Chapter 4 outline, Econometrics

In this section we will discuss the multiple regression model. We use the term multiple regression model to refer to a regression model where there is more than one independent variable.

1 The model

We can write the model as:

$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + ... + \beta_k X_{ki} + \varepsilon_i$, where

the βs represent the respective coefficients on each of the independent variables (which are the Xs), α is still the intercept, and ε is still the error term, with $\varepsilon\tilde{}N(0, \sigma_\varepsilon^2)$

The book will write the model as:

$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + ... + \beta_k X_{ki} + \varepsilon_i$, where

β_1 is the intercept and we have $k-1$ independent variables.

We have the same basic assumptions for this model as we had in chapter 3 for the two-variable model, except for one small change. The change is that no two independent variables can have an exact linear relationship.

1.1 Exact linear relationship

What is an exact linear relationship? Suppose that $X_1 = a + bX_2$ for all N observations of X_1 and X_2. This would be an exact linear relationship between X_1 and X_2, and we cannot have that if we wish to estimate our model. For example, suppose that X_2 was the variable AGE and I wished to create a variable called "$2AGE10$", where $2AGE10 = 2*AGE+10$ for every observation. In that case, I could NOT run a regression model with both AGE and $2AGE10$ as independent variables (actually, SAS might run this model but it may give you an error message and "throw out" one of your variables – we'll do an example in class).

It is also possible that you can have an exact linear relationship between multiple variables. We will see an example of this in chapter 5.

One example of a transformation that is NOT an exact linear relationship is:

$X_2 = (X_1)^2$

In chapter 5 we will discuss why we want to include squared terms in detail (usually because there are some nonlinearities in the relationship between X and Y). Students often wonder why we can include $(X_1)^2$ in a model with X_1 while we cannot include some term bX_1 in a model with X_1. Look at the tables below:

X_1	X_1	$(X_1)^2$
6	6	36
2	2	4
4	4	16
-5	-5	25
10	10	100

Table 1

b	X_1	bX_1
2	6	12
2	2	4
2	4	8
2	-5	-10
2	10	20

Table 2

We can use a model with $(X_1)^2$ and X_1 because we cannot find an equation in the form of $(X_1)^2 = a + bX_1$, where a and b are constants, for EVERY single observation of $(X_1)^2$ and X_1. We can find one for the first observation, $36 = a + b6$, if we let $a = -12$ and $b = 6$. However, this equation won't work for the 2^{nd} observation. We would have $4 = 2 * 6$, which is not true. So now we need an equation that will give us $36 = a + b6$ and $4 = a + b2$. We can also find a linear transformation for the first 2 data points if we let $a = 0$ and $b = 8$. However, try to find one for the 1^{st}, 2^{nd}, and 4^{th} data points. You can't, so we don't have an exact linear relationship. If we look at Table 2, we can see that $bX_1 = a + bX_1$, where $a = 0$ and $b = 2$ for EVERY single observation. Thus, there is an exact linear relationship between X_1 and bX_1, which means we cannot use both of them in our regression model.

1.2 Formulas for estimating the βs

In chapter 1 we were able to obtain formulas to estimate α and β. Suppose we have the following model:

$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$

We can obtain our estimates for β_1, β_2 and β_3 by minimizing the sum of squared deviations, just like we did in chapter 1. What we would want to do is:

$Minimize \sum_{i=1}^{N} (Y_i - predicted\ value\ of\ Y)^2$, where our predicted value of is given by the regression equation.

So we have:

$Minimize \sum_{i=1}^{N} (Y_i - (\beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i}))^2$

We know that to minimize means to take the derivative (partial derivative in this case) and set the derivative equal to zero. For this model we need 3 partial derivatives.

$\frac{\partial \sum_{i=1}^{N} (Y_i - (\beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i}))^2}{\partial \beta_1} = 0$

$\frac{\partial \sum_{i=1}^{N} (Y_i - (\beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i}))^2}{\partial \beta_2} = 0$

$\frac{\partial \sum_{i=1}^{N} (Y_i - (\beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i}))^2}{\partial \beta_3} = 0$

We will NOT do this differentiation. Once the differentiation is done, we will have 3 equations and 3 unknowns ($\hat{\beta}_1, \hat{\beta}_2$ and $\hat{\beta}_3$). We will NOT do the algebra to solve for the unknowns, but I will provide you with the formulas for $\hat{\beta}_1, \hat{\beta}_2$ and $\hat{\beta}_3$.

