# Estimation of Multiple Regression Model

Christopher Taber

Department of Economics
University of Wisconsin-Madison

February 16, 2011

# Outline

The model we considered before with omitted variable bias was

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

and

$$u_i = \beta_2 z_i + \varepsilon_i$$

then we can write this as

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \varepsilon_i.$$

We were assuming that

$$E\left(\varepsilon_i \mid x_i, z_i\right) = 0.$$

The fact that we are calling this $\varepsilon_i$ rather than $u_i$ should not make any real difference.

Lets try to estimate this model.

We have three parameters to estimate $\hat{\beta}_0, \hat{\beta}_1$, and $\hat{\beta}_2$.

How can we do this?

The natural thing to do is based on what we did before except now we know the following three things

$$\begin{aligned} E(\varepsilon_i) &= 0 \\ E(x_i\varepsilon_i) &= 0 \\ E(z_i\varepsilon_i) &= 0 \end{aligned}$$

We write the sample regression function as

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{\beta}_2 z_i + \hat{\varepsilon}_i$$

We can take the sample analogue of this as before.

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\hat{\varepsilon}_i &= 0 \\
\frac{1}{N}\sum_{i=1}^{N}x_i\hat{\varepsilon}_i &= 0 \\
\frac{1}{N}\sum_{ii=1}^{N}z_i\hat{\varepsilon}_i &= 0
\end{aligned}$$

We can write this as

$$
\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i - \hat{\beta}_2 z_i\right) &= 0 \\
\frac{1}{N}\sum_{i=1}^{N} x_i\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i - \hat{\beta}_2 z_i\right) &= 0 \\
\frac{1}{N}\sum_{ii=1}^{N} z_i\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i - \hat{\beta}_2 z_i\right) &= 0
\end{aligned}
$$

This gives us three equations in the three unknowns $\hat{\beta}_0, \hat{\beta}_1$, and $\hat{\beta}_2$.

One could go through the algebra and solve explicitly these three equations for the three unknowns, but we are not going to bother to do that.

Instead we will just use stata

Rather than saying

**reg y x**

we say

**reg y x z**

Lets go back and look at the examples we thought about before

# Outline

# Adding Many Variables

Now we may be worried about more than one omitted variable at a time.

You can handle that in a straight forward way.

Lets write the model more generally with $K$ different regressors as

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_K x_{iK} + u_i.$$

The interpretation here is the same as before for the causal model

"If I increase $x_{i1}$ by one unit, $y_i$ increases by $\beta_1$ units"

# Estimation

How do we estimate this model?

It really is exactly the same thing once again.

We assume that

$$E\left(u_i \mid x_{i1}, ..., x_{iK}\right) = 0$$

We now have $K+1$ parameters, we need $K+1$ equations for estimation.

We start with the K+1 conditions

$$
\begin{aligned}
E(u_i) &= 0 \\
E(x_{i1} u_i) &= 0 \\
E(x_{i2} u_i) &= 0 \\
&\vdots \\
E(x_{iK} u_i) &= 0
\end{aligned}
$$

# Terminology

We define the sample regression function as

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + ... + \hat{\beta}_K x_{iK} + \hat{u}_i.$$

We will talk about and use the same notation for “fitted value”

$$\widehat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + ... + \hat{\beta}_K x_{iK}$$

and “residual”

$$\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - ... - \hat{\beta}_K x_{iK}.$$

We start with the $K+1$ conditions

$$
\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\hat{u}_i &= 0 \\
\frac{1}{N}\sum_{i=1}^{N}x_{i1}\hat{u}_i &= 0 \\
\frac{1}{N}\sum_{i=1}^{N}x_{i2}\hat{u}_i &= 0 \\
&\vdots \\
\frac{1}{N}\sum_{i=1}^{N}x_{iK}\hat{u}_i &= 0
\end{aligned}
$$

And you can rewrite these as

$$
\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\left(y_i-\hat{\beta}_0-\hat{\beta}_1 x_{i1}-\hat{\beta}_2 x_{i2}-\ldots-\hat{\beta}_K x_{iK}\right) &= 0 \\
\frac{1}{N}\sum_{i=1}^{N}x_{i1}\left(y_i-\hat{\beta}_0-\hat{\beta}_1 x_{i1}-\hat{\beta}_2 x_{i2}\ldots-\hat{\beta}_K x_{iK}\right) &= 0 \\
\frac{1}{N}\sum_{i=1}^{N}x_{i2}\left(y_i-\hat{\beta}_0-\hat{\beta}_1 x_{i1}-\hat{\beta}_2 x_{i2}\ldots-\hat{\beta}_K x_{iK}\right) &= 0 \\
&\vdots \\
\frac{1}{N}\sum_{i=1}^{N}x_{iK}\left(y_i-\hat{\beta}_0-\hat{\beta}_1 x_{i1}-\hat{\beta}_2 x_{i2}\ldots-\hat{\beta}_K x_{iK}\right) &= 0
\end{aligned}
$$

Figuring out the algebra is extremely tedious, so once again we will let stata do it.

