Multiple Regression Models
ECON 30331

Bill Evans
Fall 2014

Assumption 1: Model is linear in parameters

$$y_i = \beta_0 + x_{1i}\beta_1 + x_{2i}\beta_2 + x_{3i}\beta_3 + ...x_{ki}\beta_k + \varepsilon_i$$

Assumption 2: All parameters are estimable

For this to be the case, at least two things have to be true.

First, for each independent variable, there must be variation within the n observations for each of the x_{ji}'s. What does this mean? Well, suppose we have a simple bivariate model of the form

$$y_i = \beta_0 + x_{1i}\beta_1 + \varepsilon_i$$

We know the OLS estimate for β_1 would be

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_{1i} - \bar{x}_1)}{\sum_{i=1}^{n}(x_{1i} - \bar{x}_1)^2}$$

Now suppose that there is no variation in the x_{1i} that is x_{1i}=a constant k for all observations i. In this case, x_{1i}= $\bar{x}_1$ and the denominator in the estimate for $\hat{\beta}_1$ is zero. Therefore, when there is no variation in the sample in x_{1i} one cannot estimate the model.

Second, each variable added to the model must provide some new piece of information. Another way of saying this is that none of the variables can be a linear combination of the remaining variables in the model. Suppose there is a model with 3 covariates

$$y_i = \beta_0 + x_{1i}\beta_1 + x_{2i}\beta_2 + x_{3i}\beta_3 + \varepsilon_i$$

But variable β_3 is a linear combination of the other variables in the model. So for example, suppose $x_{3i} = a + bx_{1i} + cx_{2i}$ where a, b and c are constants. As a concrete example, suppose that x_1 is the fraction < 24 years of age, x_2 is the fraction25 to 64, and x_3 is the fraction 65 and over. Note that by construction $x_1+x_2+x_3=1$. Therefore if you know x_1 and x_2, you know exactly x_3. Let's see what this does to the model.

Substituting the definition $x_{3i} = a + bx_{1i} + cx_{2i}$ into the equation above, we see that

$$y_i = \beta_0 + x_{1i}\beta_1 + x_{2i}\beta_2 + (a + bx_{1i} + cx_{2i})\beta_3 + u_i$$

But when we collect like terms, we see that

$$y_i = (\beta_0 + a\beta_3) + x_{1i}(\beta_1 + b\beta_3) + x_{2i}(\beta_2 + c\beta_3) + \varepsilon_i$$

Finally note that in actuality, we can only estimate the following model

$$y_i = \theta_0 + x_{1i}\theta_1 + x_{2i}\theta_2 + u_i$$

Where $\theta_0 = (\beta_0 + a\beta_3),\ \theta_1 = (\beta_1 + b\beta_3),\ and\ \theta_2 = (\beta_2 + c\beta_3)$. In this case, the model has 4 parameters (the betas) but you only have 3 degrees of freedom (the thetas) so the parameters are under-identified – 3 equations but 4 unknowns. Therefore the original 3 covariate model CANNOT be estimated.

Deriving estimates in the multivariate case:

Model: $y_i = \beta_0 + x_{1i}\beta_1 + x_{2i}\beta_2 + x_{3i}\beta_3 + \ldots.x_{ki}\beta_k + \varepsilon_i$

There are: n observations
k continuous exogenous variables
1 constant

Estimated error: $\hat{\varepsilon}_i = y_i - \hat{\beta}_0 - x_{1i}\hat{\beta}_1 - x_{2i}\hat{\beta}_2 - \ldots.x_{ki}\hat{\beta}_k$

Objective: $SSR = \sum_{i=1}^{n}\hat{\varepsilon}_i^2 = \sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - x_i\hat{\beta}_1 - x_{2i}\hat{\beta}_2 - \ldots.x_{ki}\hat{\beta}_k\right)^2$

There are k+1 unknowns so we need k+1 first order conditions (FOCs) to identify the model

FOC 1: $\dfrac{\partial SSR}{\partial\hat{\beta}_1} = -2\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - x_{1i}\hat{\beta}_1 - x_{2i}\hat{\beta}_2 - \ldots.x_{ki}\hat{\beta}_k\right)x_{1i} = 0$

