1 Another method of estimation: least squares

erm: ls-estim.tex, Dec8, 2009: 6 p.m. (draft - typos/writos likely exist)

Corrections, comments, suggestions welcome.

1.1 Least squares in general

Assume Y_i is some rv with finite mean μ_{y_i} and variance σ_y^2, where we might, or might not, know the form of $f_{Y_i}(y : \theta)$.

If one has a sample of n observations from this population, the least-squares estimator(s) of θ are those θ, θ_{ls}, that minimize[1]

$$\sum_{i=1}^{n} \left[(y_i - E[y_i : \mathbf{x}_i, \theta])^2 \right]$$

where the $\mathbf{x}_i$ is a vector of observed explanatory variables, not random variables (*fixed in repeated samples*).

Finding the least-squares estimate of θ requires that we specify the form of $E[y_i : \mathbf{x}_i, \theta]$ but does **not** require that we specify $f_{Y_i}(y_i, \mathbf{x}_i, \theta)$. Note that maximum likelihood estimation typically requires that we specify $f_{Y_i}(y_i, \mathbf{x}_i, \theta)$, which implies $E[y_i : \mathbf{x}_i, \theta]$.

For example, consider the following common additive specification for y_i

$$y_i = g(\mathbf{x}_i : \theta) + \varepsilon_i \quad i = 1, 2, ..., n$$

and ε is a rv with zero mean ($E[\varepsilon] = 0$) and finite variance, $\sigma_\varepsilon^2 = \sigma_y^2$.[2][3]

We have a data set that consists of n $\{y_i, \mathbf{x}_i\}$ pairs.

Since

$$E[y_i : \mathbf{x}_i] = g(\mathbf{x}_i : \theta)$$

the least-squares estimator(s) of θ are those θ that minimize

$$\sum_{i=1}^{n} \left[(Y_i - g(\mathbf{x}_i : \theta))^2 \right] = SSR$$

where SSR denotes the *sum of squared residuals*. Some books call it RSS (for example, Gujarati, page 171)

[1] Note that I did not say *random sample*. While a random sample would be nice, least-squares estimation is well defined even if the sample is not random. That said, the least-squares estimators might lack desirable properties if the sample is not random.

[2] For a given $\mathbf{x}_i$, all the randomness in y_i is invoked by the randomness in ε

[3] Note that here I am not being completely general. I am assuming the random component is additive, which is not required for l.s. estimation.

Things to note about least-squares estimators if one is willing to assume $y_i = g(\mathbf{x}_i : \theta) + \varepsilon_i \;\; i = 1, 2, ..., n$ where ε_i is a rv with zero mean ($E[\varepsilon] = 0$) and finite variance,$\sigma_\varepsilon^2 = \sigma_y^2$:

- One does not need to assume a specific distribution for ε (normal or otherwise), but one needs to put the above few restrictions on ε.
- $g(\mathbf{x}_i : \theta)$ does not have to be linear in the θ, but that is the specification that you are most accustomed to.
- Some of the properties of the estimators of the θ will depend on what one assumes about the disribution of ε (normal or otherwise) and/or whether one assumes the Y_i in the sample are independent of one another.

1.1.1 An aside:

Note that while we are not accustomed to thinking this way, all that one needs to do least squares is to assume the rv of interest Y, has a density function such that $E[Y]$ exists. For example, one could assume Y has the Poisson distribution, $f_Y(y) = \frac{e^{-\lambda}\lambda^y}{y!}$ for $y = 0, 1, 2, 3, ...,$ and use least squares to estimate λ , the expected value of Y (and also its variance).[4] The least-squares estimator of λ, λ_{ls}, is that λ that minimizes $\sum_{i=1}^{n}\left[(Y_i - \lambda)^2\right]$.[5]

A fun, and instructive exercise would be to find the least squares estimator(s) for a rv Y assuming a few different forms for $f_Y(y : \theta)$. For example, could one proceed with least-square assuming Y has a Bernoulli distribution? Try it and see what happens.

1.2 Revert to the standard assumption that $y_i = g(\mathbf{x}_i : \theta) + \varepsilon_i$, but now be more restrictive: assume linearity and x_i a scalar

$$g(x_i : \theta) = \alpha + \beta x_i$$

In which case

$$y_i = \alpha + \beta x_i + \varepsilon_i \;\; i = 1, 2, ..., n$$

[4] Note that here the random term is not additive. While assuming an additive term ($y_i = g(x_i : \theta) + \varepsilon_i \;\; i = 1, 2, ..., n$) is typical in least-squares, it, as I noted above, it is not necessary.

[5] We know from earlier, that the maximium likelihood estimator of λ, λ_{ml}, is the sample average. Is the least-squares estimator of λ also the sample average?

where $E[\varepsilon] = 0$ and ε has finite variance σ_ε^2. This model is called the *linear regression model* (MGB 485, 486). It has three parameters: α, β and $\sigma_\varepsilon^2 = \sigma_Y^2$.

Contrast the linear regression model with the *classical linear regression model*, which adds the assumption $\varepsilon \sim N(0, \sigma_\varepsilon^2)$.

The least-squares estimates of α and β are those estimates, α_{ls} and β_{ls}, that minimize

$$\sum_{i=1}^{n} \left[(y_i - E(y : x_i))^2\right] = \sum_{i=1}^{n} \left[(y_i - (\alpha + \beta x_i))^2\right]$$

1.2.1 Let's find these estimates. Minimize

(let me know if you find an typos in the following derivations)

$$SSR = \sum_{i=1}^{n} \left[(y_i - (\alpha + \beta x_i))^2\right]$$

wrt α and β. Since, we have put no restrictions on the ranges of α and β, we are looking for an interior solution in terms of these two variables

$$\begin{aligned}
\frac{\partial SSR}{\partial \alpha} &= \sum_{i=1}^{n} 2(y_i - \alpha - \beta x_i)(-1) \\
&= -2\sum_{i=1}^{n}(y_i - \alpha - \beta x_i) \\
&= -2\left[\sum_{i=1}^{n} y_i - n\alpha - \beta \sum_{i=1}^{n} x_i\right] \\
&= -2\left[n\bar{y} - n\alpha - \beta n\bar{x}\right]
\end{aligned}$$

Set this equal to zero, and solve for α to obtain

$$\alpha = \bar{y} - \beta\bar{x}$$

Now consider

$$\begin{aligned}
\frac{\partial SSR}{\partial \beta} &= \sum_{i=1}^{n} 2(y_i - \alpha - \beta x_i)(-x_i) \\
&= -2\left[\sum_{i=1}^{n} y_i x_i - \alpha x_i - \beta x_i^2\right]
\end{aligned}$$

Set this equal to zero and solve for β[6]

$$\begin{aligned} 0 &= \sum_{i=1}^{n} y_i x_i - \alpha x_i - \beta x_i^2 \\ &= \sum_{i=1}^{n} y_i x_i - \alpha \sum_{i=1}^{n} x_i - \beta \sum_{i=1}^{n} x_i^2 \\ &= \sum_{i=1}^{n} y_i x_i - \alpha n\bar{x} - \beta \sum_{i=1}^{n} x_i^2 \end{aligned}$$

which Implies

$$\beta = \frac{\sum_{i=1}^{n} y_i x_i - \alpha n\bar{x}}{\sum_{i=1}^{n} x_i^2}$$

