Simple Regression Model[1]

When we say that as the price of a good increases quantity demanded falls, or that an increases in Foreign Direct Investment into an economy, growth rate of GDP increases etc, what we are claiming is a that one factor/variable is affected by changes in another. More generally, we are trying ask the question how would a change in a variable y (**dependent variable/regressand**) be affected by changes in another variable x (**independent variable/regressor/covariate/explanatory variable**). y is a predetermined variable we are trying to study, and we are trying to use changes in x to explain changes by y, consequently x is a stochastic or random variable.

In finding out this relationship between the two variables we have to recognize three issues;

1. Since there is never a exact relationship, how do we for other factors to affect y.
2. What is form of the relationship; that is the form of the equation that would describe their relationship?
3. How can we be sure we are capturing a ceteris paribus relationship.

The simplest and most often used functional form is the simple linear regression model;

$$y = \beta_0 + \beta_1 x + \varepsilon$$

where ε (reads as epsilon) is just all other factors that may affect y, and you can think of this factor as capturing unobserved factors. This is also call the population model, or what we believe we would see should we examine the entire population of observations. If we hold ε constant then we would have effectively found the effect x has on y, that is $\Delta y = \beta_1 \Delta x$ if $\Delta \varepsilon = 0$, where Δ stand for change (reads as delta). Since the functional form is linear, β_0 is just a constant term. The principal measure of the relationship is β_1 and it is the slope parameter. It tells us that for a unit change in x, the level of change in y. Note that this is the purported relationship in the population, so that ε is the population idiosyncratic term.

Monotonic Transformations of Economic Relationship

Matter of fact, the functional form of the relationship between x and y need not strictly be in a linear form. As long as the function can be transformed into a linear function (a monotonic transformation that is one that preserves the relationship.) will suffice. I list some of the most common below;

$$y = \beta_0 x^{\beta_1} \Rightarrow \log y = \log \beta_0 + \beta_1 \log x$$
$$y = \exp(\beta_0 + \beta_1 x) \Rightarrow \log y = \beta_0 + \beta_1 x$$
$$\exp y = \beta_0 x^{\beta_1} \Rightarrow y = \log \beta_0 + \beta_1 \log x$$

However, note that the interpretation after the transformation is not the same. For example in the first equation, y is linear in log x not x. In the second equation, log y is

[1] **Read chapter 2 and the appendix C of your text.**

linear in x but y is not linear in x but exp x. We will discuss shortly what each of this various forms mean for us in greater detail.

In the equation such as the one below, y is not linear in x but the reciprocal of x.

$$y = \beta_0 + \beta_1 \frac{1}{x}$$

The regression equation to be used then is dependent on the economics underlying the relationship between the regressand and the regressor. Can you think of other simple economic relationships that could be transformed into a linear functional form.

Assumptions to Obtained Parameter Estimates

Once we have a functional form of the relationship, we are faced with the problem of obtaining an estimate for the parameters that would tell us how the regressor affects the regressand, and the magnitude of the effect. As noted above, we need a term ε that captures all other variables we cannot observe but exerts an effect on y. If there were no such unobservables then to find the parameter values, which in the simplest case is the first linear equation, all we would are two points to obtain the intercept and the slope of the economic relationship, but that possibly be reality. We then need some number, n, of observations to fit our model.

Consider a simple example for individual choice of the number of years of education. Part of such a decision is due to parental income. However, to say that parental income fully determines a child's choice of years of education ignores aspects such as the child's ability, the decision of her peers, etc. These variables/arguments are not observed by the researcher and are assumed to be captured by the catch all ε. Because ε is largely thought of as capturing un-observables or idiosyncratic elements to each observation, we typically think of it as capturing a randomness of the observation under study, which in the example, is the seemingly random choice in years of education once we account of some of the observed factors (For now, we will assume there is only one factor, but will generalize it in our subsequent discussions. This is because there are other factors that are measurable such as her opportunity cost should she enter the labor force, the availability of temporary job that would cover the difference between the schools fees and the ability of her parents to pay those fees, and the number of sibling of school age.). We hence call ε a disturbance or error term (not that we think the observations choice of the dependent variable is an error, but something not explained by our model.).

This random error/disturbance term can push the observation's decision one way or the other. However, because it is a factor that is thought of as idiosyncratic and unable to explain the changes in y or years of education in our example, we think of the error as having no effect on the average. This then gives us our first assumption;

Assumption 1: $E(\varepsilon) = 0$
This assumption says that on the average, the idiosyncratic factors will not affect changes in *y* systematically.

