and plug into the objective function to obtain

$$\|Y - X\beta\|^2 = \left(\hat{U} + (\hat{Y} - X\beta)\right)' \left(\hat{U} + (\hat{Y} - X\beta)\right)$$

then distribute terms to obtain

$$\|Y - X\beta\|^2 = \hat{U}'\hat{U} + 2\hat{U}'(\hat{Y} - X\beta) + (\hat{Y} - X\beta)'(\hat{Y} - X\beta).$$

Rewrite the middle term to see

$$2\hat{U}'(\hat{Y} - X\beta) = 2\hat{U}'(X\hat{\beta}^{\text{OLS}} - X\beta) = 2\underbrace{\hat{U}'X}_{0}(\hat{\beta}^{\text{OLS}} - \beta) = 0$$

where the final equality follows from the (transposed) first order condition. Thus the decomposition simplifies to Pythagorean theorem and we have

$$\|Y - X\beta\|^2 = \|\hat{U}\|^2 + \|\hat{Y} - X\beta\|^2.$$

Because the first term $\|\hat{U}\|^2$ is determined entirely by the OLS projection, it is fixed with respect to β. Minimizing the entire expression is therefore equivalent to minimizing the second term, $\|\hat{Y} - X\beta\|^2$ which achieves its minimum value of zero if and only if $X\beta = \hat{Y}$.

4 Geometric Intuition and Similar Triangles

Now consider a simple 2-dimensional example. Suppose we have

$$Y = \hat{Y} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}, \quad X_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

In this case, the remaining covariates matrix X_{-1} is a vector X_2. The OLS optimization problem is to determine which weights (β_1 and β_2) to put on these vectors to make them sum up exactly to $\hat{Y}$. Y happens to equal $\hat{Y}$ here since $n = K$ but this is irrelevant to the argument. Regardless of n we could restrict our attention to the plane spanned by X_1 and X_2 which by definition contains $\hat{Y}$.

The problem is tricky because the optimal choice of β_1 varies with the choice of β_2 and vice versa. The solution is to view the optimization problem as a two-step problem: first to choose the optimal value of β_1, and second to choose the optimal value of β_2 given the selected value of β_1.[2]

Observe that for any given value of β_1 the optimal value of β_2 is the value such that the horizontal position of $X\beta$ is exactly equal to the horizontal position of $\hat{Y}$. For instance, see figure 3. Even though the choice of $\beta_1 = 4$ is suboptimal, the optimal choice of β_2 given β_1 still results in a βX with the same horizontal coordinate as $\hat{Y}$. Any other choice of β_2 would make βX even farther from $\hat{Y}$.