FOC 2: $\dfrac{\partial SSR}{\partial\hat{\beta}_2} = -2\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - x_{1i}\hat{\beta}_1 - x_{2i}\hat{\beta}_2 - \ldots.x_{ki}\hat{\beta}_k\right)x_{2i} = 0$

: :

: :

FOC k: $\dfrac{\partial SSR}{\partial\hat{\beta}_k} = -2\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - x_{1i}\hat{\beta}_1 - x_{2i}\hat{\beta}_2 - \ldots.x_{ki}\hat{\beta}_k\right)x_{ki} = 0$

FOC k+1: $\dfrac{\partial SSR}{\partial\hat{\beta}_0} = -2\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - x_{1i}\hat{\beta}_1 - x_{2i}\hat{\beta}_2 - \ldots.x_{ki}\hat{\beta}_k\right) = 0$

This is a system of k+1 unknowns and k equations.

Each of these equations has a -2 in front which can be eliminated by multiplying both sides by a -1/2.

$$\text{FOC 1: } \sum_{i=1}^{n}\left(y_i - \widehat{\beta}_0 - x_{1i}\widehat{\beta}_1 - x_{2i}\hat{\beta}_2 - \ldots . x_{ki}\hat{\beta}_k\right)x_{1i} = 0$$

$$\text{FOC 2: } \sum_{i=1}^{n}\left(y_i - \widehat{\beta}_0 - x_{1i}\widehat{\beta}_1 - x_{2i}\hat{\beta}_2 - \ldots . x_{ki}\hat{\beta}_k\right)x_{2i} = 0$$

$$\text{FOC k: } \sum_{i=1}^{n}\left(y_i - \widehat{\beta}_0 - x_{1i}\widehat{\beta}_1 - x_{2i}\hat{\beta}_2 - \ldots . x_{ki}\hat{\beta}_k\right)x_{ki} = 0$$

$$\text{FOC K+1: } \sum_{i=1}^{n} y_i - n\hat{\beta}_0 - \hat{\beta}_1\sum_{i=1}^{n} x_{1i} - \hat{\beta}_2\sum_{i=1}^{n} x_{2i} - \ldots \hat{\beta}_k\sum_{i=1}^{n} x_{ki} = 0$$

Expanding the terms under the summation:

$$\text{FOC 1: } \hat{\beta}_0\sum_{i=1}^{n} x_{1i} + \hat{\beta}_1\sum_{i=1}^{n} x_{1i}^2 + \hat{\beta}_2\sum_{i=1}^{n} x_{2i}x_{1i} + \ldots \hat{\beta}_k\sum_{i=1}^{n} x_{ki}x_{1i} = \sum_{i=1}^{n} y_i x_{1i}$$

$$\text{FOC 2: } \hat{\beta}_0\sum_{i=1}^{n} x_{2i} + \hat{\beta}_1\sum_{i=1}^{n} x_{1i}x_{2i} + \hat{\beta}_2\sum_{i=1}^{n} x_{2i}^2 + \ldots \hat{\beta}_k\sum_{i=1}^{n} x_{ki}x_{2i} = \sum_{i=1}^{n} y_i x_{2i}$$

: :

: :

: :

$$\text{FOC k: } \hat{\beta}_0\sum_{i=1}^{n} x_{ki} + \hat{\beta}_1\sum_{i=1}^{n} x_{1i}x_{ki} + \hat{\beta}_2\sum_{i=1}^{n} x_{2i}x_{ki} + \ldots \hat{\beta}_k\sum_{i=1}^{n} x_{ki}{}^2 = \sum_{i=1}^{n} y_i x_{ki}$$

$$\text{FOC k+1: } n\hat{\beta}_0 + \hat{\beta}_1\sum_{i=1}^{n} x_{1i} + \hat{\beta}_2\sum_{i=1}^{n} x_{2i} + \ldots \hat{\beta}_k\sum_{i=1}^{n} x_{ki} = \sum_{i=1}^{n} y_i$$