However, this assumption alone is insufficient for us to study how *y* varies with changes in *x*. Consider this, what if x is affected by ε? If true we would not be able to discern between whether the effect on *y* is coming from *x*, or from ε through *x*, and which part of the effect is purely from *x* alone. That is we need the two, the idiosyncratic element and the regressor or explanatory variable to be uncorrelated, that does not have any relationship nor effect on each other. We include this by making the following assumption;

Assumption 2: $E(\varepsilon \mid x) = E(\varepsilon) = 0$

Translating this to our example, what this says is that given the educational level (or years of education) of an individual ability, the average ability are the same.

To be precise, we assume that the error/disturbance/idiosyncratic term to be unimodal, and invoking the central limit theorem, assume that they are normally distributed with variance σ_ε^2. Further, we assume each of the disturbance terms are independently and identically distributed (i.i.d.). Together they can succinctly express as follows:

Assumption 3: $\varepsilon \sim N(0, \sigma_\varepsilon^2)$.

The significance of these assumptions is seem when we take the conditional expectation of *y* given *x*. Using the simplest linear regression model,

$$E(y \mid x) = \beta_0 + \beta_1 x + E(\varepsilon \mid x) = \beta_0 + \beta_1 x$$

This is the population regression function.

Assumption 4: Homogeneity.

Embedded in assumption 3 is something that might seem subtle, that we have assumed explicitly the each of the idiosyncratic error terms have in fact the same variance. This assumption is strong, and may be relaxed, and we will do so subsequently. In the first place, do you even thing this is reasonable intuitively.

The Ordinary Least Squares (OLS) Estimates

We can estimate the relationship of variables by taking a sample of observations from the population to draw the relationship. Let $\{(x_i, y_i) : i = 1,2,...,n\}$ denote a random sample of

size *n* from the population of possible observations, and $\hat{\beta}_0, \hat{\beta}_1$ are the estimated parameters.

We can write the relationship of *x* and *y* for each observation as follows:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i$$

Let's think about a way we can solve for the relationship underlying the population using the sample. First suppose the sample numbers we need is sufficient for us to do so (we will be a little bit more precise shortly about how we can gauge if our sample size is sufficient. That part is part art and part science). Now if our model is correct, the error or disturbance term should not figure into the description of the value that *y* takes, and *x* will be sufficient a predictor of *y*. Let's write this predicted part as follows;

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Then what we want is to find parameter values of β_0 and β_1 that minimizes the difference between reality and our model, that is

$e_i = y_i - \hat{y}_i, \forall i \in \{1,2,...,n\}$

So it would seem we can minimize the sum of these error terms, and that would be it. That is;

$$\min_{\hat{\beta}_0,\hat{\beta}_1} \sum_{i=1}^{n} e_i = \min_{\hat{\beta}_0,\hat{\beta}_1} \sum_{i=1}^{n} y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$

Or is it? Consider this the values of the error term can either be positive or negative. So the objective would be minimized when the sum of error terms is zero.

$$\sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right) = 0$$

$$\Rightarrow \sum_{i=1}^{n} \left(y_i \right) = \sum_{i=1}^{n} \left(\hat{\beta}_0 + \hat{\beta}_1 x_i \right)$$

$$\Rightarrow \bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$$

This objective then merely requires us to find two parameter values than fits the last equality. Given that we have two parameters, and one equation, this means that there are several estimates, to be precise an infinite number of possible parameter values could fit that last equation. Another way to think about it is that the objective is strictly decreasing in the parameter values, and we cannot possibly find a pair of parameter values for β_0, β_1 here that could do it.

However, this can be easily solved by making this objective function convex in the parameters (since we are minimizing the sum of errors). That is

$$\min_{\hat{\beta}_0,\hat{\beta}_1} \sum_{i=1}^{n} \left(e_i \right)^2 = \min_{\hat{\beta}_0,\hat{\beta}_1} \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right)^2$$

It is easy to see then that the parameter estimates are:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\text{cov}(x, y)}{\text{var}(x)} = \frac{\sigma_{xy}^2}{\sigma_x^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Can you solve for it?