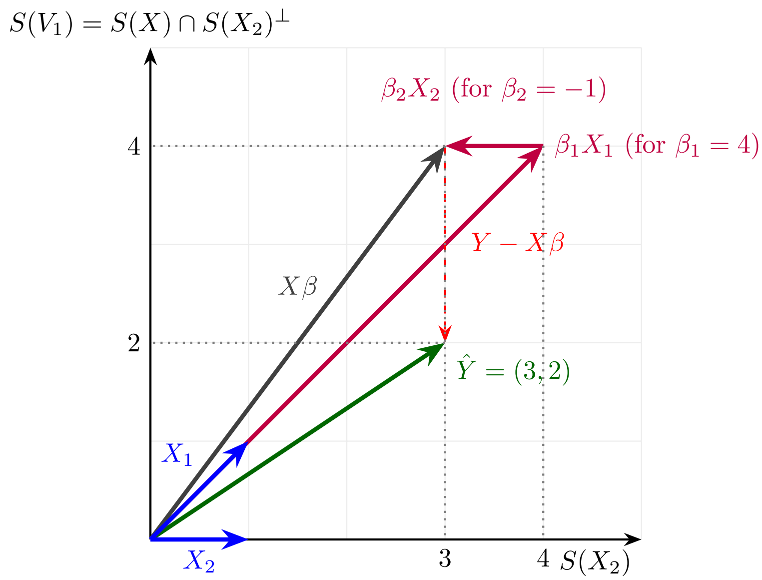

Figure 3: Optimal β_2 for a suboptimal β_1

This means that when we choose the value of β_1 in the first step, we should ignore the horizontal position entirely since we can always move horizontally (and only horizontally) along X_2. This makes the problem trivial. In the first step we choose β_1 to fit the vertical position of $\hat{Y}$, ignoring the horizontal position entirely. Then in the second step we choose β_2 to move the appropriate amount along X_2. This results in the OLS estimates and is depicted in figure 4.

[2] This is justified formally since the joint minimization problem can be nested as:

$$\min_{\beta_1, \beta_2} \|Y - X_1\beta_1 - X_2\beta_2\|^2 = \min_{\beta_1} \left[\min_{\beta_2} \|Y - X_1\beta_1 - X_2\beta_2\|^2 \right]$$

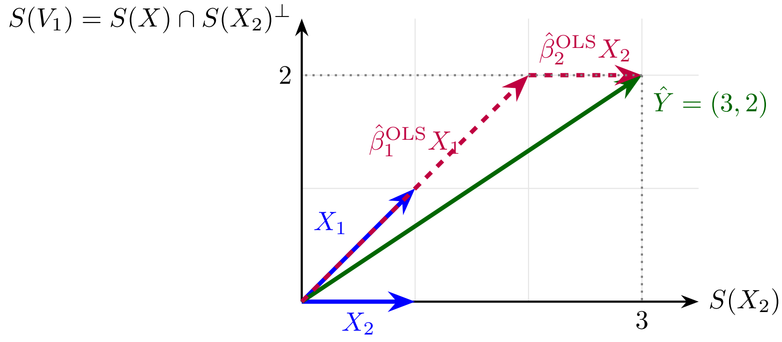

Figure 4: OLS Estimates

The FWL theorem is simply a formalization of this approach using a change of basis. Instead of simply ignoring the horizontal variation in X_1 we eliminate it altogether by regressing X_1 on X_2 in the first step to obtain V_1. We then project $\hat{Y}$ onto V_1 to determine $\hat{\beta}_1^{\text{FWL}}$. See figure 5 for a visual depiction.

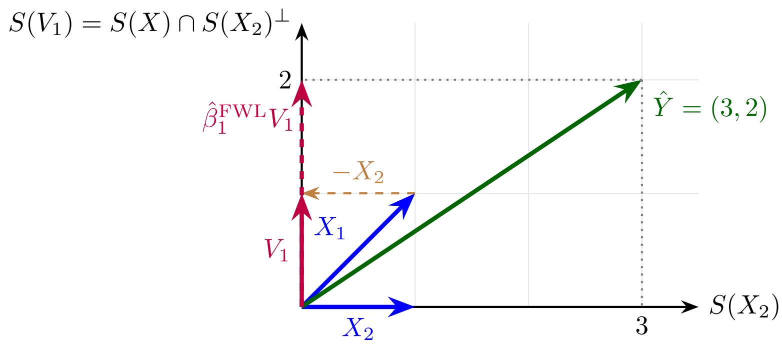

Figure 5: FWL Estimates

The final thing to show is that $\hat{\beta}_1^{\text{OLS}} = \hat{\beta}_1^{\text{FWL}}$. This is simply a similar triangles argument. See figure 6. Both the blue and red shaded triangles share two common angles: the first (θ) at the origin, and the second where they each intersect the y-axis (a right angle by construction, since both horizontal dashed lines denote movements along X_2). This ensures that the triangles are similar and that $\hat{\beta}_1$ is well-defined since $\hat{\beta}_1^{\text{OLS}} = \frac{\|\hat{\beta}_1^{\text{OLS}} X_1\|}{\|X_1\|} = \frac{\|\hat{\beta}_1^{\text{FWL}} V_1\|}{\|V_1\|} = \hat{\beta}_1^{\text{FWL}}$.