This is a system of k+1 equations and k+1 unknowns. For k+1 parameters, it is difficult to show what the solution to this system is without linear algebra. [For those who have NOT had linear algebra-skip to the next section]. However for those who have had linear algebra, this is a pretty straightforward problem. Write the k+1 equations in matrix notation

$$\begin{bmatrix} \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{1i}^2 & \sum_{i=1}^{n} x_{1i}x_{2i} \dots & \sum_{i=1}^{n} x_{1i}x_{ki} \\ \sum_{i=1}^{n} x_{2i} & \sum_{i=1}^{n} x_{1i}x_{2i} & \sum_{i=1}^{n} x_{2i}^2 \dots & \sum_{i=1}^{n} x_{2i}x_{ki} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{i=1}^{n} x_{ki} & \sum_{i=1}^{n} x_{1i}x_{ki} & \sum_{i=1}^{n} x_{2i}x_{ki} & \sum_{i=1}^{n} x_{ki}^2 \\ n & \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{2i} & \sum_{i=1}^{n} x_{k} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \\ \beta_0 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} x_{1i}y_i \\ \sum_{i=1}^{n} x_{2i}y_i \\ \vdots \\ \sum_{i=1}^{n} x_{ki}y_i \\ \sum_{i=1}^{n} y_i \end{bmatrix}$$

And therefore, the key problem is the inversion of a (k+1) x (k+1) matrix

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \\ \hat{\beta}_0 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{1i}^2 & \sum_{i=1}^{n} x_{1i}x_{2i} \dots & \sum_{i=1}^{n} x_{1i}x_{ki} \\ \sum_{i=1}^{n} x_{2i} & \sum_{i=1}^{n} x_{1i}x_{2i} & \sum_{i=1}^{n} x_{2i}^2 \dots & \sum_{i=1}^{n} x_{2i}x_{ki} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{i=1}^{n} x_{ki} & \sum_{i=1}^{n} x_{1i}x_{ki} & \sum_{i=1}^{n} x_{2i}x_{ki} & \sum_{i=1}^{n} x_{ki}^2 \\ n & \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{2i} & \sum_{i=1}^{n} x_{k} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} x_{1i}y_i \\ \sum_{i=1}^{n} x_{2i}y_i \\ \vdots \\ \sum_{i=1}^{n} x_{ki}y_i \\ \sum_{i=1}^{n} y_i \end{bmatrix}$$

Some properties of the OLS estimates

The mean of $\hat{\varepsilon}_i$ is still zero

Recall that the k+1 FOC is

$$\frac{\partial SSR}{\partial \widehat{\beta}_0} = -2\sum_{i=1}^{n}\left(y_i - \widehat{\beta}_0 - x_{1i}\widehat{\beta}_1 - x_{2i}\hat{\beta}_2 - \dots . x_{ki}\hat{\beta}_k\right) = 0$$

Which can be written as

$$\sum_{i=1}^{n}\left(y_i - \widehat{\beta}_0 - x_{1i}\widehat{\beta}_1 - x_{2i}\hat{\beta}_2 - \dots . x_{ki}\hat{\beta}_k\right) = 0$$

Recall also that $\hat{\varepsilon}_i = y_i - \hat{\beta}_0 - x_{1i}\hat{\beta}_1 - x_{2i}\hat{\beta}_2 - \dots . x_{ki}\hat{\beta}_k$ so

$$\sum_{i=1}^{n}\left(\hat{\varepsilon}_i\right) = 0$$

Which means that once again $\overline{\hat{\varepsilon}} = 0$

It is still a mean regression

Recall that the k+1 FOC

$$\frac{\partial SSR}{\partial \widehat{\beta}_0} = -2\sum_{i=1}^{n}\left(y_i - \widehat{\beta}_0 - x_{1i}\widehat{\beta}_1 - x_{2i}\hat{\beta}_2 - \ldots . x_{ki}\hat{\beta}_k\right) = 0$$

Can be written as

$$\sum_{i=1}^{n} y_i - n\hat{\beta}_0 - \hat{\beta}_1\sum_{i=1}^{n} x_{1i} - \hat{\beta}_2\sum_{i=1}^{n} x_{2i} - \ldots \hat{\beta}_k\sum_{i=1}^{n} x_{ki} = 0$$