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3$$

$$\hat{\beta}_2 = \frac{\left(\sum\left(X_{2i}-\bar{X}_2\right)\left(Y_i-\bar{Y}\right)\right)\left(\sum\left(X_{3i}-\bar{X}_3\right)^2\right)-\left(\sum\left(X_{3i}-\bar{X}_3\right)\left(Y_i-\bar{Y}\right)\right)\left(\sum\left(X_{2i}-\bar{X}_2\right)\left(X_{3i}-\bar{X}_3\right)\right)}{\left(\sum\left(X_{2i}-\bar{X}_2\right)^2\right)\left(\sum\left(X_{3i}-\bar{X}_3\right)^2\right)-\left(\sum\left(X_{2i}-\bar{X}_2\right)\left(X_{3i}-\bar{X}_3\right)\right)^2}$$

$$\hat{\beta}_3 = \frac{\left(\sum\left(X_{3i}-\bar{X}_3\right)\left(Y_i-\bar{Y}\right)\right)\left(\sum\left(X_{2i}-\bar{X}_2\right)^2\right)-\left(\sum\left(X_{2i}-\bar{X}_2\right)\left(Y_i-\bar{Y}\right)\right)\left(\sum\left(X_{2i}-\bar{X}_2\right)\left(X_{3i}-\bar{X}_3\right)\right)}{\left(\sum\left(X_{2i}-\bar{X}_2\right)^2\right)\left(\sum\left(X_{3i}-\bar{X}_3\right)^2\right)-\left(\sum\left(X_{2i}-\bar{X}_2\right)\left(X_{3i}-\bar{X}_3\right)\right)^2}$$

(I'm using $\sum$ to mean $\sum\limits_{i=1}^{N}$ to cut down on the clutter).

Clearly, these formulas are a mess. We can rewrite the equations for $\hat{\beta}_2$ and $\hat{\beta}_3$ in a somewhat more manageable form as:

$$\hat{\beta}_2 = \frac{Cov(X_2,Y)Var(X_3)-Cov(X_3,Y)Cov(X_2,X_3)}{Var(X_2)Var(X_3)-(Cov(X_2,X_3))^2}$$

$$\hat{\beta}_3 = \frac{Cov(X_3,Y)Var(X_2)-Cov(X_2,Y)Cov(X_2,X_3)}{Var(X_2)Var(X_3)-(Cov(X_2,X_3))^2}$$

Still, these formulas are a bit useless. (Although if you look closely enough you can see what the units of $\hat{\beta}_3$ are. The numerator has units of $X_3 * Y * (X_2)^2$. The denominator has units of $(X_2)^2(X_3)^2$. Thus, our $\hat{\beta}_3$ has units of $\frac{X_3 * Y * (X_2)^2}{(X_2)^2(X_3)^2} = \frac{Y}{X_3}$, which are the units of a slope coefficient between X_3 and Y.) However, this is what the computer is calculating when it calculates the coefficients for a multiple regression model.

1.2.1 Interpreting these results

$\hat{\beta}_1$ still measures the intercept. Again, think about this as what Y would be equal to if ALL of the Xs equalled zero.

$\hat{\beta}_2$ measures the effect of X_2 on Y, holding X_3 constant. $\hat{\beta}_3$ measures the effect of X_3 on Y, holding X_2 constant. In the two variable model, we didn't hold anything else constant. However, now we are holding some other variable constant. Suppose our model was:

$$Wage_i = \beta_1 + \beta_2 Tenure_i + \beta_3 Age_i + \varepsilon_i$$

β_2 would tell us the effect of tenure on wages, holding age constant. This means how much higher (or lower) would your wage be if you had one more year of tenure GIVEN that you are a certain age. Suppose you are 26 years old and you have 4 years of tenure. The coefficient on tenure (β_2) would tell you how much higher (or lower) your wage would be IF you were 26 years old and had 5 years of tenure. Similarly, the coefficient on age (β_3) would tell you how much higher (or lower) your wage would be IF you were 27 years old with 4 years of tenure. It becomes a little more difficult to interpret the results once we move beyond the 2-variable model (remember the 2-variable model is the model with one independent variable). We can also have what appear to be counterintuitive results if we forget how to interpret the coefficients.