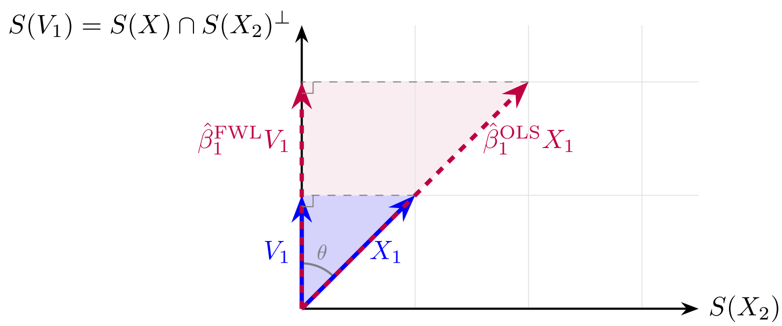

Figure 6: Geometric Proof via Similar Triangles

5 Algebraic Proof

To formalize the geometric intuition, define the residualization matrix

$$M_{-1} = I - X_{-1}(X'_{-1}X_{-1})^{-1}X'_{-1}.$$

Intuitively, this matrix residualizes a vector by projecting out any variation contained in $S(X_{-1})$.

We first premultiply X_1 by M_{-1} to obtain variation $V_1 = M_{-1}X_1$. This is exactly equivalent to projecting X_1 onto $S(V_1) = S(X) \cap S(X_2)^{\perp}$ in the earlier section.

Next we restrict to only the variation in Y that is contained in $S(V_1)$. To do so we premultiply by V_1' and obtain

$$V_1'Y = V_1'(X_1\hat{\beta}_1^{\text{OLS}} + X_{-1}\hat{\beta}_{-1}^{\text{OLS}} + \hat{U}).$$

By construction, V_1 is orthogonal to $S(X_{-1})$ implying that $V_1'X_{-1} = 0$. Moreover because the final OLS residual vector $\hat{U}$ is orthogonal to $S(X)$ we also have $V_1'\hat{U} = 0$. Premultiplying by V_1' thus acts as a directional filter that ignores movements in $S(V_1)^{\perp}$. The expression thus simplifies to

$$V_1'Y = V_1'X_1\hat{\beta}_1^{\text{OLS}}.$$

Moreover, because M_{-1} is symmetric ($M'_{-1} = M_{-1}$) and idempotent ($M_{-1}M_{-1} = M_{-1}$), it satisfies $M'_{-1}M_{-1} = M_{-1}$. Thus, we have $V_1'X_1 = (M_{-1}X_1)'X_1 = X_1'M'_{-1}M_{-1}X_1 = V_1'V_1$ and can write

$$V_1'Y = (V_1'V_1)\hat{\beta}_1^{\text{OLS}}.$$

Intuitively, we can replace X_1 with V_1 since the variation in X_1 not contained in V_1 will be projected out regardless. Last, we can rearrange this equation to isolate the OLS parameter estimate

$$\hat{\beta}_1^{\text{OLS}} = (V_1'V_1)^{-1}V_1'Y.$$

Since the right-hand side is exactly the definition of our FWL estimator, this algebraically verifies that $\hat{\beta}_1^{\text{OLS}} = \hat{\beta}_1^{\text{FWL}}$. This is the algebraic equivalent of our similar triangles argument.

6 Implications for Identification

The key lesson of the FWL theorem is that the variation used to estimate β_1 excludes all the variation linearly explained by X_{-1}. For example, when X_{-1} denotes a set of industry-specific dummies, we use only variation within an industry to estimate the effect of β_1. If no such variation exists, β_1 isn't identified. Economists often discard a lot of variation to restrict to the variation they want! See, for instance, John Cochrane's Causation Does not Imply Variation.

References

[DM04] Russell Davidson and James G. MacKinnon. *Econometric Theory and Methods.* Oxford University Press, 2004.

[Han22] Bruce E. Hansen. *Econometrics.* Princeton University Press, 2022.