Decomposition of Sum of Squares

But how do we know if our model fit the data well enough? That is we need a measure to tell us how "true" in a statistical sense is our model. Well we can examine how far does y varies from its mean, and see how well our econometric (here our simple regression model) model measures up by checking how much of that variation was our model able to capture. To see this, note that the total sum of squares is

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}\left(\hat{\beta}_1(x_i-\bar{x})+\varepsilon_i\right)^2=\sum_{i=1}^{n}\left(\hat{\beta}_1(x_i-\bar{x})^2+\hat{\beta}_1(x_i-\bar{x})\varepsilon_i+\varepsilon_i^2\right)$$

$$=\left(\hat{\beta}_1\right)^2\sum_{i=1}^{n}\left((x_i-\bar{x})^2\right)+\hat{\beta}_1\sum_{i=1}^{n}\left((x_i-\bar{x})\varepsilon_i\right)+\sum_{i=1}^{n}\left(\varepsilon_i^2\right)$$

$$=\left(\hat{\beta}_1\right)^2\sum_{i=1}^{n}(x_i-\bar{x})^2+\sum_{i=1}^{n}\left(\varepsilon_i^2\right)$$

$$\Rightarrow \text{Total Sum of Squares (TSS)}=\begin{pmatrix}\text{Explained Sum of Squares (ESS)}\\+\text{ Residual Sum of Squares (RSS)}\end{pmatrix}$$

This is because $\hat{\beta}_1\sum_{i=1}^{n}\left((x_i-\bar{x})\varepsilon_i\right)=0$, which can derive from assumption 2 since assumption two essentially says that the error term is independent of the regressor.

$$E(e\mid x)=E(e)=0$$

$$\operatorname{cov}(x,e)=\frac{\sum_{i=1}^{n}\left((x_i-\bar{x})\varepsilon_i\right)}{n-1}=0$$

From the above, we can see that part of the regression movement or changes in *y* is explained by *x*, and we call it the explained sum of squares of the regressor (ESS or SSE), while the other part is the residual sum of squares (RSS or SSR), or that part of the regression model that is not explained. It is using RSS and ESS that we can define some form of goodness of fit to gauge how good our model is. What would you think is a good measure? Well, if think it should be the ratio between ESS and TSS, you would be right!

$$R^2=\frac{ESS}{TSS}=1-\frac{RSS}{TSS}$$

Expected Values and Variances of OLS estimators

We will now study some of the statistical properties of our estimator for the population parameters $\hat{\beta}_0, \hat{\beta}_1$. What this means is that we would like to understand how the estimators would behave as we keep redrawing samples from the population to run our model.

The OLS Estimators are Unbiased

There are three essential features to the sampling distribution which are the mean, variance, and mean square variance.

First the mean is nothing but

$$E(\hat{\beta}_i) = \int \hat{\beta} f(\hat{\beta}_i) d\hat{\beta}_i$$

Now let us define what the bias is; suppose we are able to resample from the population as many times we want, we can then obtain an average of all the estimators, $bias(\hat{\beta}_i) = E(\hat{\beta}_i) - \beta_i$. If this bias is zero, we can then say that our estimator is unbiased, otherwise it is bias.

Second the variance of the distribution

$$\text{var}(\hat{\beta}_i) = E\left[\hat{\beta}_i - E(\hat{\beta}_i)\right]^2 = \int \left(\hat{\beta}_i - E(\hat{\beta}_i)\right)^2 f(\hat{\beta}_i) d\hat{\beta}_i = \sigma^2_{\hat{\beta}_i}$$

which measures the spread of the distribution about the mean. Now would you prefer an estimator that has a small variance or a large one? Well, if you think you would prefer a small one, you would be right. Of all possible estimators, the one which yields the smallest variance is to be the *Best Unbiased Estimator*.

However, it is possible for an estimator to have a low bias but a higher variance, while it is also a possibility to have a large bias but small variance. How would you make your optimal choice? What would be the optimal tradeoff so that you could distinguish between two estimators, so that you can pick one? We can measure this by calculating the mean squared error (*MSE*) , which is