Solving for $n\hat{\beta}_0$, we get that

$$n\hat{\beta}_0 = \sum_{i=1}^{n} y_i - \hat{\beta}_1\sum_{i=1}^{n} x_{1i} - \hat{\beta}_2\sum_{i=1}^{n} x_{2i} - \ldots \hat{\beta}_k\sum_{i=1}^{n} x_{ki}$$

Dividing through by n, and recognizing that $\frac{1}{n}\sum_{i=1}^{n} x_{ji} = \overline{x}_j$ we get that

$$\hat{\beta}_0 = \overline{y} - \overline{x}_1\hat{\beta}_1 - \overline{x}_2\hat{\beta}_2 - \ldots\ldots \overline{x}_k\hat{\beta}_k$$

Any therefore

$$\overline{y} = \hat{\beta}_0 + \overline{x}_1\hat{\beta}_1 + \overline{x}_2\hat{\beta}_2 + \ldots\ldots \overline{x}_k\hat{\beta}_k$$

The regression still fits the means of x's through the means of the y's

The correlation between $\hat{\varepsilon}_i$ and x_{ji} is still zero

Recall the definition of the first order condition for a particular variable k

$$\frac{\partial SSR}{\partial \widehat{\beta}_k} = -2\sum_{i=1}^{n}\left(y_i - \widehat{\beta}_0 - x_{1i}\widehat{\beta}_1 - x_{2i}\hat{\beta}_2 - \ldots . x_{ki}\hat{\beta}_k\right)x_{ki} = 0$$

Which can be written as

$$\sum_{i=1}^{n}\left(y_i - \widehat{\beta}_0 - x_{1i}\widehat{\beta}_1 - x_{2i}\hat{\beta}_2 - \ldots . x_{ki}\hat{\beta}_k\right)x_{ki} = 0$$

Recall also that $\hat{\varepsilon}_i = y_i - \hat{\beta}_0 - x_{1i}\hat{\beta}_1 - x_{2i}\hat{\beta}_2 - \ldots . x_{ki}\hat{\beta}_k$ so

$$\sum_{i=1}^{n}\left(\hat{\varepsilon}_i\right)x_{ki} = 0$$

So by construction, the estimated residuals are uncorrelated (independent) of the actual x's

The Bias Associated with Omitted Variables in the Multivariate model

Recall that the estimated parameters are a function of the Y_i's and the Y_i's are a function of ε_i's which is the true source of randomness in the model. Therefore, the properties of the estimates will be a function of the properties of the estimates are a function of the ε_i's .

The linear model is $y_i = \beta_0 + x_{1i}\beta_1 + x_{2i}\beta_2 + x_{3i}\beta_3 + ...x_{ki}\beta_k + \varepsilon_i$ and so we will have to simply expand some of the assumptions. Our assumptions concerning the errors are now conditioned on all the covariates in the model

1) $E[\varepsilon_i] = E[\varepsilon_i \mid x_{1i}, x_{2i}, ...x_{ki}] = 0$
2) $\text{Cov}\left(\varepsilon_i,\ \varepsilon_j\right) = 0 \text{ for all } i \neq j$
3) $Var(\varepsilon_i) = Var(\varepsilon_i \mid x_{1i}, x_{2i}, ...x_{ki}) = \sigma_\varepsilon^2$

In the simple bivariate model, we assumed ε_i and x_i we uncorrelated. Now, we assume ε_i is uncorrelated with each of the x's.

Assumptions 2) and 3) are essentially used to identify the variance of the estimate. As before, the key assumption for whether the estimates are unbiased is assumption 1.

To illustrate the properties of the multivariate model, let's reduce the dimension of the problem somewhat. Suppose the true model is one with two variables $y_i = \beta_0 + x_{1i}\beta_1 + x_{2i}\beta_2 + \varepsilon_i$

But the researcher only estimates a regression with one variable, x_{1i}. Maintain assumption 1) above which implies that $E(x_{1i}\varepsilon_i)=0$ and $E(x_{2i}\varepsilon_i)=0$ (the real errors are uncorrelated with covariates). Let $\tilde{\beta}_1$ represent the OLS estimate from the simple bivariate regression of y on x_1. From the previous section, we know that