We just say

$$\textbf{reg}\ y\ x_1\ x_2\ ...\ x_K$$

Lets look at some examples in which we control for a number of variables at the same time.

# Descriptive Interpretation

We are focusing on the causal interpretation of this, but one can interpret it in a descriptive way as well

We just interpret the model as

$$E\left(y \mid x_1, x_2, ..., x_K\right) = \beta_0 + \beta_1 x_1 + ... + \beta_K x_K$$

# Forecasting

We talked before of forecasting with one variable, but that is kind of crazy.

Generally we have a lot of variables to forecast with so lets think about using them

We will use the same sort of notation from before, we want to choose $\hat{\beta}_0$,$\hat{\beta}_1$,...,$\hat{\beta}_K$ to minimize

$$\sum_{i=1}^{N}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - ... - \hat{\beta}_K x_{iK}\right)^2$$

# How do we do this?

As before we take the derivative of this expression with respect to each $\hat{\beta}_j$ and set it to zero.

That will give us K+1 equations in K+1 unknowns

Lets look at them

First consider the derivative with respect to $\hat{\beta}_0$

$$
\begin{aligned}
0 &= \frac{\partial \sum_{i=1}^{N} \left( y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - ... - \hat{\beta}_K x_{iK} \right)^2}{\partial \hat{\beta}_0} \\
&= \sum_{i=1}^{N} \frac{\partial \left( y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - ... - \hat{\beta}_K x_{iK} \right)^2}{\partial \hat{\beta}_0} \\
&= \sum_{i=1}^{N} -2 \left( y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - ... - \hat{\beta}_K x_{iK} \right)
\end{aligned}
$$

which after dividing each side by -2N we get

$$
0 = \frac{1}{N} \sum_{i=1}^{N} \left( y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - ... - \hat{\beta}_K x_{iK} \right)
$$

This is the first expression from above.

Next take the derivative with respect to $\hat{\beta}_j$

This will have the same form for $j = 1, ..., K$

$$
\begin{aligned}
0 &= \frac{\partial \sum_{i=1}^{N} \left( y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - ... - \hat{\beta}_K x_{iK} \right)^2}{\partial \hat{\beta}_j} \\
&= \sum_{i=1}^{N} \frac{\partial \left( y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - ... - \hat{\beta}_K x_{iK} \right)^2}{\partial \hat{\beta}_j} \\
&= \sum_{i=1}^{N} -2 x_{ij} \left( y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - ... - \hat{\beta}_K x_{iK} \right)
\end{aligned}
$$

which after dividing each side by -2N we get

$$0 = \frac{1}{N}\sum_{i=1}^{N} x_{ij}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - ... - \hat{\beta}_K x_{Ki}\right)$$

There are K different expressions like this

They are the other K expressions we had from above

Thus we won't explicitly derive the OLS estimator for this case

However, you get exactly the same answer as if you did it the other way

# Normal Equations

Also the normal equations hold exactly as before

Now there are K+1 of them rather than just 2

That is we know that for any given sample

$$
\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\hat{u}_i &= 0 \\
\frac{1}{N}\sum_{i=1}^{N}x_{i1}\hat{u}_i &= 0 \\
\frac{1}{N}\sum_{i=1}^{N}x_{i2}\hat{u}_i &= 0 \\
&\vdots \\
\frac{1}{N}\sum_{i=1}^{N}x_{iK}\hat{u}_i &= 0
\end{aligned}
$$

# Outline

Wooldridge talks about three facts about these Models

They all come from the Normal Equations

Lets go through them

# 1. Sample average of the predicted values

Notice that

$$\begin{aligned}
\overline{\widehat{y}} &= \frac{1}{N}\sum_{i=1}^{N}\widehat{y}_i \\
&= \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \widehat{u}_i\right) \\
&= \frac{1}{N}\sum_{i=1}^{N} y_i - \frac{1}{N}\sum_{i=1}^{N}\widehat{u}_i \\
&= \overline{y}
\end{aligned}$$