$$\begin{aligned}
MSE(\hat{\beta}_i) &= E(\hat{\beta}_i - \beta_i)^2 = E\left(\left[\hat{\beta}_i - E(\hat{\beta}_i)\right] + \left[E(\hat{\beta}_i) - \beta_i\right]\right)^2 \\
&= E\left[\hat{\beta}_i - E(\hat{\beta}_i)\right]^2 + 2E\left[\hat{\beta}_i - E(\hat{\beta}_i)\right]\left[E(\hat{\beta}_i) - \beta_i\right] + E\left[E(\hat{\beta}_i) - \beta_i\right]^2 \\
&= E\left[\hat{\beta}_i - E(\hat{\beta}_i)\right]^2 + 2\left[E(\hat{\beta}_i) - E(\hat{\beta}_i)\right]\left[E(\hat{\beta}_i) - \beta_i\right] + \left[E(\hat{\beta}_i) - \beta_i\right]^2 \\
&= \sigma^2_{\hat{\beta}_i} + (bias)^2
\end{aligned}$$

We can bring the expectation operator into the cross product because $\left[E(\hat{\beta}_i) - \beta_i\right]$ is a constant.

We can now examine to derive the means, and variances of our estimator to see if our estimator is a *Best Linear Unbiased Estimator* (B.L.U.E.).

We know from before that

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

where $w_i = \frac{(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$. Further note that the following are true,

$$\sum_{i=1}^{n} w_i = \sum_{i=1}^{n} \frac{(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{1}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \sum_{i=1}^{n}(x_i - \bar{x})$$

$$= \frac{1}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \left\{ \left(\sum_{i=1}^{n} x_i \right) - n\bar{x} \right\} = \frac{1}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \left\{ \left(\sum_{i=1}^{n} x_i \right) - n \frac{\left(\sum_{i=1}^{n} x_i \right)}{n} \right\} = 0$$

and

$$\sum_{i=1}^{n} w_i x_i = \frac{\sum_{i=1}^{n}(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} x_i = \frac{\sum_{i=1}^{n}\left((x_i)^2 - \bar{x}x_i\right)}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}\left((x_i)^2 - \bar{x}x_i\right)}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$= \frac{\sum_{i=1}^{n}(x_i)^2 - n(\bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = 1$$

and,

$$\sum_{i=1}^{n} \left(\frac{(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \right)^2 = \frac{1}{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2 \right]^2} \sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Now note that,

$$\hat{\beta}_1 = \sum_{i=1}^{n} w_i y_i = \sum_{i=1}^{n} w_i (\beta_0 + \beta_1 x_i + \varepsilon_i)$$

$$= \beta_1 \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} w_i (\varepsilon_i) = \beta_1 + \sum_{i=1}^{n} w_i (\varepsilon_i) = \beta_1 + bias$$

since $\sum_{i=1}^{n} w_i \beta_0 = 0$. Taking expectations of the above, you would then get

$$E\left(\hat{\beta}_1\right) = \beta_1 + \sum_{i=1}^{n} w_i \left(\varepsilon_i\right) = \beta_1$$

since $E\left(\varepsilon_i\right) = 0$. This means the OLS slope estimator is unbiased. How about the constant term then?

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \beta_0 + \left(\beta_1 - \hat{\beta}_1\right)\bar{x} + \bar{\varepsilon}$$
$$E\left(\hat{\beta}_0\right) = \beta_0 + \left(\beta_1 - E\left(\hat{\beta}_1\right)\right)\bar{x} + E\left(\bar{\varepsilon}\right) = \beta_0$$

Which means that the constant term is also unbiased. We have to check the variance of the parameter estimates still.

$$\text{var}\left(\hat{\beta}_1\right) = E\left(\hat{\beta}_1 - \beta_1\right)^2 = \left(E\left(\hat{\beta}_1\right)^2 - 2\beta_1 E\left(\hat{\beta}_1\right) - \left(\beta_1\right)^2\right)$$

$$= \left(E\left(\hat{\beta}_1\right)^2 - \left(\beta_1\right)^2\right) = \left(E\left(\beta_1 + \sum_{i=1}^{n} w_i \varepsilon_i\right)^2 - \left(\beta_1\right)^2\right)$$

$$= E\left(\left(\beta_1\right)^2 + 2\beta_1 \sum_{i=1}^{n} w_i \varepsilon_i + \left(\sum_{i=1}^{n} w_i \varepsilon_i\right)^2\right) - \left(\beta_1\right)^2$$

$$= 2\beta_1 E\left(\sum_{i=1}^{n} w_i \varepsilon_i\right) + E\left(\left(\sum_{i=1}^{n} w_i \varepsilon_i\right)^2\right)$$

$$= \frac{\sigma_\varepsilon^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$$