Substitute in $\bar{y} - \beta\bar{x}$ for α to obtain

$$\begin{aligned} \beta &= \frac{\sum_{i=1}^{n} y_i x_i - n\bar{x}(\bar{y} - \beta\bar{x})}{\sum_{i=1}^{n} x_i^2} \\ &= \frac{\sum_{i=1}^{n} y_i x_i - n\bar{x}\bar{y} + \beta n\bar{x}^2}{\sum_{i=1}^{n} x_i^2} \\ &= \frac{\sum_{i=1}^{n} y_i x_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2} + \frac{\beta n\bar{x}^2}{\sum_{i=1}^{n} x_i^2} \end{aligned}$$

which implies

$$\beta - \frac{\beta n\bar{x}^2}{\sum_{i=1}^{n} x_i^2} = \frac{\sum_{i=1}^{n} y_i x_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2}$$

Note the following rearrangement of the lhs, $\beta - \frac{\beta n\bar{x}^2}{\sum_{i=1}^{n} x_i^2}$

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

[6] Note that

$$\begin{aligned} \frac{\partial SSR}{\partial \beta} &= \sum_{i=1}^{n} 2(y_i - \alpha - \beta x_i)(-x_i) = 0 \\ &\Rightarrow \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)(x_i) = 0 \\ &\Rightarrow \sum_{i=1}^{n} (y_i - \alpha_{ls} - \beta_{ls} x_i)(x_i) = 0 \\ &\Rightarrow \sum_{i=1}^{n} (y_i - y_{ls_i})(x_i) = 0 \\ &\Rightarrow \sum_{i=1}^{n} (\hat{\varepsilon}_i)(x_i) = 0 \end{aligned}$$

One could check that one's least squares estimates imply this. It is a good check on your math.

so, replacing $\beta - \frac{\beta n\bar{x}^2}{\sum_{i=1}^n x_i^2}$ with$\frac{\beta\left[\sum_{i=1}^n x_i^2 - n\bar{x}^2\right]}{\sum_{i=1}^n x_i^2}$, one obtains.

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

multiplying through one gets

$$\beta\left[\sum_{i=1}^n x_i^2 - n\bar{x}^2\right] = \sum_{i=1}^n y_i x_i - n\bar{x}\bar{y}$$

which implies that

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

This is the least-squares estimate for β assuming $g(x_i : \theta) = \alpha + \beta x_i$.

By substitution, the least-squares estimate for α, α_{ls}, is

$$\alpha_{ls} = \bar{y} - \beta_{ls}\bar{x}$$

Note that, in this case,

$$\beta_{ls} = \beta_{ml}$$

and

$$\alpha_{ls} = \alpha_{ml}$$

where α_{ml} and β_{ml} are the maximum likelihood estimates assuming the classical linear regression model. That is, if one assumes a classic linear-regression model, the *ml* estimators exist and are equal to the *ls* estimators, but if one assumes only the linear regression model (don't add the assumption that $\varepsilon \sim N(0, \sigma_\varepsilon^2)$), the *ls* estimators exist, but not the *ml* estimators.

There are a number of different ways to write β_{ls}, they are all equal.

$$
\begin{aligned}
\beta_{ls} &= \frac{\sum_{i=1}^{n} y_i x_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} \\
&= \frac{\sum_{i=1}^{n} y_i x_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \\
&= \frac{\sum_{i=1}^{n} \tilde{x}_i y_i}{\sum_{i=1}^{n} \tilde{x}_i^2} \\
&= \frac{\sum_{i=1}^{n} \tilde{x}_i \tilde{y}_i}{\sum_{i=1}^{n} \tilde{x}_i^2}
\end{aligned}
$$

where $\tilde{x}_i = x_i - \bar{x}$ and $\tilde{y}_i = y_i - \bar{y}$. One uses different characterizations in different situations - depending on what one wants to demonstrate.

1.2.2 There is no least-squares estimate of σ_y^2

Note that since $\sum_{i=1}^{n}(y_i - \alpha - \beta x_i))^2$ is not a function of σ_y^2, there is not a least-squares estimator for σ_y^2. That is, what one minimizes to obtain the ls estimates is not a function of σ_y^2.

However, given α_{ls} and β_{ls}, one can estimate σ_y^2 with

$$\hat{\sigma}_y^2 = \hat{\sigma}_\varepsilon^2 = \frac{\sum_{i=1}^{n}(y_i - \alpha_{ls} - \beta_{ls} x_i))^2}{n-2}$$

The intuition for dividing by $n-2$ is that one loses two degrees of freedom in the calculation of α_{ls} and β_{ls}.

$\hat{\sigma}_\varepsilon^2$ is not a least-squares estimator, but is based on the least-squares estimators of α and β.

It is possible to show that $E[\hat{\sigma}_\varepsilon^2] = \sigma_\varepsilon^2$. See, for example, Gujarati Basic Econometrics, the appendix to chapter 3.

1.3 Remember

Least-squares estimators exist even if

$$g(x_i : \theta) \neq \alpha + \beta x_i$$

That is, $g(x_i : \theta)$ can be nonlinear in β. For example, one could assume

$$g(x_i : \theta) = e^{\beta x_i}$$

so

$$\frac{\partial e^{\beta x_i}}{\partial \beta} = x_i e^{\beta x_i}$$

which is highly nonlinear in β.

In which case,

$$y_i = e^{\beta x_i} + \varepsilon_i \quad i = 1, 2, ..., n$$

where $E[\varepsilon] = 0$ and ε has finite variance σ^2_ε. and the least-squares estimates of β is that estimate, β_{ls}, that minimizes

$$SSR = \sum_{i=1}^{n} \left[(y_i - e^{\beta x_i})^2 \right]$$

This is an example of *nonlinear least squares*.[7]

1.4 Some properties of least-squares estimators of the form $y_i = \alpha + \beta x_i + \varepsilon_i$ where $E[\varepsilon] = 0$ and ε has finite variance σ^2_ε. $i = 1, 2, ..., n$

Assume that the the Y_i in the sample are independent of one another (we have a random sample)

From above, and assuming the above linear form

$$\beta_{ls} = \frac{\sum_{i=1}^{n} \tilde{x}_i y_i}{\sum_{i=1}^{n} \tilde{x}_i^2} = \sum_{i=1}^{n} \frac{\tilde{x}_i}{k} y_i$$

where $k = \sum_{i=1}^{n} \tilde{x}_i^2$

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

where $w_i = \frac{\tilde{x}_i}{k}$. In words, β_{ls} is a linear combination (weighted sum) of the n random variables, $y_1, y_2, ..., y_n$, where the weights are a function of the $x's$.

We call estimators with this property *linear* estimators.[8]. Note that determining it was a linear estimator did not require that $f(\varepsilon)$ have a particular form.

[7] In contrast, note that if one assumed $y_i = \beta x_i^2 + \varepsilon_i$ it would still be linear least-squares because the function is linear in β.

[8] Looking ahead, this is part of the famous, Gauss-Markov theorem.