# 2. Sample covariance of predicted value and residual is zero

This comes immediately

$$
\begin{aligned}
&\frac{1}{N-1}\sum_{i=1}^{N}\widehat{y}_i\widehat{u}_i \\
&= \frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\beta}_0+\hat{\beta}_1 x_{i1}+\hat{\beta}_2 x_{i2}+...+\hat{\beta}_K x_{iK}\right)\widehat{u}_i \\
&= \frac{1}{N-1}\left[\sum_{i=1}^{N}\hat{\beta}_0\widehat{u}_i+\sum_{i=1}^{N}\hat{\beta}_1 x_{i1}\widehat{u}_i+\sum_{i=1}^{N}\hat{\beta}_2 x_{i2}\widehat{u}_i+...+\sum_{i=1}^{N}\hat{\beta}_K x_{iK}\widehat{u}_i\right] \\
&= \frac{1}{N-1}\left[\hat{\beta}_0\sum_{i=1}^{N}\widehat{u}_i+\hat{\beta}_1\sum_{i=1}^{N}x_{i1}\widehat{u}_i+\hat{\beta}_2\sum_{i=1}^{N}x_{i2}\widehat{u}_i+...+\hat{\beta}_K\sum_{i=1}^{N}x_{iK}\widehat{u}_i\right] \\
&= 0
\end{aligned}
$$

# 3. Mean of Regressors and Regression Line

First recall that in the simple regression model we showed that

$$\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x}$$

Which also can be written as

$$\overline{y} = \widehat{\beta}_0 + \widehat{\beta}_1 \overline{x}$$

Do we get something similar here?

The answer is yes.

Lets take the equation

$$
\begin{aligned}
0 =& \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - ... - \hat{\beta}_K x_{iK}\right) \\
=& \frac{1}{N}\sum_{i=1}^{N} y_i - \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_0 - \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_1 x_{i1} - \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_2 x_{i2} - ... - \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_K x_{iK} \\
=& \overline{y} - \hat{\beta}_0 - \hat{\beta}_1 \overline{x}_1 - \hat{\beta}_2 \overline{x}_2 - ... - \hat{\beta}_K \overline{x}_K
\end{aligned}
$$

or

$$\overline{y} = \hat{\beta}_0 + \hat{\beta}_1 \overline{x}_1 + \hat{\beta}_2 \overline{x}_2 + ... + \hat{\beta}_K \overline{x}_K$$

This means that this point must be on the OLS regression line

# Outline

# Goodness of Fit

Recall that before we defined

$$
\begin{aligned}
SST &= \sum_{i=1}^{N} (y_i - \overline{y})^2 \\
SSE &= \sum_{i=1}^{N} \left(\widehat{y}_i - \overline{y}\right)^2 \\
SSR &= \sum_{i=1}^{N} \widehat{u}_i^2
\end{aligned}
$$

We define them exactly the same was for the multiple regression model

It is straight forward to show that once again

$$SST = SSE + SSR.$$

We can still use the

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

However there is an interesting property here

What happens to the $R^2$ when we add a regressor to the model?

Intuitively it seems like it should go up-in fact that is mathematically true.

Consider the sample regression functions:

$$\begin{aligned} y_i &= \widehat{\beta}_0 + \widehat{\beta}_1 x_{i1} + \widehat{u}_i \\ y_i &= \widehat{\gamma}_0 + \widehat{\gamma}_1 x_{i1} + \widehat{\gamma}_2 x_{i2} + \widehat{v}_i \end{aligned}$$

Both models have the same SST, consider the two residual sum of squares

$$\sum_{i=1}^{N}\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_{i1}\right)^2$$

$$\sum_{i=1}^{N}(y_i - \widehat{\gamma}_0 - \widehat{\gamma}_1 x_{i1} - \widehat{\gamma}_2 x_{i2})^2$$

# How do they compare?

Recall that $\widehat{\gamma}_0, \widehat{\gamma}_1$,and $\widehat{\gamma}_2$ were chosen to minimize this sum of squares.

I could pick

$$\begin{aligned} \widehat{\gamma}_0 &= \widehat{\beta}_0 \\ \widehat{\gamma}_1 &= \widehat{\beta}_1 \\ \widehat{\gamma}_2 &= 0 \end{aligned}$$

If that was the case then the SSR would be the same

However, I can almost certainly do better in which case the SSR would be lower in the second model and thus the $R^2$ would be higher.