and for the constant term,

$$\text{var}\left(\hat{\beta}_0\right) = E\left(\hat{\beta}_0 - \beta_0\right)^2 = E\left(\bar{y} - \hat{\beta}_1 \bar{x} - \beta_0\right)^2$$
$$= E\left(\left(\beta_1 - \hat{\beta}_1\right)\bar{x} + \bar{\varepsilon}\right)^2 = \bar{x}^2 E\left(\beta_1 - \hat{\beta}_1\right)^2 + 2\bar{x}E\left(\left(\beta_1 - \hat{\beta}_1\right)\bar{\varepsilon}\right) + E\left(\bar{\varepsilon}\right)^2$$

$$= \frac{\bar{x}^2 \sigma_\varepsilon^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} + \frac{\sigma_\varepsilon^2}{n} = \sigma_\varepsilon^2 \left(\frac{\bar{x}^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} + \frac{1}{n}\right)$$

The last equality is because,

$$E\left(\left(\beta_1 - \hat{\beta}_1\right)\bar{\varepsilon}\right) = \beta_1 E\left(\bar{\varepsilon}\right) - E\left(\hat{\beta}_1 \bar{\varepsilon}\right) = -E\left(\beta_1 + \sum_{i=1}^{n} w_i \varepsilon_i\right)\bar{\varepsilon}$$

$$= -\left(\sum_{i=1}^{n} w_i \varepsilon_i\right)\frac{\sum_{i=1}^{n} \varepsilon_i}{n} = -\frac{1}{n}\left(\left(\sum_{i=1}^{n} w_i \left(\varepsilon_i\right)^2\right) + \sum_{i \neq j}\left(w_i + w_j\right)\varepsilon_i \varepsilon_j\right) = 0$$

Note that the above variances would be the variance of the parameters. We can then write the distribution of the parameters as

$$\hat{\beta}_0 \sim N\left(\beta_0, \sigma_\varepsilon^2\left\{\frac{1}{n}+\frac{\bar{x}}{\sum(x_i-\bar{x})}\right\}\right)$$
$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma_\varepsilon^2}{\sum(x_i-\bar{x})}\right)$$

Detour

However we need to augment the formula because we need to estimate σ_e^2 if it is unknown. We can do so by calculating $\hat{\sigma}_e^2 = s^2 = \frac{\sum_{i=1}^{n} e_i}{n-2}$. We typically call the estimate of the variance for the idiosyncratic error term as s^2. Why?

The answer is because,

$$e_i = y_i - \hat{y}_i = \beta_0 + \beta_1 x_i + \varepsilon_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$
$$= \left(\beta_0 - \hat{\beta}_0\right) + \left(\beta_1 - \hat{\beta}_1\right)x_i + \varepsilon_i$$
$$= \left(\varepsilon_i - \bar{\varepsilon}\right) + \left(\beta_1 - \hat{\beta}_1\right)\left(x_i - \bar{x}\right)$$
$$\therefore \sum(e_i)^2 = \sum(\varepsilon_i - \bar{\varepsilon})^2 + 2\left(\beta_1 - \hat{\beta}_1\right)\sum(\varepsilon_i - \bar{\varepsilon})(x_i - \bar{x}) + \left(\beta_1 - \hat{\beta}_1\right)^2 \sum(x_i - \bar{x})^2$$
$$= \sum(\varepsilon_i)^2 - n\bar{\varepsilon}^2 + \left(\beta_1 - \hat{\beta}_1\right)^2 \sum(x_i - \bar{x})^2 + 2\left(\beta_1 - \hat{\beta}_1\right)\left\{\sum \varepsilon_i x_i - n\bar{\varepsilon}\bar{x}\right\}$$
$$= \sum(\varepsilon_i)^2 - n\bar{\varepsilon}^2 + \left(\beta_1 - \hat{\beta}_1\right)^2 \sum(x_i - \bar{x})^2 + 2\left(\beta_1 - \hat{\beta}_1\right)\left\{\sum \varepsilon_i (x_i - \bar{x})\right\}$$
$$\therefore E\left(\sum(e_i)^2\right) = n\sigma_\varepsilon^2 - n\frac{\sigma_\varepsilon^2}{n} + \frac{\sigma_\varepsilon^2}{\sum(x_i-\bar{x})^2}\sum(x_i-\bar{x})^2 - 2E\left(\sum w_i\varepsilon_i\right)\left\{\sum \varepsilon_i(x_i-\bar{x})\right\}$$
$$= n\sigma_\varepsilon^2 - 2\sigma_\varepsilon^2 = (n-2)\sigma_\varepsilon^2$$
$$\Rightarrow \sigma_\varepsilon^2 = E\left(\frac{\sum(e_i)^2}{n-2}\right)$$