(1) $$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_{1i} - \bar{x}_1)}{\sum_{i=1}^{n}(x_{1i} - \bar{x}_1)^2}$$

Use the fact that $\sum_{i=1}^{n}(y_i - \bar{y})(x_{1i} - \bar{x}_1) = \sum_{i=1}^{n} y_i(x_{1i} - \bar{x}_1)$ and substitute the true value for y into the numerator, $\sum_{i=1}^{n} y_i(x_{1i} - \bar{x}_1) = \sum_{i=1}^{n}(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i)(x_{1i} - \bar{x}_1)$. Finally, let the denominator equal SST_{x1}. Equation (1) can be written as

(2) $$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_{1i} - \bar{x}_1)}{\sum_{i=1}^{n}(x_{1i} - \bar{x}_1)^2} = \frac{\sum_{i=1}^{n}(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i)(x_{1i} - \bar{x}_1)}{SST_{x1}}$$

Expand the numerator:

$$(3) \qquad \tilde{\beta}_1 = \frac{\sum_{i=1}^{n}\beta_0(x_{1i}-\bar{x}_1)+\sum_{i=1}^{n}\beta_1 x_{1i}(x_{1i}-\bar{x}_1)+\sum_{i=1}^{n}\beta_2 x_{2i}(x_{1i}-\bar{x}_1)+\sum_{i=1}^{n}\varepsilon_i(x_{1i}-\bar{x}_1)}{SST_{x1}}$$

There are four terms in the numerator

$$(term\,1) \qquad \frac{\sum_{i=1}^{n}\beta_0(x_{1i}-\bar{x}_1)}{SST_{x1}} = \frac{\beta_0\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)}{SST_{x1}} = \frac{\beta_0 0}{SST_{x1}} = 0$$

$$(term\,2) \qquad \frac{\sum_{i=1}^{n}\beta_1 x_{1i}(x_{1i}-\bar{x}_1)}{SST_{x1}} = \frac{\sum_{i=1}^{n}\beta_1(x_{1i}-\bar{x}_1)(x_{1i}-\bar{x}_1)}{SST_{x1}} = \frac{\beta_1\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2}{SST_{x1}} = \beta_1$$

$$(term\,3) \qquad \frac{\sum_{i=1}^{n}\beta_2 x_{2i}(x_{1i}-\bar{x}_1)}{SST_{x1}} = \frac{\beta_2\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)(x_{1i}-\bar{x}_1)}{SST_{x1}}$$

$$(term\,4) \qquad \frac{\sum_{i=1}^{n}\varepsilon_i(x_{1i}-\bar{x}_1)}{SST_{x1}} = \frac{\sum_{i=1}^{n}\varepsilon_i(x_{1i}-\bar{x}_1)}{SST_{x1}}$$

Term 1 drops out, term 2 reduces to β_1, and substituting the definitions for terms 3 and 4 into equation (3).

$$(3) \qquad \tilde{\beta}_1 = \beta_1 + \frac{\beta_2\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)(x_{1i}-\bar{x}_1)}{SST_{x1}} + \frac{\sum_{i=1}^{n}\varepsilon_i(x_{1i}-\bar{x}_1)}{SST_{x1}}$$

Now, take the expectations of both sides

$$(5) \qquad E[\tilde{\beta}_1] = E[\beta_1] + E\left[\frac{\beta_2\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)(x_{1i}-\bar{x}_1)}{SST_{x1}}\right] + E\left[\frac{\sum_{i=1}^{n}\varepsilon_i(x_{1i}-\bar{x}_1)}{SST_{x1}}\right]$$

By definition: $E[\beta_1]=\beta_1$ because β_1 is a constant

Note also in the final term that because we still maintain $\text{cov}(x_{1i}, \varepsilon_i)=0$

$$E\left[\frac{\sum_{i=1}^{n}\varepsilon_i(x_{1i}-\bar{x}_1)}{SST_{x1}}\right]=0$$

Now, work on the middle term

$$E\left[\frac{\beta_2\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)(x_{1i}-\bar{x}_1)}{SST_{x1}}\right]=\left[\frac{\beta_2\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)(x_{1i}-\bar{x}_1)}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2}\right]$$

because all variables are assumed to be fixed

$$E[\tilde{\beta}_1]=\beta_1+\beta_2\frac{\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)(x_{1i}-\bar{x}_1)}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2} \tag{6}$$

