Interpreting Coefficients in Nested Models as Directions in the Full Model

Ella Foster-Molina

Contents

1 Introduction

A visual description and example showing the numeric equivalence of the coefficients in a nested model with the slope of a direction in the full model can be found at https://ellafostermolina.github.io/interpret-nested-models/. This document lays out the proofs that show that these equivalencies hold for all cases, not just the example in that document.

2 Defining the Models

For parsimony, assume each variable in each model is centered. This ensures that the constants are all zero and can be dropped from the models.

Let the full model be defined as

$$\begin{aligned} y &= X\beta' \\ &= x_1\beta'_1 + x_2\beta'_2 + \epsilon' \end{aligned} \tag{1}$$

and

$$\hat{\beta}' = (X^T X)^{-1} X^T y. \tag{2}$$

The predicted values of Y from the full model are

$$\hat{y}' = x_1\hat{\beta}'_1 + x_2\hat{\beta}'_2 \tag{3}$$

$$= y - y'_{resid} \tag{4}$$

where Y'_{resid} are the residuals from the full model. Let $\hat{\sigma}'_{full}$ be the estimated variance of the full model, and σ'_{full} be the true variance.

The nested model is

$$y = x_1\hat{\beta}_1 + \epsilon \tag{5}$$

where

$$\hat{\beta}_1 = (x_1^T x_1)^{-1} x_1^T y \tag{6}$$

and

$$E(\hat{\beta}_1) = \hat{\beta}_1. \tag{7}$$

Let the relationship between x_2 and x_1 be defined by

$$x_2 = x_1\gamma + \epsilon_\gamma. \tag{8}$$

where

$$\hat{\gamma} = (x_1^T x_1)^{-1} x_1^T x_2 \tag{9}$$

and

$$\hat{x}_2 = x_1\hat{\gamma}. \tag{10}$$

The residuals of the model estimating x_2 are

$$\begin{aligned}\widehat{x_{resid}} &= x_2 - \widehat{x_2} \\ &= x_2 - x_1\hat{\gamma}.\end{aligned} \tag{11}$$

3 Relationships between coefficients in the models

This section recaps the relationship between β_1 in the nested model and the β' in the full model. One relationship between β_1, β', and γ can be found by substituting Equation 5 into Equation 1.

$$\begin{aligned}y &= x_1\beta_1' + x_2\beta_2' + \epsilon' \\ &= x_1\beta_1' + (x_1\gamma + \epsilon_\gamma)\beta_2' + \epsilon' \\ &= x_1(\beta_1' + \gamma\beta_2') + (\epsilon' + \beta_2'\epsilon_\gamma) \\ &= x_1\beta_1 + \epsilon.\end{aligned}$$

Therefore,

$$\beta_1 = \beta_1' + \gamma\beta_2'. \tag{12}$$

More rigorously, following Greene (2003):

$$
\begin{aligned}
y &= x_1\beta_1' + x_2\beta_2' + \epsilon' \\
(x_1^T x_1)^{-1} x_1^T y &= (x_1^T x_1)^{-1} x_1^T x_1 \beta_1' + (x_1^T x_1)^{-1} x_1^T x_2 \beta_2' + (x_1^T x_1)^{-1} x_1^T \epsilon' \\
\hat{\beta}_1 &= \beta_1' + \hat{\gamma}\beta_2' + (x_1^T x_1)^{-1} x_1^T \epsilon' \\
E(\hat{\beta}_1) &= \beta_1' + \gamma\beta_2' \\
\beta_1 &= \beta_1' + \gamma\beta_2'. \qquad (13)
\end{aligned}
$$

This is a well established result, and has traditionally been used to quantify the 'bias' that exists in models with omitted variables. In the remainder of this paper I show that this can meaningfully interpreted as a change in a particular direction of the regression plane defined by the full model. Whether this represents bias or not is an open question.

4 Relationship between the estimated coefficients in the models

Knowing the relationship between the 'true' coefficients is only part of the battle. This section demonstrates that relationship for the 'true' coefficients shown in Equation 13 has a parallel for the estimated coefficients. The derivation in this section will be used to interpret the estimated coefficient $\hat{\beta}_1$ as a directional derivative in the full model.

Note that

$$x_1^T y_{resid} = 0 \qquad (14)$$

because the residuals of y are required to be orthogonal to x_1 in a least squares regression solution.