Which shows that our suggested estimator is unbiased. **Can you think of a better way to show it?**

We are however not done with showing that our estimator is B.L.U.E. To finalize the proof, note that

$$\hat{\beta}_1 = \sum_{i=1}^{n} w_i y_i$$

which means that the estimator for the slope parameter is a linear combination of the *y*'s. Let there be another similar estimator,

$$b_1 = \sum_{i=1}^{n} c_i y_i = \sum_{i=1}^{n} c_i \left(\beta_0 + \beta_1 x_i + \varepsilon_i\right)$$

$$\therefore E\left(b_1\right) = \beta_0 \sum_{i=1}^{n} c_i + \beta_1 \sum_{i=1}^{n} c_i x_i$$

For b_1 to be unbiased we need the following to be true,

$$\sum_{i=1}^{n} c_i = 0, \sum_{i=1}^{n} c_i x_i = \sum_{i=1}^{n} c_i \left(x_i - \bar{x}\right) = 1$$

However with these conditions,

$$\text{var}\left(b_1\right) = E\left(b_1 - \beta_1\right)^2 = \sigma_\varepsilon^2 \sum_{i=1}^{n} \left(c_i\right)^2$$

Next note that

$$c_i = w_i + \left(c_i - w_i\right)$$

$$\therefore \sum_{i=1}^{n} \left(c_i\right)^2 = \sum_{i=1}^{n} \left(w_i\right)^2 + 2\sum_{i=1}^{n} w_i \left(c_i - w_i\right) + \sum_{i=1}^{n} \left(c_i - w_i\right)^2$$

But,

$$\sum_{i=1}^{n} w_i \left(c_i - w_i\right) = \sum_{i=1}^{n} w_i c_i - \sum_{i=1}^{n} \left(w_i\right)^2 = \frac{1}{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2} - \frac{1}{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2} = 0$$

Therefore,

$$\text{var}\left(b_1\right) = \sigma_\varepsilon^2 \sum_{i=1}^{n} \left(c_i\right)^2 = \sigma_\varepsilon^2 \sum_{i=1}^{n} \left(w_i\right)^2 + \sigma_\varepsilon^2 \sum_{i=1}^{n} \left(c_i - w_i\right)^2$$

It is then clear that the only way this new estimate could do as well is if $c_i = w_i$. By a similar argument we can show that $c_i = w_i$ must be true for the new estimator to do as well. **Can you see?** Therefore we have established that our estimator is B.L.U.E.

Units of Measurement and Functional Form
We will now examine and rationalize why we may want to use certain forms of functions in our regressions.

Effects of Changing Units of Measurement on OLS statistics:

1. If you multiply the dependent variable, *y*, by a constant, say changing millions of dollars to thousands of dollars, you are effectively changing the slope and constant parameter by that same constant.
2. If the independent variable x is multiplied by a constant, the slope parameter is divided by that same constant. The intercept term does not change.
3. The units of measure of the variables will not affect our goodness of fit measure, R^2.

Incorporating Non-linearities in OLS:
We have seen various other forms of regression equation that are linear in the parameters but not in the original variables. We will now examine when we would like to use one over the other.

1. If we wish to answer the question what is percentage effect on *y* for a 1 unit change in x, we can run the model, $\log y = a + bx + e$, where a and b are the parameters we would like to estimate. Why? This is known as the log-levels model.
2. If we wish to answer the question what is the effect on *y* as measured by its natural measure for a 1% change in *x*, we would run the following model, $y = a + b(\log x) + e$. This is known as the levels-log model.
3. If we wish to estimate a constant elasticity model where we can examine the percentage effect on *y* for a 1% change in *x*, we estimate the following, $\log y = a + b(\log x) + e$. This is known as the log-log model.

This then tells us what we are really concerned with in OLS is that the equation be linear in the parameters we are estimating and not the variables we are estimating. The crux of the matter is then how to we interpret the parameters we estimate. We will examine this as we go along.

Regression through Origin

Read your text on this part.