Given that $\beta_{ls} = \sum_{i=1}^{n} w_i y_i$, and given the x_i the y_i are independent

$$\begin{aligned} E[\beta_{ls}] &= E[\sum_{i=1}^{n} w_i y_i] \\ &= \sum_{i=1}^{n} w_i E[y_i] \end{aligned}$$

since the w_i are constants: they vary with x but the x are assumed fixed in repeated samples. Since $E[y_i] = \alpha + \beta x_i$

$$\begin{aligned} &= \sum_{i=1}^{n} w_i(\alpha + \beta x_i) \\ &= \alpha \sum_{i=1}^{n} w_i + \beta \sum_{i=1}^{n} w_i x_i \end{aligned}$$

Since $w_i = \frac{\tilde{x}_i}{k}$

$$= \frac{\alpha}{k} \sum_{i=1}^{n} \tilde{x}_i + \frac{\beta}{k} \sum_{i=1}^{n} \tilde{x}_i x_i$$

since$\sum_{i=1}^{n} \tilde{x}_i = \sum_{i=1}^{n}(x_i - \bar{x}) = 0$

$$= \frac{\beta}{k} \sum_{i=1}^{n} \tilde{x}_i x_i$$

Because $\tilde{x}_i = x_i - \bar{x} \ \Rightarrow x_i = \tilde{x}_i + \bar{x}$

$$\begin{aligned} &= \frac{\beta}{k} \sum_{i=1}^{n} \tilde{x}_i(\tilde{x}_i + \bar{x}) \\ &= \frac{\beta}{k} \left[\sum_{i=1}^{n} \tilde{x}_i^2 + \bar{x} \sum_{i=1}^{n} \tilde{x}_i \right] \end{aligned}$$

Because$\sum_{i=1}^{n} \tilde{x}_i = 0$

$$= \frac{\beta}{k} \sum_{i=1}^{n} \tilde{x}_i^2$$

And because $k = \sum_{i=1}^{n} \tilde{x}_i^2$

$$\begin{aligned} &= \frac{\beta}{\sum_{i=1}^{n} \tilde{x}_i^2} \sum_{i=1}^{n} \tilde{x}_i^2 \\ &= \beta \end{aligned}$$

That is

$$E[\beta_{ls}] = \beta$$

In words, β_{ls} is an unbiased estimator of β. Note that this proof did not require that we assume a specific distribution for ε. We need only the assumptions of the linear regression model, and a random sample (independent y_i).

Note that at this point we have demonstrated that β_{ls} is a *linear unbiased estimator*, and this result does not depend on a normality assumption.

It is also possible to show that

$$E[\alpha_{ls}] = \alpha$$

I leave that as an exercise for you.

In summary, the least-squares estimators of the parameters are linear and unbiased estimators.

It is possible to show that $E[\hat{\sigma}_\varepsilon^2] = \sigma_\varepsilon^2$, but remember that $\hat{\sigma}_\varepsilon^2 = \frac{\sum_{i=1}^{n}(y_i - \alpha_{ls} - \beta_{ls} x_i))^2}{n-2}$ is not a least-squares estimate.

So, assuming the linear regression model and a random sample, β_{ls} and α_{ls} are linear estimators and unbiased estimators. This is good.

It is possible to further show that in the class of linear unbiased estimators, the least-squares estimators have minimum variance. This earns them the adjective *best*.

So, assuming the linear regression model and a random sample, β_{ls} and α_{ls} are BLUE (*best linear unbiased estimators*). This is the Gauss-Markov theorem.

If one assumes $E[\varepsilon] = N(0, \sigma_\varepsilon^2)$ the estimators gain more desirable properties because they are also the *ml* estimators.

1.4.1 One can use α_{ls} and β_{ls} to predict values of y_j conditional on x_j

$$y_{ls_j} = \alpha_{ls} + \beta_{ls} x_j$$

y_{ls_j} is a random variable that, for fixed x's, will vary from sample to sample. Since α_{ls} and β_{ls} are both unbiased estimates, y_{ls_j} is an unbiased estimate of y_j; that is $E[y_{ls_j}] = y_j$. Think about the sampling distribution of y_{ls_j}, which is conditioned on x_j

1.5 The variances of the least-squares estimators

1.5.1 The variance of β_{ls}

The least-squares estimate of β is a statistic and will vary from sample to sample, so β_{ls} has a sampling distribution, $f_{\beta_{ls}}(v)$. An issue at hand is determining the variance of this sampling distribution.[9]

An important issue is whether we proceed assuming a knowledge of σ_ε^2, or only knowledge of its estimate, $\widehat{\sigma}_\varepsilon^2$. We will start assuming knowledge of σ_ε^2, and afterwards discuss how the variance of β_{ls} differs when it is expressed as a function of $\widehat{\sigma}_\varepsilon^2$ rather than σ_ε^2. Knowing σ_ε^2 is atypical, but easier, so we start there.

To emphasize that we are conditioning on σ_ε^2, in the shortrun, denote $f_{\beta_{ls}}(v)$ more specifically as $f_{\beta_{ls}}(v \left| \sigma_\varepsilon^2 \right.)$ and write $var(\beta_{ls} \left| \sigma_\varepsilon^2 \right.) \equiv \sigma_{\beta_{ls}}^2(\sigma_\varepsilon^2)$.[10]

Above we showed that β_{ls} is a linear estimator. That is, it can be written

$$\beta_{ls} = \sum_{i=1}^{n} w_i y_i$$

where the w_i can be treated as constants. We also know that $var(ax) = a^2 var(x)$ if a is a constant. Combining these two pieces of information, along with knowledge of σ_ε^2:

$$var(\beta_{ls} \left| \sigma_\varepsilon^2 \right.) = \sum_{i=1}^{n} w_i^2 \sigma_y^2 = \sum_{i=1}^{n} w_i^2 \sigma_\varepsilon^2$$

Recollect that $y_i = \alpha + \beta x_i + \varepsilon_i$ where $E[\varepsilon] = 0$ and ε has a finite variance so $\sigma_y^2 = \sigma_\varepsilon^2$.

[9] More generally, we would like to know the form of $f_{\beta_{ls}}(v)$.

[10] In contrast to $f_{\beta_{ls}}(v \left| \widehat{\sigma}_\varepsilon^2 \right.)$ and $var(\beta_{ls} \left| \widehat{\sigma}_\varepsilon^2 \right.) \equiv \sigma_{\beta_{ls}}^2(\widehat{\sigma}_\varepsilon^2)$

Proceeding,

$$\begin{aligned}
var(\beta_{ls}\left|\sigma_\varepsilon^2\right.) &= \sum_{i=1}^{n} w_i^2 \sigma_\varepsilon^2 \\
&= \sum_{i=1}^{n} \frac{\tilde{x}_i^2}{k^2} \sigma_\varepsilon^2 \\
&= \sigma_\varepsilon^2 \sum_{i=1}^{n} \frac{\tilde{x}_i^2}{k^2} \\
&= \sigma_\varepsilon^2 \sum_{i=1}^{n} \frac{\tilde{x}_i^2}{(\sum_{i=1}^{n} \tilde{x}_i^2)^2} \\
&= \frac{\sigma_\varepsilon^2}{\sum_{i=1}^{n} \tilde{x}_i^2} = \frac{\sigma_\varepsilon^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}
\end{aligned}$$

since $k = \sum_{i=1}^{n} \tilde{x}_i^2$, and the "standard error" of β_{ls} is

$$se\, of\ \beta_{ls}(\sigma_\varepsilon^2) = [var(\beta_{ls}\left|\sigma_\varepsilon^2\right.)]^{.5}$$

Notice that $var(\beta_{ls}\left|\sigma_\varepsilon^2\right.)$ decreases as $\sum_{i=1}^{n}(x_i - \bar{x})^2$ increases

What did we assume to derive $var(\beta_{ls}\left|\sigma_\varepsilon^2\right.) \equiv \sigma_{\beta_{ls}}^2(\sigma_\varepsilon^2) = \frac{\sigma_\varepsilon^2}{\sum_{i=1}^{n} \tilde{x}_i^2}$? We assumed that $y_i = \alpha + \beta x_i + \varepsilon_i$ where $E[\varepsilon] = 0$ and ε has a finite variance, and the Y_i are independent.

We did not need to assume that ε has a specific distribution, such as the normal.