This means if you choose your regression model by choosing the one with the highest $R^2$ you will always add more variables

# Adjusted $R^2$

This is the justification of the adjusted $R^2$. It is described in Wooldridge chapter 6 and is defined as

$$\overline{R}^2 = 1 - \frac{SSR/(N-K-1)}{SST/(N-1)} = 1 - \frac{(1-R^2)(N-1)}{N-K-1}$$

Notice that holding the $R^2$ fixed, when $K$ rises the $\overline{R}^2$ falls

Thus there is a "penalty" for adding parameters

Some people choose their models by maximizing the adjusted R-squared

This is not something economists worry about much at all

It is important to note that the argument included only nested models

That is the variables in the first model were a subset of the variables in the second model

This argument would not work if we compared the two models

$$\sum_{i=1}^{N}\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_{i1}\right)^2$$

$$\sum_{i=1}^{N}\left(y_i - \widehat{\gamma}_0 - \widehat{\gamma}_2 x_{i2} - \widehat{\gamma}_3 x_{i3}\right)^2$$

In this case we can't say anything about which is better

Lets look at some examples

# Outline

Before we derived the fact that OLS was unbiased.

We can use similar conditions here

Lets go through them

**Assumption MLR.1-Linear in Parameters**

The model in the population can be written as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_K x_K + u,$$

where $\beta_0$,$\beta_1, .., \beta_K$ are the unknown parameters (constants) of interest and $u$ is an unobservable random error or disturbance term.

**Assumption MLR.2-Random Sampling**

We have a random sample of $n$ observations, $\{(x_{i1}, x_{i2}, ..., x_{iK}, y_i)\}, i = 1, ..., n$ following the population model defined in MLR.1.

**Assumption MLR.3-No Perfect Collinearity**

In the sample (and therefore in the population), none of the independent variables is constant, and there are no exact linear relationships among the independent variables.

This one is different but related to the assumption that there is variation in $x$. Now not only do we need variation, but we need them to vary separately.

One case which is clearly a problem is if

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i + u_i$$

How could you possibly tell $\beta_1$ apart from $\beta_2$?

This generalizes and is still a problem if

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$$

and

$$x_{i2} = \alpha_0 + \alpha_1 x_{i1}.$$

To see why note that we could just rewrite this model as

$$\begin{aligned}
y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i \\
&= \beta_0 + \beta_1 x_{i1} + \beta_2 \left(\alpha_0 + \alpha_1 x_{i1}\right) + u_i \\
&= \left(\beta_0 + \beta_2 \alpha_0\right) + \left(\beta_1 + \beta_2 \alpha_1\right) x_{i1} + u_i
\end{aligned}$$

which is really just a univariate regression.

## Assumption MLR.4-Zero Conditional Mean

The error $u$ has an expected value of zero given any value of the independent variables. In other words

$$E\left(u \mid x_1, x_2, ..., x_K\right) = 0.$$

Putting these together we get

**Theorem 3.1 Unbiasedness of OLS**

Under Assumptions MLR.1 through MLR.4,

$$E\left(\widehat{\beta}_j\right) = \beta_j, j = 0, 1, ..., K,$$

for any values of the population parameters $\beta_j$.In other words the OLS estimators are unbiased estimators of the population parameters.

# Outline

# Omitted Variable Bias

Wooldridge talks about omitted variables in the more general model.

I want to do it a somewhat different manner

Think about omitted variables in the following way

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + u_i$$

where

$$E\left(u_i \mid x_{i1}, x_{i2}, x_{i3}\right) = 0$$

# What happens if we don't have data on $x_{i3}$?

Think about regressing $x_{i3}$ on $x_{i2}$ and $x_{i3}$.

Write this as

$$x_{i3} = \delta_0 + \delta_1 x_{i1} + \delta_{i2} x_{i2} + \xi_i$$

where

$$E\left(\xi_i \mid x_{i1}, x_{i2}\right) = 0$$

then

$$
\begin{aligned}
y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + u_i \\
&= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 \left(\delta_0 + \delta_1 x_{i1} + \delta_{i2} x_{i2} + \xi_i\right) + u_i \\
&= \left(\beta_0 + \beta_3 \delta_0\right) + \left(\beta_1 + \beta_3 \delta_1\right) x_{i1} + \left(\beta_2 + \beta_3 \delta_2\right) x_{i2} + \left(\beta_3 \xi_i + u_i\right)
\end{aligned}
$$

Thus the omitted variable bias is close to what it is before: $\beta_3 \delta_1$

The only real difference is that $\delta_1$ is more complicated because it comes from a multiple regression rather than a simple one.

# Outline

# Gauss Markov

There is an important theorem to look at, but we need one more assumption first

**Assumption MLR.5 Homoskedasticity**

The error $u$ has the same variance given any values of the explanatory variables. In other words, $Var(u \mid x_1, ..., x_K) = \sigma^2$.

When We combine all of these assumptions we get the Gauss-Markov Theorem

**Theorem 3.4 Gauss-Markov Theorem**

Under Assumptions MLR.1-MLR.5, $\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_K$ are the Best Linear Unbiased Estimators (BLUEs) of $\beta_0, \beta_1, ..., \beta_K$, respectively.