2 Regression statistics

When we obtain our estimates for $\hat{\beta}_1, \hat{\beta}_2$ and $\hat{\beta}_3$, we do not know how reliable the estimates are, so we need to perform some statistical tests. We can show that:

$\frac{\hat{\beta}_1 - \beta_1}{s_{\hat{\beta}_1}} \sim t_{N-k}$, $\frac{\hat{\beta}_2 - \beta_2}{s_{\hat{\beta}_2}} \sim t_{N-k}$, and $\frac{\hat{\beta}_3 - \beta_3}{s_{\hat{\beta}_3}} \sim t_{N-k}$,

where $\hat{\beta}_1, \hat{\beta}_2$ and $\hat{\beta}_3$ are our coefficient estimates; β_1, β_2 and β_3 are our null hypotheses; $s_{\hat{\beta}_1}, s_{\hat{\beta}_2}$ and $s_{\hat{\beta}_3}$ are the standard errors of $\hat{\beta}_1, \hat{\beta}_2$ and $\hat{\beta}_3$; and k is the number of independent variables, INCLUDING the constant (intercept) term. In this model, $Wage_i = \beta_1 + \beta_2 Tenure_i + \beta_3 Age_i + \varepsilon_i$, we would have $k = 3$. We then use the same format for hypothesis testing.

1. Set up the null and alternative hypotheses.
2. Construct your test statistic (remember to take the absolute value for a two-tailed test)
3. Pick a significance level
4. Look up the critical value in the table – remember how to count your degrees of freedom $(N - k)$
5. Reject or fail to reject the null hypothesis

We can use some shortcuts to testing hypotheses. Remember that the table for the t-distribution goes from 120 degrees of freedom to ∞ degrees of freedom. When we have ∞ degrees of freedom, we can use the critical values for the normal distribution, which are 1.96 and 2.57 for the 5% significance level and the 1% significance level respectively. If the absolute value of our test statistic is greater than 1.96, we can say the estimate is significant at the 5% level. If the absolute value of our test statistic is greater than 2.57, we can say the estimate is significant at the 1% level. While the shortcut method is useful for large data sets, we will need to know how to count degrees of freedom for small data sets and for some of the statistical tests we do later in the course.

3 R^2, corrected R^2 (or $\bar{R}^2$), and F-tests

We have seen how we measure goodness of fit of the model with R^2 in chapter 3. However, R^2 did not give us any more information than the simple t-statistic in the two-variable model. When we construct a multiple regression model, R^2 is a bit more useful. We can use R^2 to test if the regression is significant. With the multiple regression model, we can have a significant regression model with insignificant individual coefficients.

We have seen that $R^2 = \frac{RSS}{TSS}$, which is the explained variation (regression sum of squares) divided by the total variation (total sum of squares). Should our goal be to maximize R^2? While this seems like an appropriate goal, consider the following.

We draw a sample of the dependent variable, Y. The sample that we draw has a specific numerical value for its total variation (or total sum of squares). So let the total sum of squares of our sample of Y be TSS_Y. We know we can break TSS_Y into the portion of the variation explained by the model and the portion of the variation that is not explained by the model. Suppose $TSS_Y = 100$. We run a regression with one independent variable, X_1. We find that the RSS of the model is equal to 20 with just the X_1 variable. In this simple model, $R^2 = .2$. Now, suppose we want to add another independent variable to our model, X_2. Suppose that X_2 has very little to do with Y. The question is, will the new model $Y = \beta_1 + \beta_2 X_1 + \beta_3 X_2 + \varepsilon$, explain LESS than the old model, $Y = \beta_1 + \beta_2 X_1 + \varepsilon$? The answer is no. The new model will explain at least 20% of the variation in Y, due to the fact that X_1 is included in the model. If X_2 has ZERO effect on Y, then it will explain ZERO of the variation in Y, which means that we will still only be explaining 20% of the model. We can never explain less of the variation in Y by adding more variables. So if our goal was to maximize R^2, then we would use what is called the "kitchen sink approach". The kitchen sink approach means we throw in every single variable that we can find and this will maximize R^2 because adding additional regressors to the model can NEVER lower the amount of explained variation in the model.