It is also possible to derive the $var(\alpha_{ls}\left|\sigma_\varepsilon^2\right.)$ as a function of σ_ε^2

$$var(\alpha_{ls}\left|\sigma_\varepsilon^2\right.) \equiv \sigma_{\alpha_{ls}}^2(\sigma_\varepsilon^2) = (\frac{\sigma_\varepsilon^2}{n})(\frac{\sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} \tilde{x}_i^2})$$

Note again that we cannot calculate $var(\alpha_{ls}\left|\sigma_\varepsilon^2\right.)$ or $var(\beta_{ls}\left|\sigma_\varepsilon^2\right.)$ unless we assume a specific value for σ_ε^2.[11]

---

[11] Note that one can also calculate $cov\left(\alpha_{ls}, \beta_{ls}\left|\sigma_\varepsilon^2\right.\right)$. It is not 0 because both α_{ls} and β_{ls} are a function of σ_ε^2.

$$\begin{aligned}
& cov\left(\alpha_{ls}, \beta_{ls}\left|\sigma_\varepsilon^2\right.\right) \\
=\ & E\left[\left(\alpha_{ls} - E[\alpha_{ls}]\right)\left(\beta_{ls} - E[\beta_{ls}]\right)\right] \\
=\ & E\left[\left(\alpha_{ls} - \alpha\right)\left(\beta_{ls} - \beta\right)\right]
\end{aligned}$$

Note that $var(\beta_{ls} \left| \sigma_\varepsilon^2 \right.) \equiv \sigma_{\beta_{ls}}^2(\sigma_\varepsilon^2)) = \frac{\sigma_\varepsilon^2}{\sum_{i=1}^n \tilde{x}_i^2}$ **is a function of the** x**'s in the sample, but not the** y**'s.** Therefore, if one makes the typical assumption that the x's are "fixed in repeated samples, $\sigma_{\beta_{ls}}^2(\sigma_\varepsilon^2)$ is not a random variable. By the same argument, neither is $\sigma_{\alpha_{ls}}^2(\sigma_\varepsilon^2)$. This is because we are assuming knowledge of σ_ε^2.

This is an important point. $\sigma_{\beta_{ls}}^2(\sigma_\varepsilon^2)$ and $\sigma_{\alpha_{ls}}^2(\sigma_\varepsilon^2)$ are not statistics and not things that are estimated; they are calculated given knowledge of σ_ε^2 and knowledge of the x levels in the data.

Said another way, while the least-squares estimates of β and α will vary from sample to sample, $\sigma_{\beta_{ls}}^2(\sigma_\varepsilon^2)$ and $\sigma_{\alpha_{ls}}^2(\sigma_\varepsilon^2)$ do not vary from sample to sample (assuming the x's are "fixed in repeated samples").

Soon, we we will consider the problem of estimating $var(\beta_{ls} \left| \widehat{\sigma}_\varepsilon^2 \right.)$ and $var(\alpha_{ls} \left| \widehat{\sigma}_\varepsilon^2 \right.)$. But first,

1.5.2 The variance of y_{ls_j} as a function of σ_ε^2

One can also show that (for example, Gujarati, Essentials page 185)

$$var\left(y_{ls_j} \left| \sigma_\varepsilon^2 \right.\right) \equiv \sigma_{y_{ls_j}}^2(\sigma_\varepsilon^2) = \sigma_\varepsilon^2 \left[\frac{1}{n} + \frac{(x_j - \bar{x})^2}{\sum_{i=1}^n \tilde{x}_j^2} \right]$$

Note that this is a function of σ_ε^2 and the x's but not the y's, so not something that is estimated. It is not a random variable.

But, $\alpha_{ls} = \bar{y} - \beta_{ls}\bar{x}$ and $\alpha = \bar{y} - \beta\bar{x}$, so

$$\begin{aligned} & cov\left(\alpha_{ls}, \beta_{ls} \left| \sigma_\varepsilon^2 \right.\right) \\ = \;& E\left[\bar{y} - \beta_{ls}\bar{x} - (\bar{y} - \beta\bar{x})\right](\beta_{ls} - \beta)] \\ = \;& E\left[(-\beta_{ls}\bar{x} + \beta\bar{x})(\beta_{ls} - \beta)\right] \\ = \;& -\bar{x}E\left[(\beta_{ls} - \beta)^2\right] \\ = \;& \frac{-\bar{x}\sigma_\varepsilon^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{-\bar{x}\sigma_\varepsilon^2}{\sum_{i=1}^n \tilde{x}_i^2} \end{aligned}$$

The covariance decreases as the variation in the x's increases.

1.6 What is implied if one adds to the above the assumption that $\varepsilon \tilde{} N(o, \sigma_\varepsilon^2)$.

If $y_i = \alpha + \beta x_i + \varepsilon_i$ where $\varepsilon \tilde{} N(o, \sigma_\varepsilon^2)$, then $y_i \tilde{} N(\alpha + \beta x_i, \sigma_\varepsilon^2)$. From earlier we know that $\beta_{ls} = \sum_{i=1}^n w_i y_i$, so β_{ls} is a linear combination of normally distributed random variables, so β_{ls} is normally distributed. Specifically

$$\beta_{ls} \tilde{} N(\beta, \frac{\sigma_\varepsilon^2}{\sum_{i=1}^n \tilde{x}_i^2})$$

By the same logic

$$\alpha_{ls} \tilde{} N(\alpha, (\frac{\sigma_\varepsilon^2}{n})(\frac{\sum_{i=1}^n x_i^2}{\sum_{i=1}^n \tilde{x}_i^2}))$$

When σ_ε^2 is known, neither β_{ls} or α_{ls} has a t distribution, both are normally distributed. I say this here because some people incorrectly believe that β_{ls} and α_{ls} always have a t distribution

If $\beta_{ls} \tilde{} N(\beta, \sigma_{\beta_{ls}}^2)$ then

$$\frac{\beta_{ls} - \beta}{\sigma_{\beta_{ls}}} = \frac{\beta_{ls} - \beta}{\sigma_\varepsilon / (\sum_{i=1}^n \tilde{x}_i^2)^{.5}} \tilde{} N(0, 1)$$

So, if we assumed a value for σ_ε^2, we could calculate $\sigma^2_{\beta_{ls}}$ (not estimate it) and then calculate a confidence interval for β and test the null hypothesis that β takes some specific value such as zero.

For example, since

$$\frac{\beta_{ls}-\beta}{\sigma_{\beta_{ls}}}\,\tilde{}\,N(0,1)$$

$$\begin{aligned}
\Pr\left(-1.96 \leq \frac{\beta_{ls}-\beta}{\sigma_{\beta_{ls}}} \leq 1.96\right) &= .95 \\
&\Rightarrow \Pr\left(-1.96\sigma_{\beta_{ls}} \leq \beta_{ls}-\beta \leq 1.96\sigma_{\beta_{ls}}\right) = .95 \\
&\Rightarrow \Pr\left(-1.96\sigma_{\beta_{ls}} - \beta_{ls} \leq -\beta \leq 1.96\sigma_{\beta_{ls}} - \beta_{ls}\right) = .95 \\
&\Rightarrow \Pr\left(\beta_{ls} - 1.96\sigma_{\beta_{ls}} \leq \beta \leq \beta_{ls} + 1.96\sigma_{\beta_{ls}}\right) = .95
\end{aligned}$$

$\left(\beta_{ls} - 1.96\sigma_{\beta_{ls}} \leq \beta \leq \beta_{ls} + 1.96\sigma_{\beta_{ls}}\right)$ is the 95% confidence interval for β based on $\sigma_{\beta_{ls}}$, and the assumption that ε is normally distributed.