We can use the equation for the estimation of $\hat{y}$ to demonstrate the following relationship between the estimated coefficients:

$$
\begin{aligned}
&\hat{y} = x_1\hat{\beta}_1^{'} + x_2\hat{\beta}_2^{'} \\
&\implies (x_1^T x_1)^{-1}x_1^T\hat{y} = (x_1^T x_1)^{-1}x_1^T x_1\hat{\beta}_1^{'} + (x_1^T x_1)^{-1}x_1^T x_2\hat{\beta}_2^{'} \\
&\implies (x_1^T x_1)^{-1}x_1^T\hat{y} = \hat{\beta}_1^{'} + \hat{\gamma}\hat{\beta}_2^{'} \\
&\implies (x_1^T x_1)^{-1}x_1^T(y - y_{resid}) = \hat{\beta}_1^{'} + \hat{\gamma}\hat{\beta}_2^{'} \\
&\implies (x_1^T x_1)^{-1}x_1^T y - (x_1^T x_1)^{-1}x_1^T y_{resid} = \hat{\beta}_1^{'} + \hat{\gamma}\hat{\beta}_2^{'} \\
&\implies \hat{\beta}_1 - (x_1^T x_1)^{-1}x_1^T y_{resid} = \hat{\beta}_1^{'} + \hat{\gamma}\hat{\beta}_2^{'}. \qquad (15)
\end{aligned}
$$

The term $(x_1^T x_1)^{-1}x_1^T y_{resid}$ drops out using Equation 14, so

$$\hat{\beta}_1 = \hat{\beta}_1^{'} + \hat{\gamma}\hat{\beta}_2^{'}. \qquad (16)$$

This is a simple but useful expansion on the omitted variable formula in Greene (2003). He describes the relationship between the true values of the coefficients. This shows that the relationship for the true values holds for the estimated values as well, which is necessary to understand the relationship between the estimated coefficients derived from regressions of the full model against those from omitted variable models. The next section will show how these results can be interpreted in 3 dimensions as directional derivatives of the regression plane in the full model.

5 Directional Derivatives

So far we have equations that demonstrate the relationship between β_1 and β', as well as their estimated values. Equation 15 can also be interpreted as a directional derivative in the regression plane defined by Equation 17. This interpretation highlights that although $\hat{\beta}_1$ is not estimated using information from the omitted variable x_2, it nonetheless reflects changes in x_2. This section reframes these equations as directional derivatives in a multivariate regression surface.

The first step is to restate the formulas above in terms of partial and directional derivatives.

$$\hat{y} = x_1\hat{\beta}_1' + x_2\hat{\beta}_2' \tag{17}$$

$$= x_1\frac{\partial \hat{y}}{\partial x_1} + x_2\frac{\partial \hat{y}}{\partial x_2} \tag{18}$$

That is, $\hat{\beta}_1'$ and $\hat{\beta}_2'$ are both partial derivatives of a multivariate regression plane. Partial derivatives assume that all else remains equal, and are sufficient to define a linear plane in terms of x_1 and x_2 as an orthogonal basis.

This can be equivalently framed as directional derivatives in the direction of the x_1 and x_2 axes. Consider the direction of the x_1 axis, represented by the unit vector $u = \left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$. Then $\hat{\beta}_1'$ is the directional derivative of $\hat{y}$ in the direction $\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$:

$$\begin{aligned}
\hat{\beta}_1 &= \begin{bmatrix} \frac{\partial \hat{y}}{\partial x_1} \frac{\partial \hat{y}}{\partial x_2} \end{bmatrix} \begin{bmatrix} 1 \\ \\ 0 \end{bmatrix} \\
&= \frac{\partial \hat{y}}{\partial x_1}, && (19) \\
\hat{\beta}_2 &= \begin{bmatrix} \frac{\partial \hat{y}}{\partial x_1} \frac{\partial \hat{y}}{\partial x_2} \end{bmatrix} \begin{bmatrix} 0 \\ \\ 1 \end{bmatrix} \\
&= \frac{\partial \hat{y}}{\partial x_2}. && (20)
\end{aligned}$$

Just as the partial derivatives from the full regression model can be interpreted as directional derivatives in the full regression plane, so to can $\hat{\beta}_1$.

$$\begin{aligned}
\hat{\beta}_1 &= \hat{\beta}_1' + \hat{\gamma}\hat{\beta}_2' \\
&= \begin{bmatrix} \hat{\beta}_1' \hat{\beta}_2' \end{bmatrix} \begin{bmatrix} 1 \\ \\ \hat{\gamma} \end{bmatrix} \\
&= \begin{bmatrix} \frac{\partial \hat{y}}{\partial x_1} \frac{\partial \hat{y}}{\partial x_2} \end{bmatrix} \begin{bmatrix} 1 \\ \\ \hat{\gamma} \end{bmatrix} && (21)
\end{aligned}$$

That is, $\hat{\beta}_1$ as a directional derivative in the regression plane in the direction $\left[\begin{smallmatrix} 1 \\ \hat{\gamma} \end{smallmatrix}\right]$.

References

Greene, W. H. (2003). *Econometric analysis* (6th ed.). Pearson Education India.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis* (Vol. 821). John Wiley & Sons.