Since R^2 can never fall when we add additional regressors, we need a different way to measure "goodness of fit" when we are using multiple regression models. The statistic that we will use is called corrected R^2 or $\bar{R}^2$. We define $\bar{R}^2$ as: $\bar{R}^2 = 1 - \frac{\hat{V}ar(\varepsilon)}{\hat{V}ar(Y)}$, where $\hat{V}ar(\varepsilon)$ is the estimated error variance and $\hat{V}ar(Y)$ is the estimated variance of Y. Notice that this looks very similar to our definition of R^2. Recall that $R^2 = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS}$. We can show how R^2 is related to $\bar{R}^2$.

We know $\hat{V}ar(\varepsilon) = \frac{\sum(\hat{\varepsilon}_i)^2}{N-k}$ and $\hat{V}ar(Y) = \frac{\sum\left(Y_i - \bar{Y}\right)^2}{N-1}$. So $ESS = \frac{\sum(\hat{\varepsilon}_i)^2}{N-k} * N - k$ and $TSS = \frac{\sum\left(Y_i - \bar{Y}\right)^2}{N-1} * N - 1$. If we plug these in, we get:

$R^2 = 1 - \frac{\hat{V}ar(\varepsilon)}{\hat{V}ar(Y)} \frac{N-k}{N-1}$. Now multiply both sides by $\frac{N-1}{N-k}$ to get:

$\frac{N-1}{N-k} R^2 = \frac{N-1}{N-k} - \frac{\hat{V}ar(\varepsilon)}{\hat{V}ar(Y)}$. Now, add 1 to both sides to get:

$1 + \frac{N-1}{N-k} R^2 = \frac{N-1}{N-k} + 1 - \frac{\hat{V}ar(\varepsilon)}{\hat{V}ar(Y)}$. Realizing that $1 - \frac{\hat{V}ar(\varepsilon)}{\hat{V}ar(Y)} = \bar{R}^2$, we get:

$1 + \frac{N-1}{N-k} R^2 = \frac{N-1}{N-k} + \bar{R}^2$. Isolate $\bar{R}^2$,

$1 + \frac{N-1}{N-k} R^2 - \frac{N-1}{N-k} = \bar{R}^2$. A little algebra gives us,

$1 + (R^2 - 1)\frac{N-1}{N-k} = \bar{R}^2$. This is equivalent to the equation in the book, which is: $1 - (1 - R^2)\frac{N-1}{N-k} = \bar{R}^2$. Both of these will be useful to show the 3 results below.

Looking at this equation gives us 3 results:

1. If $k = 1$, $R^2 = \bar{R}^2$ (this is easy to see)
2. If $k > 1$, $R^2 \geq \bar{R}^2$ (this is a little more difficult to see)
3. $\bar{R}^2$ can be negative (also a little more difficult to see)

Start with result number 3, and use $1 + (R^2 - 1)\frac{N-1}{N-k} = \bar{R}^2$. We know that $R^2 < 1$, so $(R^2 - 1) < 0$. This fact alone will not cause $\bar{R}^2$ to be negative unless $\frac{N-1}{N-k}$ is "very large". And when will $\frac{N-1}{N-k}$ be very large? When k is very large. As an example, suppose $N = 51$ and $k = 50$. Also, suppose $R^2 = .9$, which looks like a good fit. What will $\bar{R}^2$ be? $1 + (.9 - 1)\frac{50}{1} = -4.0$

As for result 2, use $1 - (1 - R^2)\frac{N-1}{N-k} = \bar{R}^2$. The important fact to realize is that we are subtracting a number greater than $(1 - R^2)$ from 1. If we were only subtracting $(1 - R^2)$ from 1, then we would get R^2. However, since we are subtracting a number greater than $(1 - R^2)$ from 1, we get a number smaller than R^2.