How do we interpret this interval? Note that this interval depends on β_{ls} which is a random variable, so the confidence interval is a random variable. 95% of these intervals will contain β.[12][13]

Assuming that $\varepsilon\,\tilde{}\,N(o,\sigma_\varepsilon^2)$, it follows that y_{ls_j} is also normally distributed (it is a linear function of two normally distributed random variables (α_{ls} and β_{ls}). Specifically,

$$y_{ls_j}\,\tilde{}\,N\left(\alpha+\beta x_j, \sigma_\varepsilon^2\left[\frac{1}{n}+\frac{(x_j-\bar{x})^2}{\sum_{i=1}^n \tilde{x}_j^2}\right]\right)$$

So one can also get a confidence interval for y_i

[12] Note that one cannot say that there is a 95% chance that the true β is between $(\beta_{ls} - 1.96\sigma_{\beta_{ls}})$ and $(\beta_{ls} + 1.96\sigma_{\beta_{ls}})$. Further note that since $\sigma_{\beta_{ls}}$ is not a random variable if the x's are fixed in repeated sample, the position of this confidence interval is a random variable, but not its length.

[13] Note that none of the above has anything to do with the t distribution.

1.7 However, we don't typically assume a value for σ_ε^2 but estimate it with $\hat{\sigma}_\varepsilon^2$

Continue, for now, to assume that $\varepsilon\tilde{}N(o,\sigma_\varepsilon^2)$, so assume the CLR model, but that we do not know σ_ε^2, so have to estimate it

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_{i=1}^{n}\left(y_i - \alpha_{ls} - \beta_{ls}x_i\right)^2}{(n-2)}$$

and note the important distinction between $\hat{\sigma}_\varepsilon^2$ and σ_ε^2, the first is a rv, the second is a constant.

The first thing to note, as we demonstrate below, is even though $\varepsilon\tilde{}N(o,\sigma_\varepsilon^2)$

$$\frac{\beta_{ls} - \beta}{\hat{\sigma}_{\beta_{ls}}} \nsim N(0,1)$$

where

$$\hat{\sigma}_{\beta_{ls}}^2 = \frac{\hat{\sigma}_\varepsilon^2}{\sum_{i=1}^{n}\tilde{x}_i^2}$$

Toooooo bad.[14]

[14] Note that $\beta_{ls}\tilde{}N(\beta,\sigma_{\beta_{ls}}^2)$ because we are assuming $\varepsilon\tilde{}N(o,\sigma_\varepsilon^2)$

Since it is not normal, what distribution does

$$\frac{\beta_{ls} - \beta}{\hat{\sigma}_{\beta_{ls}}}$$

have?

Let's try and demonstrate that it has a t distribution. The following is a bit difficult - think of it as walking backwards from the end of the trail back to your car, forgetting where you started. What I am doing is deriving the distribution of $\frac{\beta_{ls}-\beta}{\hat{\sigma}_{\beta_{ls}}}$.

Remember that that $\frac{\beta_{ls}-\beta}{\sigma_{\beta_{ls}}} \tilde{} N(0,1)$.

Now define another random variable, G, remember we are going backwards, such that

$$
\begin{aligned}
G &\equiv \frac{(n-2)\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} \text{ note that I have defined a function that is a linear function of the ratio } \frac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} \\
&= \frac{(n-2)\frac{\sum(Y-\alpha_{ls}-\beta_{ls}x_i)^2}{(n-2)}}{\sigma_\varepsilon^2} \text{ the reason for } (n-2) \text{ above was so it would cancel here} \\
&= \frac{\sum(Y_i-\alpha_{ls}-\beta_{ls}x_i)^2}{\sigma_\varepsilon^2} \\
&= \frac{\sum(Y_i-E[y_{ls_j}\,|x_i\,])^2}{\sigma_\varepsilon^2} \\
&= \frac{\sum(Y_i-E[Y_i\,|x_i\,])^2}{\sigma_\varepsilon^2} \text{ because } y_{ls_j}\,|x_i \text{ is an unbiased estimate of } E[Y_i\,|x_i\,] \\
&= \frac{\sum(Y_i-E[Y_i\,|x_i\,])^2}{\sigma_y^2} \\
&= \sum_{i=1}^{n}\left(\frac{(Y_i-E[Y_i\,|x_i\,])}{\sigma_y}\right)^2
\end{aligned}
$$

Note that $\hat{\sigma}_\varepsilon^2$ does not explicitly appear in this last term - we started with it, but it disappeared.

Further note that

$$\left(\frac{(y_i-E[y_i\,|x_i\,])}{\sigma_y}\right)\tilde{\ }N(0,1)$$

because $y_i\tilde{\ }N(E[y_i\,|x_i\,],\sigma_y^2)$.

Note, this is critical, our created random variable, $G = \sum_{i=1}^{n} \left(\frac{(Y_i - E[Y_i|x_i])}{\sigma_y} \right)^2$, is the sum of the squares of a bunch of standard normal random variables. That means it has a χ^2 distribution.[15]

The important thing to remember at this point is that we have created a random variable G that is a linear function of the ratio $\frac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2}$ and we know its density function. You want to learn what you can about the Chi-squared distribution (keep in mind, saying k is the *degrees of freedom* of the χ^2 distribution is just another way of saying the χ^2 density function has one parameter, k).

Specifically,

$$\begin{aligned} G &= \sum_{i=1}^{n} \left(\frac{(Y_i - E[Y_i|x_i])}{\sigma_y} \right)^2 \\ &= \frac{(n-2)\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} \sim \chi_{n-2}^2 \end{aligned}$$

It is $(n-2)$ because of the parameter (number of degrees of freedom) is not the number of terms in the sum, but the number of independent terms in the sum, which is $(n-2)$ because we lose two degrees of freedom to get $E[Y_i|x_i] = \alpha_{ls} + \beta_{ls} x_i$. That is, $G = \frac{(n-2)\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2}$ is a rv with a Chi-squared distribution with parameter $(n-2)$.

(The bottom line is someone worked backward and figured out a rv that was a function of $\hat{\sigma}_\varepsilon^2$ and σ_ε^2, and that had a Chi-square distribution.) Note that neither $\hat{\sigma}_\varepsilon^2$ nor σ_ε^2 is a **parameter** in the Chi-square, which is important.

[15] See Gujarati page 114 and MGB pages 241-243). Theorem 7 (MGB page 242) states that "If random variables X_i, $i = 1, 2, .., k$, are normally distributed with means μ_i and variances σ_i^2, then $U = \sum_{i=1}^{n} \left(\frac{X_i - \mu_i}{\sigma_i} \right)^2$ has a chi-square distribution with parameter k (k degress of freedom). A collarary is that if $X_1, X_2, .., X_n$ is a random sample from a normal distribution with mean mean μ and variance σ^2 then $\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2$ has a chi-square distribution with n degrees of freedom. A special case is that $\left(\frac{X_i - \mu}{\sigma} \right)^2$ has a chi-square distribution with 1 degree of freedom.