Note that if we were to run a synthetic regression of x_{2i} on x_{1i}, $x_{2i}=\delta_0+\delta_1 x_{1i}+\phi_i$, the estimate for $\hat{\delta}_1$ would be

$$\hat{\delta}_1=\frac{\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)(x_{1i}-\bar{x}_1)}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2} \tag{7}$$

which is exactly the final term in equation (6). Substituting equation (7) into equation (8) reduces the definition to

$$E[\tilde{\beta}_1]=\beta_1+\beta_2\hat{\delta}_1 \tag{8}$$

The bias in $E[\tilde{\beta}_1]$ generated by NOT including x_{2i} in the model is therefore a function of two thing: The covariance between x_{1i} and x_{2i} and the impact of x_{2i} on y

The following table summarizes the results

Direction in the bias for $E[\tilde{\beta}_1]$

	$Cov(x_1,x_2)>0$	$Cov(x_1,x_2)<0$
$\beta_2>0$	Positive	Negative
$\beta_2<0$	Negative	Positive

The Partialing out Properties of Multivariate Regression Models

For this example, we are going to examine a simple regression with only two covariates. The model is of the form

(1) $y_i = \beta_0 + x_{1i}\beta_1 + x_{2i}\beta_2 + \varepsilon_i$

Using the first-order conditions for this model, one can show, solving three equations and three unknowns that the estimate for β_1 is

(2) $$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)(y_i-\bar{y})\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2 - \sum_{i=1}^{n}(x_{1i}-\bar{x}_1)(x_{2i}-\bar{x})\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)(y_i-\bar{y})}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2 - \left(\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)(x_{2i}-\bar{x})\right)^2}$$

This is a complicated equation and you will NOT be asked to derive it. Note one thing. Suppose that in your sample, x_{1i} and x_{2i} are uncorrelated. This means that $\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)(x_{2i}-\bar{x}) = 0$ and equation (2) reduces to read

(3) $$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)(y_i-\bar{y})\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2} = \frac{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)(y_i-\bar{y})}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2}$$

which is the estimate we would obtain from a regression of y_i on x_{1i}. When covariates are uncorrelated, we do not need multivariate regression models, bivariate regressions will do.

But if the covariates are correlated, what variation in x_{1i} is used to produce the estimate for $\hat{\beta}_1$?

To answer this, consider a different regression. Regress x_{1i} on x_{2i} which is a model of the form

(4) $x_{1i} = \gamma_0 + x_{2i}\gamma_1 + \upsilon_i$

We know the OLS estimates of the parameters for this model will be

(5) $\hat{\gamma}_1 = \dfrac{\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)(x_{1i}-\bar{x}_1)}{\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2}$ and (6) $\hat{\gamma}_0 = \bar{x}_1 - \bar{x}_2\hat{\gamma}_1$

And construct the estimated error from this regression

(7) $\hat{\upsilon}_i = x_{1i} - \hat{\gamma}_0 - x_{2i}\hat{\gamma}_1$

Consider what $\hat{\upsilon}_i$ represents. Suppose x_{1i} and x_{2i} are correlated. This means that some of the value of x_{1i} is predictable by the value of x_{2i} and vice versa. The variable $\hat{\upsilon}_i$ measures the variation in x_{1i} that is NOT predictable by x_{2i}. This is the unique component of x_{1i} that the model will use when generating an estimate of $\hat{\beta}_1$. To see this, consider a third regression: suppose we were to regress the dependent variable y on the predicted error $\hat{\upsilon}_i$, which is a model of the form

(8) $y_i = \pi_0 + \hat{\upsilon}_i \pi_i + u_i$

Note that the estimate for $\hat{\pi}_1$ will be of the form

(9) $$\hat{\pi}_1 = \frac{\sum_{i=1}^{n}(\hat{\upsilon}_i - \bar{\hat{\upsilon}}_i)(y_i - \bar{y})}{\sum_{i=1}^{n}(\hat{\upsilon}_i - \bar{\hat{\upsilon}}_i)^2}$$

