Omitted variable bias

Aswin Thiruvengadam

3/31/2021

Omitted variable bias

Setting up the equations for the discussions

$$\text{Consider the equation}$$
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$
$$\text{Here X3 is the omitted variable and hence you do not know the value of } \beta_3$$
$$\text{We now write the relationship between } X_3 \text{ and } X_2 \text{ as:}$$
$$X_3 = \alpha_0 + \alpha_1 X_2$$
$$\text{From the omitted variable math we know:}$$
$$\beta_2^{measured} - \beta_2^{true} = \beta_3 \delta_1$$
$$\text{Rearranging:}$$
$$\beta_2^{true} = \beta_2^{measured} - \beta_3 \delta_1 \rightarrow (1)$$
$$\text{To save keystrokes I will now make these shortcuts:}$$
$$\beta_2^{true} \rightarrow \beta_2^{T}$$
$$\beta_2^{measured} \rightarrow \beta_2^{M}$$
$$\text{We can rewrite 1 using the shorthand:}$$
$$\beta_2^{T} = \beta_2^{M} - \beta_3 \delta_1 \rightarrow (2)$$

Couple of things to note here:

1. We do not know the magnitude of the omitted variable components but have a qualitative understanding of the direction, i.e. the sign
2. Sign of β_3 can be negative or positive independent of the sign of δ_1
3. The sign of the omitted variable bias is the sign of $\beta_3 \delta_1$

There can only be five scenarios now

Scenario 1:

$$\beta_3 \delta_1 = 0$$

This equates to no relationship either:

a. Between X3 and Y

b. X3 and X2
c. Or both

In any of these situations β_2 is unaffected. Whatever association was made between Y and X2 holds.

Scenario 2:

$$\beta_3\delta_1 > 0$$
$$\text{and}$$
$$\beta_2 > 0$$

Ignore X3 for now, what was the claim we made about Y's relationship with X2?

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

As $\beta_2 > 0$ our claim was "1 unit increase in X2 is associated with a β_2 increase in Y" OR in more general terms; **Increase in X2 is associated with an increase in Y**

Now lets include the information we have on $\beta_3\alpha_1$, i.e. it is greater than zero. Let's create some examples to explain what one should expect. In all of these example $\beta_2 = 5$

Example 1:

$$\beta_3\delta_1 = 2$$
$$\text{Using equation (2) from the top of this document}$$
$$\beta_2^T = \beta_2^M - \beta_3\delta_1$$
$$\beta_2^T = 5 - 2 = 3$$

In this context let's look at the general claim we made, **Increase in X2 is associated with an increase in Y**. Does that still hold good? The answer is yes because β_2^T is greater than zero

Example 2:

$$\beta_3\delta_1 = 7$$
$$\text{Using equation (2) from the top of this document}$$
$$\beta_2^T = \beta_2^M - \beta_3\delta_1$$
$$\beta_2^T = 5 - 7 = -2$$

In this context let's look again at the general claim we made, **Increase in X2 is associated with an increase in Y**. Does that still hold good? The answer is NO. As a matter of fact the relationship has flipped signs!! i.e. β_2^T flipped signs.

This is an example of **away from zero** bias, which is simply a shorthand statement to say the sign of the omitted variable bias(i.e. sign of $\beta_3\delta_1$) is the same as the sign of the coefficient of interest β_2. Note that it makes our claim about the association between β_2 and Y weaker. In the worst case it could make the claim completely wrong. Another way of saying the above is that the inclusion of the omitted variable makes β_2 move towards 0 and can even make it flip signs.

Scenario 3:

$$\beta_3\delta_1 < 0$$
$$\text{and}$$
$$\beta_2 < 0$$

Ignore X3 for now, what was the claim we made about Y's relationship with X2?

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

As $\beta_2 < 0$ our claim was "1 unit increase in X2 is associated with a β_2 decrease in Y" OR in more general terms; **Increase in X2 is associated with a decrease in Y**

Now lets include the information we have on $\beta_3\alpha_1$, i.e. it is lower than zero. Let's create some examples to explain what one should expect. In all of these example $\beta_2 = -5$.

Example 1:

$$\beta_3\delta_1 = -2$$
$$\text{Using equation (2) from the top of this document}$$
$$\beta_2^T = \beta_2^M - \beta_3\delta_1$$
$$\beta_2^T = -5 - (-2) = -3$$

In this context let's look at the general claim we made, **Increase in X2 is associated with a decrease in Y**. Does that still hold good? The answer is yes because β_2^T is lower than zero

Example 2:

$$\beta_3\delta_1 = -7$$
$$\text{Using equation (2) from the top of this document}$$
$$\beta_2^T = \beta_2^M - \beta_3\delta_1$$
$$\beta_2^T = 5 - (-7) = +2$$

In this context let's look again at the general claim we made, **Increase in X2 is associated with a decrease in Y** . Does that still hold good? The answer is NO. As a matter of fact the relationship has flipped signs!! i.e β_2^T flipped signs.

This is an example of **away from zero** bias, which is simply a shorthand statement to say the sign of the omitted variable bias(i.e. sign of $\beta_3\delta_1$) is the same as the sign of the coefficient of interest β_2. Note that it makes our claim about the association between β_2 and Y weaker. In the worst case it could make the claim completely wrong. Another way of saying the above is that the inclusion of the omitted variable makes β_2 move towards 0 and can even make it flip signs.