3.1 F-tests

We would like to know if the R^2 we receive is statistically significant. We can perform a statistical test to answer this question. Formally, we are testing:

$H_0 : \beta_2 = \beta_3 = \beta_4 = ... = \beta_k = 0$

H_A : At least one $\beta \neq 0$

Note that we do not include the intercept in our null hypothesis, meaning that we are testing to see if $k - 1$ coefficients are equal to zero. What we wish to test is that all the regression coefficients are JOINTLY equal to zero. This is different than looking at each parameter estimate and seeing if it is (individually) different than zero, so we need a different statistical test than the t-test.

The statistical test that we use is an F-test. We calculate our F-statistic as:

$\frac{\frac{RSS}{k-1}}{\frac{ESS}{N-k}} \tilde{} F_{k-1,N-k}$

Alternatively, we could write:

$\frac{\frac{R^2}{k-1}}{\frac{1-R^2}{N-k}} \tilde{} F_{k-1,N-k}$

You should convince yourself that you will obtain the same F-statistic regardless of which formula you use to calculate it.

Why do we use the F-distribution? We can show that our F-statistic is the ratio of 2 independent χ^2 random variables to their respective degrees of freedom.

To finish the test you just need to look up the critical value in the table in the back of the book. If your F-statistic is greater than the critical value then you reject the null hypothesis. As for choosing a significance level, realize that there are only tables in the back of the book for the 1% and 5% levels, so those are the only 2 significance levels you can test at unless you want to find tables for the other significance levels. Once again, if the F-statistic that you calculated was greater than the critical value, you reject the null hypothesis and conclude that at least one β is significant at the chosen significance level.

3.1.1 Final note on R^2

We can only use R^2 to compare models that have the exact same independent variable. That is, suppose we had 2 models, where one model used Y as the independent variable and the other model used $\ln Y$. Although it seems like we are using the same variable (after all, $\ln Y$ is just a transformation of Y), the models will have different total sums of squares and will NOT be comparable.

4 Multicollinearity

Multicollinearity MAY occur when we have 2 or more independent variables that are highly correlated. Recall when we interpret regression coefficients we say that we hold everything else constant. However, if 2 independent variables are highly correlated this may not be the case.

4.1 Perfect collinearity

Perfect collinearity occurs when there is an EXACT linear relationship between one or more of the independent variables. When an exact linear relationship occurs, we CANNOT obtain least squares estimates of the regression coefficients. The "real reason" we cannot obtain least squares estimates has to do with matrix algebra. Since matrix algebra is not a prerequisite for this course, I will not go into the details. Perfect collinearity rarely arises in data that is naturally occurring, and is most often researcher induced. We will examine perfect collinearity more closely when we work through chapter 5.

4.2 Effects of multicollinearity

A more realistic problem occurs when the variables are highly correlated, but a perfect linear relationship does not exist. In this case we can obtain least squares estimates for our regression coefficients. However, it may be difficult to interpret the estimates because we are supposed to be holding all else constant.

Another "problem" that occurs is the variances (and standard errors) of the β's become large when 2 independent variables are highly correlated. Consider the following model:

$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$

It can be shown that:

$Var(\hat{\beta}_2) = \frac{(N-1)s^2}{Var(X_2)(1-(corr(X_2,X_3)^2))}$ and

$Var(\hat{\beta}_3) = \frac{(N-1)s^2}{Var(X_3)(1-(corr(X_2,X_3)^2))}$

$corr(X_2, X_3)$ is the sample correlation between X_2 and X_3. What happens as the correlation gets close to 1 (perfect positive correlation) or -1 (perfect negative correlation)? The term $(1-(corr(X_2,X_3)^2))$ becomes very close to zero, meaning that the variances of our estimated coefficients go to infinity. If the variances go to infinity, and the standard errors are the square root of something that goes to infinity, then the standard errors are also becoming very large. Why is this a problem for us? Recall that when we constructed our t-statistic for the $\hat{\beta}$'s we used: $\left|\frac{\hat{\beta}-\beta}{s_{\hat{\beta}}}\right|$. The problem now is that our t-statistics will be very small since we are dividing by a number, $s_{\hat{\beta}}$, that is very large. If we have small t-statistics this means that we will often fail to reject the null hypothesis.