So what do we know at this point?

$$\frac{\beta_{ls}-\beta}{\sigma_{\beta_{ls}}} = \frac{\beta_{ls}-\beta}{\sigma_{\varepsilon}/(\sum_{i=1}^{n}\tilde{x}_i^2)^{.5}} \tilde{} N(0,1)$$

and

$$\frac{(n-2)\hat{\sigma}_{\varepsilon}^2}{\sigma_{\varepsilon}^2} \tilde{} \chi_{n-2}^2$$

So, now let's mention the t distribution. MGB 249-251 tell us

$$\frac{N(0,1)}{\left(\left(\chi_{n-2}^2\right)/(n-2)\right)^{.5}} \tilde{} t_{n-2}$$

That is, a standard normal rv, e.g. $\frac{\beta_{ls}-\beta}{\sigma_{\beta_{ls}}}$, divided by the square root of a χ^2 rv (divided by its parameter) has a t distribution with that parameter.[16]

[16] Theorem 10 (MGB page 250) states that "If the rv Z has a standard normal distribution, if the rv U has a chi-squared distribution with (degrees of freedom k), and if Z and U are independent, $Z/\left(U/k\right)^{.5}$ has a Student t distribution with parameter k (degrees of freedom). A relevant Corollary is on page 250.

So, let's divide and see what simplifies. Define the rv W

$$\begin{aligned}
W &\equiv \frac{N(0,1)}{\left(\left(\chi^2_{n-2}\right)/(n-2)\right)^{.5}} \\
&= \frac{\frac{\beta_{ls}-\beta}{\sigma_{\beta_{ls}}}}{\left(\left(\chi^2_{n-2}\right)/(n-2)\right)^{.5}} \\
&= \frac{\frac{\beta_{ls}-\beta}{\sigma_\varepsilon/(\sum_{i=1}^n \tilde{x}_i^2)^{.5}}}{\left(\left(\chi^2_{n-2}\right)/(n-2)\right)^{.5}} \\
&= \frac{\frac{\beta_{ls}-\beta}{\sigma_\varepsilon/(\sum_{i=1}^n \tilde{x}_i^2)^{.5}}}{\left(\left(\frac{(n-2)\hat{\sigma}^2_\varepsilon}{\sigma^2_\varepsilon}\right)/(n-2)\right)^{.5}} \\
&= \frac{\frac{\beta_{ls}-\beta}{\sigma_\varepsilon/(\sum_{i=1}^n \tilde{x}_i^2)^{.5}}}{\left(\left(\frac{(\hat{\sigma}^2_\varepsilon}{\sigma^2_\varepsilon}\right)\right)^{.5}} \\
&= \frac{\frac{\beta_{ls}-\beta}{\sigma_\varepsilon/(\sum_{i=1}^n \tilde{x}_i^2)^{.5}}}{\frac{\hat{\sigma}_\varepsilon}{\sigma_\varepsilon}}. \\
&= \frac{\beta_{ls}-\beta}{\sigma_\varepsilon/(\sum_{i=1}^n \tilde{x}_i^2)^{.5}} \frac{\sigma_\varepsilon}{\hat{\sigma}_\varepsilon} \text{ Note that } \sigma_\varepsilon \text{ cancels out;} \\
& \quad \text{this is critical since we don't know it.} \\
&= \frac{\beta_{ls}-\beta}{\hat{\sigma}_\varepsilon/(\sum_{i=1}^n \tilde{x}_i^2)^{.5}} \\
&= \frac{\beta_{ls}-\beta}{\hat{\sigma}_{\beta_{ls}}}\tilde{\ } t_{n-2}
\end{aligned}$$

if $y_i = \alpha + \beta x_i + \varepsilon_i$ where $\varepsilon\tilde{\ }N(0,\sigma^2_\varepsilon)$.

So, to say it explicitly, we have determined that $\frac{\beta_{ls}-\beta}{\hat{\sigma}_{\beta_{ls}}}$ **has a** t **distribution with parameter** $(n-2)$[17] It took a lot of what we have learned to derive this.

Consider an example.

If[18] $n = 32$

$$\frac{\beta_{ls} - \beta}{\hat{\sigma}_{\beta_{ls}}} \,\tilde{}\, t_{30}$$

In which case $\Pr(t_{30} > 2.042) = .025$ and $\Pr(t_{30} < -2.042) = .025$ from the t table. So,

$$\Pr\left(-2.042 < \frac{\beta_{ls} - \beta}{\hat{\sigma}_{\beta_{ls}}} < 2.042\right) = .95$$

$\Longleftrightarrow$

$$\Pr\left(\beta_{ls} - 2.042\hat{\sigma}_{\beta_{ls}} < \beta < \beta_{ls} + 2.042\hat{\sigma}_{\beta_{ls}}\right) = .95$$

The interval $\beta_{ls} - 2.042\hat{\sigma}_{\beta_{ls}}$ to $\beta_{ls} + 2.042\hat{\sigma}_{\beta_{ls}}$ is the 95% confidence interval for β based on $\hat{\sigma}_{\beta_{ls}}$ rather than $\sigma_{\beta_{ls}}$.

This interval is a random variable; 95% of these intervals will include β.

Contrast this confidence interval with

$$\Pr\left(\beta_{ls} - 1.96\sigma_{\beta_{ls}} < \beta < \beta_{ls} + 1.96\sigma_{\beta_{ls}}\right) = .95$$

which we derived earlier.

A hypothesis test How would one determine whether they can reject the null hypothesis that $\beta = 4$? One can derive the confidence interval for β and see if its includes 4. Alternatively, one can directly use

$$\frac{\beta_{ls} - \beta}{\hat{\sigma}_{\beta_{ls}}} \,\tilde{}\, t_{n-2}$$

If $\beta = 4$, the null is correct

$$\frac{\beta_{ls} - 4}{\hat{\sigma}_{\beta_{ls}}} \,\tilde{}\, t_{n-2}$$

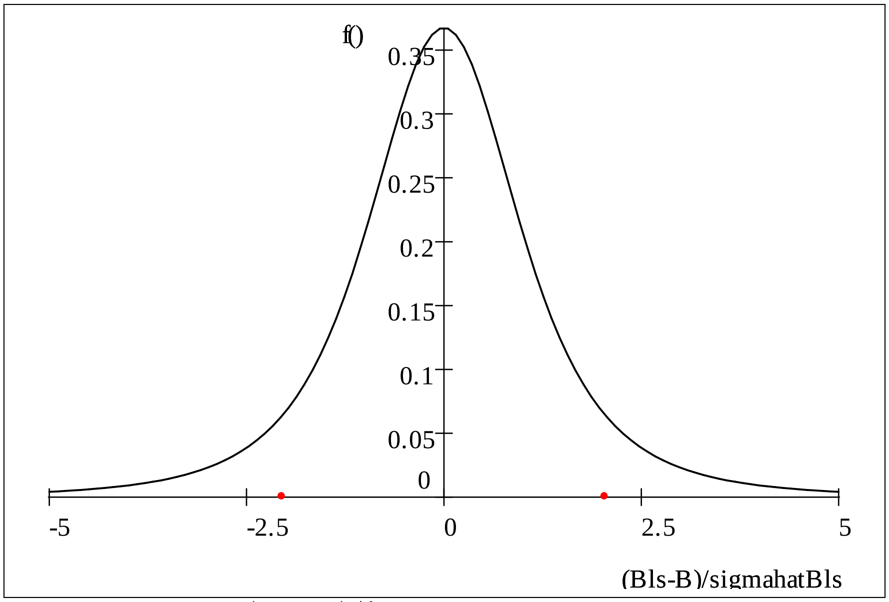
[17] This is close but different from saying that β_{ls} has a t distribution.

[18] If $\frac{\beta_{ls}-\beta}{\hat{\sigma}_{\beta_{ls}}} \,\tilde{}\, t_{n-2}$, $E[\frac{\beta_{ls}-\beta}{\hat{\sigma}_{\beta_{ls}}}] = 0$ and it variance is $\frac{n-2}{(n-2)-2} = \frac{n-2}{(n-4)}$. In explanation, all t distributions have a mean of zero, and $\frac{n-2}{(n-4)}$ is the variance of all t distributions.