Scenario 4:

$$\beta_3\delta_1 > 0$$
$$\text{and}$$
$$\beta_2 < 0$$

Ignore X3 for now, what was the claim we made about Y's relationship with X2?

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

As $\beta_2 < 0$ our claim was "1 unit increase in X2 is associated with a β_2 decrease in Y" OR in more general terms; **Increase in X2 is associated with a decrease in Y**

Now lets include the information we have on $\beta_3\alpha_1$, i.e. it is greater than zero.Let's create some examples to explain what one should expect, in all these example $\beta_2 = -5$.

Example 1:

$$\beta_3\delta_1 = 0$$
$$\beta_3\delta_1 \text{ is always greater than 0, so the lowest value it can take is 0}$$
$$\text{Using equation (2) from the top of this document}$$
$$\beta_2^T = \beta_2^M - \beta_3\delta_1$$
$$\beta_2^T = -5 - 0 = -5$$

In this context let's look at the general claim we made, **Increase in X2 is associated with a decrease in Y** . Does that still hold good? The answer is yes because when $\beta_3\alpha_1$ is 0, $\beta_2^T = \beta_2^M$

Example 2:

$$\beta_3\delta_1 = 7$$
$$\text{Using equation (2) from the top of this document}$$
$$\beta_2^T = \beta_2^M - \beta_3\delta_1$$
$$\beta_2^T = -5 - 7 = -12$$

In this context let's look at the general claim we made, **Increase in X2 is associated with a decrease in Y**. Does that still hold good? The answer is yes, yes and absolute yes because β_2^T is lower than zero as a matter of fact it is lower than -5

This is an example of **towards zero** bias, which is simply a shorthand statement to say the sign of the omitted variable bias(i.e. sign of $\beta_3\delta_1$) is the opposite to the sign of the coefficient of interest β_2. Note that it makes our claim about the association between β_2 and Y stronger. In the worst case where $\beta_3\alpha_1$ is 0, β_2^T = β_2^M. What it is is saying is that the omitted variable is suppressing the relationship between β_2 and Y. So if we measure a coefficient in the absence of the omitted variable, the coefficient will only get bigger when the omitted variable is included. Which in turn means that we can strongly say that the general comment of **Increase in X2 is associated with a decrease in Y** holds.

Scenario 5:

$$\beta_3\delta_1 < 0$$
$$\text{and}$$
$$\beta_2 > 0$$

Ignore X3 for now, what was the claim we made about Y's relationship with X2?

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

As $\beta_2 > 0$ our claim was "1 unit increase in X2 is associated with a β_2 increase in Y" OR in more general terms; **Increase in X2 is associated with an increase in Y**

Now lets include the information we have on $\beta_3\alpha_1$, i.e. it is greater than zero. Let's create some examples to explain what one should expect, in all these example $\beta_2 = 5$.
Example 1:

$$\beta_3\delta_1 \text{ is always lower than 0, so the highest value it can take is 0}$$
$$\beta_3\delta_1 = 0$$
$$\text{Using equation (2) from the top of this document}$$
$$\beta_2^T = \beta_2^M - \beta_3\delta_1$$
$$\beta_2^T = 5 - 0 = 5$$

In this context let's look at the general claim we made, **Increase in X2 is associated with a decrease in Y** . Does that still hold good? The answer is yes because when $\beta_3\alpha_1$ is 0, $\beta_2^T = \beta_2^M$

Example 2:

$$\beta_3\delta_1 = -7$$

Using equation (2) from the top of this document

$$\beta_2^T = \beta_2^M - \beta_3\delta_1$$

$$\beta_2^T = 5 - (-7) = 12$$

In this context let's look at the general claim we made, **Increase in X2 is associated with a decrease in Y**. Does that still hold good? The answer is yes, yes and absolute yes because β_2^T is greater than than zero as a matter of fact it is greater than 5

This is an example of **towards zero** bias, which is simply a shorthand statement to say the sign of the omitted variable bias(i.e. sign of $\beta_3\delta_1$) is the opposite to the sign of the coefficient of interest β_2. Note that it makes our claim about the association between β_2 and Y stronger. In the worst case where $\beta_3\alpha_1$ is 0, β_2^T = β_2^M. What it is is saying is that the omitted variable is suppressing the relationship between β_2 and Y. So if we measure a coefficient in the absence of the omitted variable, the coefficient will only get bigger when the omitted variable is included. Which in turn means that we can strongly say that the general comment of **Increase in X2 is associated with an increase in Y** holds.

A table speaks a thousand words

	$\beta_3\alpha_1 > 0$	$\beta_3\,\alpha_1 < 0$
$\beta_2^M > 0$	• Away from zero • Makes claim weaker • β_2^T moves towards 0 and can flip signs	• Towards zero • Makes claim stronger • β_2^T moves away from 0
$\beta_2^M < 0$	• Towards zero • Makes claim stronger • β_2^T moves away from 0	• Away from zero • Makes claim weaker • β_2^T moves towards 0 and can flip signs

Figure 1: One table to rule them all