4.3 Indications of multicollinearity

Multicollinearity is a difficult problem to detect (at least in a statistical manner). The best method to detecting multicollinearity is to look at the t-statistics of the individual regression coefficients and the F-statistic for the R^2. If the regression is significant but the individual coefficients are not, then this suggests that there is multicollinearity in the data. One of the remedies is to remove one of the variables and see what happens to your R^2, your t-statistics, and your estimated coefficients. If the R^2 and the estimated coefficients "do not change very much" while the t-statistics increase, then this is suggestive that you have corrected for your multicollinearity problem.

There are other methods that one might use to test for multicollinearity – unfortunately, none of them are highly accepted.

5 A few more useful items

5.1 Standardized coefficients

Suppose we run the following regression:

$price = \beta_1 + \beta_2 floorspace + \beta_3 bedrooms + \varepsilon$

We obtain estimates for β_1, β_2 and β_3. It is easy to interpret the coefficients but it is difficult to tell which coefficient has the greatest impact on the price because the coefficients are in different units (the denominator of β_2 is square feet and the denominator of β_3 is number of rooms). In order to determine which variable has the biggest impact on the dependent variable we can use standardized coefficients. If we have the original model, $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + ... + \beta_k X_k + \varepsilon$, we can create the following standardized model: $\frac{Y-\bar{Y}}{s_Y} = \beta_2^* \frac{X_2-\bar{X}_2}{s_{X_2}} + \beta_3^* \frac{X_3-\bar{X}_3}{s_{X_3}} + ... + \beta_k^* \frac{X_k-\bar{X}_k}{s_{X_k}} + \varepsilon$. The coefficients will now tell us how much Y will increase when we have a 1 standard deviation increase in a particular X. We can now tell which variables have the biggest impact on Y.

5.2 Elasticities

Recall that in other economics courses we developed this concept of elasticity. We typically discussed the price elasticity of demand or income elasticity. The price elasticity of demand told us how much quantity demanded changed when we had a change in the price of the good. One "problem" with elasticities is that they depend on where you are along the line. Recall that with price elasticity of demand if we were near the top of the demand curve then we would likely have elastic demand (large percentage changes in the quantity demanded relative to the percentage change in price) and if we were near the bottom of the demand curve we would likely have inelastic demand (small percentage changes in the quantity demanded relative to the percentage change in price). So we need to pick a price and quantity about which to measure the elasticity. We had the formula $\frac{\%\Delta Q_d}{\%\Delta P} = \frac{\frac{Q_1 - Q_0}{Q_0}}{\frac{P_1 - P_0}{P_0}}$ (actually if you used the Gwartney textbook in Principles of Micro here at FSU you likely had the following formula: $\frac{\frac{Q_1 - Q_0}{\left(\frac{Q_0 + Q_1}{2}\right)}}{\frac{P_1 - P_0}{\left(\frac{P_0 + P_1}{2}\right)}}$. The actual formula is not important, just thought I would jog your memory.)

Concentrating on, $\frac{\frac{Q_1 - Q_0}{Q_0}}{\frac{P_1 - P_0}{P_0}}$, let's rearrange a few terms. Doing some algebra gives us, $\frac{Q_1 - Q_0}{P_1 - P_0} * \frac{P_0}{Q_0}$. Now, look at the regression equation for simple linear demand, $quantity = \beta_1 + \beta_2 price + \varepsilon$. If we take the partial derivative of this equation with respect to price, we get: $\frac{\partial quantity}{\partial price} = \beta_2$. Now, what does a derivative tell us? How much Y (in this case quantity) changes when we change X (in this case price). Notice that this is similar to $\frac{Q_1 - Q_0}{P_1 - P_0}$, where $(Q_1 - Q_0)$ is the change in Y (quantity) and $(P_1 - P_0)$ is the change in X (price). So the coefficient is ONE part of the elasticity. However, elasticities are PERCENTAGE changes, and β_2 does not measure a percentage change (at least not in this model). Thus, we need to multiply our coefficient by some $\frac{P}{Q}$ to have obtain the elasticity. Now the choice of which P and which Q is arbitrary, but remember that the point that you wish to evaluate the elasticity at will influence the elasticity. By convention we evaluate the elasticity about the means of the variables. This is how "they" (authors of textbooks – well, the authors of textbooks probably take the estimates from authors of research papers and consulting work) determine that the price elasticity of demand for salt is -0.1, for movies is -0.9, for fresh green peas is -2.8.