Note that since a value of β is assumed, this is a calculable number. For example, if $n = 32$, $\beta_{ls} = 8$, and $\hat{\sigma}_{\beta_{ls}} = 2$ then $\frac{\beta_{ls}-4}{\hat{\sigma}_{\beta_{ls}}} = \frac{8-4}{2} = 2$ If one chooses a two-tailed test (.025 in each tail), the critical value of t is 2.042. In this case,

$$\frac{8-4}{3} = 2.0 < 2.045$$

and one fails to reject the null hypothesis that $\beta = 4$.[19]



$(\beta_{ls} - \beta)/\widehat{\sigma}_{\beta_{ls}}$ has a t distribution

Most basic OLS regression packages print out the t values corresponding to the null hypothesis $\beta = 0$. Be aware that these t statistics don't mean much unless you are willing to assume that ε is normally distributed.

[19] Note that these t values make no sense if one does not assume $\varepsilon\tilde{~}N\left(0, \sigma_\varepsilon^2\right)$. That is, if one does not adopt this assumption, the random variable $\frac{\beta_{ls}}{\hat{\sigma}_{\beta_{ls}}}$ does not have a t distribution. Said a different way, if you are unwilling to assume $\varepsilon\tilde{~}N\left(0, \sigma_\varepsilon^2\right)$ you better not be paying any attention to the t values your OLS package printed out.

Now derive the 95% confidence interval for σ_ε^2 assuming $n = 32$ We are still assuming the CLR model and no knowledge of σ_ε^2.

Earlier we showed that

$$G = \frac{(n-2)\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} \tilde{\ } \chi_{n-2}^2$$

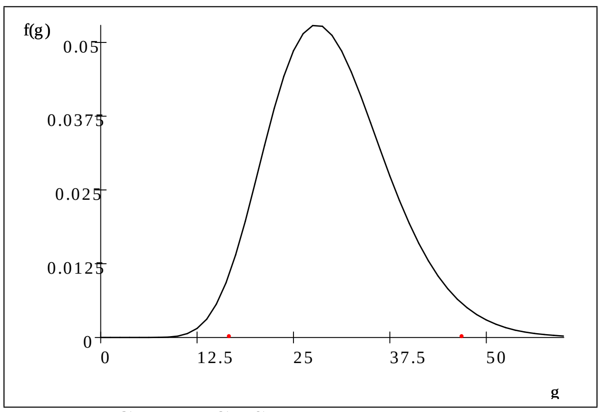
Using the χ^2 table one can determine that

$$\Pr\left(\chi_{30}^2 > 46.98\right) = .025$$

and

$$\Pr\left(\chi_{30}^2 < 16.79\right) = .025$$

Below is the density function for χ_{30}^2; 2.5% of the area is to the left of 16.79 and 2.5% is to the right of 46.98.



G has a ChiSquared distribution

We are still assuming the CLR model and no knowledge of σ_ε^2.

Earlier we showed that

$$G = \frac{(n-2)\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} \tilde{\ } \chi_{n-2}^2$$

So

$$\Pr\left(16.79 \le \frac{30\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} \le 46.98\right) = .95$$

$\Rightarrow$

$$= \Pr\left(\frac{16.79}{30\hat{\sigma}_\varepsilon^2} \le \frac{1}{\sigma_\varepsilon^2} \le \frac{46.98}{30\hat{\sigma}_\varepsilon^2}\right) = .95$$

$\Rightarrow$

$$\Pr\left(\frac{30\hat{\sigma}_\varepsilon^2}{16.79} \ge \sigma_\varepsilon^2 \ge \frac{30\hat{\sigma}_\varepsilon^2}{46.98}\right) = .95$$

$\Rightarrow$

$$\Pr\left(\frac{30\hat{\sigma}_{\varepsilon}^{2}}{46.98} \leq \sigma_{\varepsilon}^{2} \leq \frac{30\hat{\sigma}_{\varepsilon}^{2}}{16.79}\right) = .95$$

$\Rightarrow$

$$\Pr\left(.638\hat{\sigma}_{\varepsilon}^{2} \leq \sigma_{\varepsilon}^{2} \leq 1.786\hat{\sigma}_{\varepsilon}^{2}\right) = .95$$

So, we have derived a confidence interval on the population parameter σ_{ε}^{2} as a function of $\hat{\sigma}_{\varepsilon}^{2}$.

Note that the confidence interval, $.638\hat{\sigma}_{\varepsilon}^{2} \leq \sigma_{\varepsilon}^{2} \leq 1.786\hat{\sigma}_{\varepsilon}^{2}$, is a random variable; 95% of these intervals will include σ_{ε}^{2}.

If one wanted to test the null hypothesis that σ_{ε}^{2} takes some specific value, e.g. 4, one can either see if 4 is in the interval $.638\hat{\sigma}_{\varepsilon}^{2} \leq \sigma_{\varepsilon}^{2} \leq 1.786\hat{\sigma}_{\varepsilon}^{2}$. Or one can directly use the fact that

$$\frac{(n-2)\hat{\sigma}_{\varepsilon}^{2}}{\sigma_{\varepsilon}^{2}} \tilde{} \chi_{n-2}^{2}$$

Plugging in the 4 and $n = 32$

$$\frac{(30)\hat{\sigma}_{\varepsilon}^{2}}{4} = 7.5\hat{\sigma}_{\varepsilon}^{2} \tilde{} \chi_{n-2}^{2}$$

From above, for a two-tailed test at the .05 significance level, the critical values of χ_{30}^{2} are 16.79 and 46.98. So if $\hat{\sigma}_{\varepsilon}^{2} \geq \frac{46.98}{7.5} = 6.26$, one would reject the null hypothesis that $\sigma_{\varepsilon}^{2} = 4$. One would also reject this null hypothesis if $\hat{\sigma}_{\varepsilon}^{2} \leq \frac{16.798}{7.5} = 2$.

How about a confidence interval for y_i, conditional on x_i, assuming $\varepsilon \tilde{} N(0, \sigma_\varepsilon^2)$ and no knowldege of σ_ε^2? From above we know that

$$y_{ls_j} \tilde{} N\left(\alpha + \beta x_j, \sigma_\varepsilon^2 \left[\frac{1}{n} + \frac{(x_j - \bar{x})^2}{\sum_{i=1}^n \tilde{x}_j^2}\right]\right)$$

If we replace σ_ε^2 with $\hat{\sigma}_\varepsilon^2$ it no longer has a normal distribution. But, by the same argument as above

$$\frac{y_{ls_j} - E[y_{lsj} \,|x_j\,]}{\hat{\sigma}_{y_{ls_j}}} \tilde{} t_{n-2}$$

This implies, still assuming $n = 32$,

$$\Pr\left(-2.042 < \frac{y_{ls_j} - E[y_{lsj} \,|x_j\,]}{\hat{\sigma}_{y_{ls_j}}} < 2.042\right) = .95$$

$\Rightarrow$

$$\Pr\left(y_{ls_j} - 2.042\hat{\sigma}_{y_{ls_j}} < y_j \,|x_j\, < y_{ls_j} + 2.042\hat{\sigma}_{y_{ls_j}}\right) = .95$$

So, 95% of the intervals, $\left(y_{ls_j} - 2.042\hat{\sigma}_{y_{ls_j}} < y_j \,|x_j\, < y_{ls_j} + 2.042\hat{\sigma}_{y_{ls_j}}\right)$, will contain the true y_j conditional on x_j.