Note as well that $\hat{\upsilon}_i = x_{1i} - \hat{\gamma}_0 - x_{2i}\hat{\gamma}_1$ and $\bar{\hat{\upsilon}}_i = \bar{x}_1 - \hat{\gamma}_0 - \bar{x}_2\hat{\gamma}_1$ which means that $(\hat{\upsilon}_i - \bar{\hat{\upsilon}}_i) = (x_{1i} - \bar{x}_1) - (x_{2i} - \bar{x}_2)\hat{\gamma}_1$ where $\hat{\gamma}_1$ is defined above in equation (5). Substitute this value into equation (9) and we produce

(10) $$\hat{\pi}_1 = \frac{\sum_{i=1}^{n}(\hat{\upsilon}_i - \bar{\hat{\upsilon}}_i)(y_i - \bar{y})}{\sum_{i=1}^{n}(\hat{\upsilon}_i - \bar{\hat{\upsilon}}_i)^2} = \frac{\sum_{i=1}^{n}\left((x_{1i} - \bar{x}_1) - (x_{2i} - \bar{x}_2)\hat{\gamma}_1\right)(y_i - \bar{y})}{\sum_{i=1}^{n}\left((x_{1i} - \bar{x}_1) - (x_{2i} - \bar{x}_2)\hat{\gamma}_1\right)^2}$$

Working with the numerator, one can show that

(11) $$\sum_{i=1}^{n}\left((x_{1i} - \bar{x}_1) - (x_{2i} - \bar{x}_2)\hat{\gamma}_1\right)(y_i - \bar{y}) = \sum_{i=1}^{n}(x_{1i} - \bar{x}_1)(y_i - \bar{y}) - \hat{\gamma}_1\sum_{i=1}^{n}(x_{2i} - \bar{x}_2)(y_i - \bar{y})$$

Substitute the definition of $\hat{\gamma}_1$ from equation (5) into (11) and group like terms and you get

$$(12)\quad \sum_{i=1}^{n}(x_{1i}-\bar{x}_1)(y_i-\bar{y})-\frac{\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)(x_{1i}-\bar{x}_1)}{\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2}\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)(y_i-\bar{y})$$

$$=\frac{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)(y_i-\bar{y})\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)(x_{1i}-\bar{x}_1)\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)(y_i-\bar{y})}{\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2}$$

Notice this looks surprising like the numerator in equation (2). Now, work with the denominator in equation (10).

$$(13)\quad \sum_{i=1}^{n}\left((x_{1i}-\bar{x}_1)-(x_{2i}-\bar{x}_2)\hat{\gamma}_1\right)^2=\sum_{i=1}^{n}\left(x_{1i}-\bar{x}_1\right)^2+\hat{\gamma}_1^2\sum_{i=1}^{n}\left(x_{2i}-\bar{x}_2\right)^2-2\hat{\gamma}_1\sum_{i=1}^{n}\left(x_{1i}-\bar{x}_1\right)\left(x_{2i}-\bar{x}_2\right)$$

Substitute the definition of $\hat{\gamma}_1$ from equation (5) and group like terms and this term reduces to

$$(14)\quad \frac{\sum_{i=1}^{n}\left(x_{1i}-\bar{x}_1\right)^2\sum_{i=1}^{n}\left(x_{2i}-\bar{x}_2\right)^2-\left(\sum_{i=1}^{n}\left(x_{1i}-\bar{x}_1\right)\left(x_{2i}-\bar{x}_2\right)\right)^2}{\sum_{i=1}^{n}\left(x_{2i}-\bar{x}_2\right)^2}$$

Substituting (14) into the denominator and (12) into the numerator, the estimate for $\hat{\pi}_1$ now read

$$(15)\quad \hat{\pi}_1=\frac{\sum_{i=1}^{n}(\hat{\upsilon}_i-\bar{\hat{\upsilon}}_i)(y_i-\bar{y})}{\sum_{i=1}^{n}(\hat{\upsilon}_i-\bar{\hat{\upsilon}}_i)^2}=\frac{\dfrac{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)(y_i-\bar{y})\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)(x_{1i}-\bar{x}_1)\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)(y_i-\bar{y})}{\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2}}{\dfrac{\sum_{i=1}^{n}\left(x_{1i}-\bar{x}_1\right)^2\sum_{i=1}^{n}\left(x_{2i}-\bar{x}_2\right)^2-\left(\sum_{i=1}^{n}\left(x_{1i}-\bar{x}_1\right)\left(x_{2i}-\bar{x}_2\right)\right)^2}{\sum_{i=1}^{n}\left(x_{2i}-\bar{x}_2\right)^2}}$$

Notice that the $\sum_{i=1}^{n}\left(x_{2i}-\bar{x}_2\right)^2$ term in the numerator and denominator cancels out and (16) reduces to

$$(16) \quad \hat{\pi}_1 = \frac{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)(y_i-\bar{y})\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2 - \sum_{i=1}^{n}(x_{2i}-\bar{x}_2)(x_{1i}-\bar{x}_1)\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)(y_i-\bar{y})}{\sum_{i=1}^{n}\left(x_{1i}-\bar{x}_1\right)^2\sum_{i=1}^{n}\left(x_{2i}-\bar{x}_2\right)^2 - \left(\sum_{i=1}^{n}\left(x_{1i}-\bar{x}_1\right)\left(x_{2i}-\bar{x}_2\right)\right)^2}$$

The key result is that the $\hat{\beta}_1$ we obtain in a multivariate regression model is the same estimate we obtain from running a regression of *y* on the estimated residuals from equation (7). This result shows that the variation in x_{1i} used in the construction of $\hat{\beta}_1$ is only that variation that is not predictable by the other covariates in the regression. Therefore, the estimate for $\hat{\beta}_1$ is produce by holding "all else constant" – that is, the variation in the other variables in the model.

Summary of the Results for "Partialling out" Properties of Regressions

Consider a regression with two covariates: $y_i = \beta_0 + x_{1i}\beta_1 + x_{2i}\beta_2 + \varepsilon_i$

We know that the estimate for β_1 and β_2 will be a function of the covariance between x_1 and x_2

This can be seen most easily in the following example. Consider a synthetic regression of x_1 on x_2

$$x_{1i} = \theta_0 + x_{2i}\theta_1 + r_{1i}$$

Obtain the OLS estimates for $\hat{\theta}_0$ *and* $\hat{\theta}_1$ then construct the estimated residual

$$\hat{r}_{1i} = x_{1i} - \hat{\theta}_0 - x_{2i}\hat{\theta}_1$$

This residual is the portion of x_{1i} that is NOT explained by x_{2i}. Therefore, consider a regression of y_i on $\hat{r}_{1i}$ The "beta" in that regression would be of the form

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(\hat{r}_{1i}y_i)}{\sum_{i=1}^{n}\hat{r}_{1i}^2}$$

This is true because by construction residuals from OLS regressions always have zero mean so $\bar{\hat{r}}_1 = 0$

A Note about Variances in Multivariate Regression Models

Consider a basic multivariate model of the form

$$y_i = \beta_0 + x_{1i}\beta_1 + x_{2i}\beta_2 + x_{3i}\beta_3 + \ldots. x_{ki}\beta_k + \varepsilon_i$$

One can demonstrate that the variance of the estimate for $\hat{\beta}_k$ is of the form

$$V(\hat{\beta}_k) = \frac{\sigma_\varepsilon^2}{(1 - R_k^2)\sum_{i=1}^{n}(x_{ki} - \bar{x}_k)^2}$$

Where R_k^2 is the R^2 from a regression of x_k on all other exogenous variables.

$$x_{ki} = \gamma_0 + x_{1i}\gamma_1 + x_{2i}\gamma_2 + x_{3i}\gamma_3 + \ldots. x_{(k-1)i}\gamma_{k-1} + v_i$$

This result has a number of important implications. Suppose that the information contained in x_k is reflected in what is also in the model, that is, the other x's explain most of the variation in x_k. In that case, R_k^2 approaches 1, 1- R_k^2 approaches zero and $V(\hat{\beta}_k)$ explodes. The precision of an estimate is a function of how much independent variation there is in each x. If there is little 'new' information contained in the variable k, then we will have a difficult time learning anything new from having that variable in the model as the equation above illustrates.