Note that this interval takes it minimum value when $x_j = \bar{x}$, decreases as $x_j \rightarrow \bar{x}$. How do I know this?

1.7.1 What if I don't know the distribution of ε but am willing to assume $E[\varepsilon] = 0$ and that ε has a finite but unknown variance σ_ε^2 ?

We are now assuming a LRM, but not a CLRM.

In this case we still can do OLS estimation, and, as we saw, α_{ls}, β_{ls}, and y_{ls_j} are BLUE estimators. We can also calculate

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_{i=1}^n (y_i - \alpha_{ls} - \beta_{ls} x_i)^2}{(n-2)}$$

and

$$\hat{\sigma}_{\beta_{ls}}^2 = \frac{\hat{\sigma}_\varepsilon^2}{\sum_{i=1}^n \tilde{x}_i^2}$$

To do hypothesis tests or interval estimation on β, we need to determine the distribution of

$$\frac{\beta_{ls} - \beta}{\hat{\sigma}_{\beta_{ls}}}$$

Note that we cannot assume that $\frac{\beta_{ls}-\beta}{\sigma_{\beta_{ls}}} \tilde{} N(0,1)$. If it were normal, one can determine (above) that $\frac{\beta_{ls}-\beta}{\hat{\sigma}_{\beta_{ls}}} \tilde{} t_{n-2}$, but now we can't determine the distribution $\frac{\beta_{ls}-\beta}{\hat{\sigma}_{\beta_{ls}}}$. To do so we need to know the distribution of ε, which we do not.

1.7.2 What if we know the distribution of ε and it is not normal?

Now we are assuming a LRM and knowledge of $f_\varepsilon(\varepsilon)$, which is not normal. So we are not assuming the CLR model.

Assume for example, $\varepsilon \tilde{} S\left(0, \sigma_\varepsilon^2\right)$ where S denotes the Snerd distribution, where the Snerd is not the normal distribution - to start, assume you know σ_ε^2. In this case, is

$$\frac{\beta_{ls} - \beta}{\sigma_{\beta_{ls}}} \tilde{} S(0,1)\,?$$

That is, does it have a standardized Snerd distribution?

The answer is sometimes but not always.[20] If you could show that $\frac{\beta_{ls}-\beta}{\sigma_{\beta_{ls}}} \tilde{} S(0,1)$ one could do confidence intervals and hypothesis tests for assumed values of σ_ε^2.

If one replaces σ_ε^2 with $\hat{\sigma}_\varepsilon^2$,

$$\frac{\beta_{ls} - \beta}{\hat{\sigma}_{\beta_{ls}}}$$

will **definitely not** have a Snerd distribution or a student t distribution. In theory, one could figure out the distribution of this rv (along the lines we did it assuming normality) and then do hypothesis tests and confidence intervals. This could be tough.

Now again assume you know σ_ε^2, continuing to assume ε has a Snerd distribution To simulate estimated confidence intervals for β_{ls} and α_{ls}, one might proceed as follows: Assume the data-generating process for your real-world population of interest is the LRM with $\varepsilon \tilde{} S\left(0, \sigma_\varepsilon^2\right)$, where the value of σ_ε^2 is known - $S\left(0, \sigma_\varepsilon^2\right)$ is completely specified. Estimate β_{ls} and α_{ls} for this sample. Then assume β_{ls}, α_{ls} and σ_ε^2 are the population parameters; that is, your suedo data-generating process is $y_i = \alpha_{ls} + \beta_{ls} x_i + \varepsilon_i$ where $\varepsilon \tilde{} S\left(0, \sigma_\varepsilon^2\right)$. For the vector of x, $x_1, x_2, ...x_i, ..., x_n$ generate S different random samples of size n based on the suedo data-generating process; make S a large number. For each sample s, estimate β_{ls}^s and α_{ls}^s. Plot the distribution of the S β_{ls}^s and the distribution of the S α_{ls}^s. The former is an estimated sampling distribution for β_{ls}, centered on β_{ls}, the latter an estimated sampling distribution for α_{ls}, centered on α_{ls}. A 95% confidence for each can be estimated by lopping off the top and bottom 2.5% of each of these estimated distributions.

[20] For example if ε had a t distribution, $\frac{\beta_{ls}-\beta}{\sigma_{\beta_{ls}}}$ would not have a t distribution. But we know that if ε is normal then $\frac{\beta_{ls}-\beta}{\sigma_{\beta_{ls}}}$ is normal.

Note, these estimated confidence intervals are a function of the initial random sample from your population, the assumption that one has a LRM, the assumption that $\varepsilon\tilde{}S\left(0,\sigma_{\varepsilon}^{2}\right)$, that σ_{ε}^{2} is known, and n: it is definitely a function of the Snerd assumption and σ_{ε}^{2}. The larger n the shorter the confidence interval.

Note, one does not need to theoretically derive either $f(\frac{\beta_{ls}-\beta}{\sigma_{\beta_{ls}}})$ or $f(\beta_{ls})$: the latter was derived by simulation.

Now continue to assume ε has a Snerd distribution but now assum the value of σ_{ε}^{2} is unknown. To simulate estimated confidences intervals for β_{ls}, α_{ls}, $\widehat{\sigma}_{\varepsilon}^{2}$ one might proceed as follows: Assume the data-generating process for your real-world population of interest is the LRM with $\varepsilon\tilde{}S\left(0,\sigma_{\varepsilon}^{2}\right)$, where the value of σ_{ε}^{2} is unknown. Estimate β_{ls} and α_{ls} for this sample, and use these to estimate σ_{ε}^{2}, $\widehat{\sigma}_{\varepsilon}^{2}$. Then assume β_{ls}, α_{ls} and $\widehat{\sigma}_{\varepsilon}^{2}$ are the population parameters; that is, your suedo data-generating process is $y_i = \alpha_{ls} + \beta_{ls}x_i + \varepsilon_i$ where $\varepsilon\tilde{}S\left(0,\widehat{\sigma}_{\varepsilon}^{2}\right)$. For the vector of x, $x_1, x_2, ...x_i, ..., x_n$ generate S different random samples of size n based on the suedo data-generating process; make S a large number. For each sample s, estimate β_{ls}^{s} and α_{ls}^{s}, and then use them to estimate ${}^{s}\widehat{\sigma}_{\varepsilon}^{2}$ Plot the distribution of the S β_{ls}^{s}, the distribution of the S α_{ls}^{s}, and the distribution of the S ${}^{s}\widehat{\sigma}_{\varepsilon}^{2}$. The first is an estimated sampling distribution for β_{ls}, centered on β_{ls}, the second is an estimated sampling distribution for α_{ls}, centered on α_{ls}, and the third is the sampling distribution of $\widehat{\sigma}_{\varepsilon}^{2}$, centered on $\widehat{\sigma}_{\varepsilon}^{2}$. A 95% confidence for each can be estimated by lopping off the top and bottom 2.5% of each of these estimated distributions.

Note, these estimated confidence intervals are a function of the initial random sample from your population, the assumption that one has a LRM, the assumption that $\varepsilon\tilde{}S\left(0,\sigma_{\varepsilon}^{2}\right)$ and n: it is definitely a function of the Snerd assumption. It is not a function of the value of σ_{ε}^{2}. The larger n the shorter these confidence intervals.

Note, one does not need to theoretically derive either $f(\frac{\beta_{ls}-\beta}{\widehat{\sigma}_{\beta_{ls}}})$ or $f(\beta_{ls})$: the latter was derived by simulation.