14.381: Statistical Methods in Economics

OLS Basics

Michael Powell
Department of Economics, MIT
September 18th, 2006

The purpose of these notes is to provide a brief introduction and overview of the basics of the linear regression model.

1 Two Vector Calculus Results

There are two results from vector calculus which are useful in the derivation of ordinary least squares. Let $A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$ be a square $n \times n$ matrix and $t = \begin{bmatrix} t_1 \\ \vdots \\ t_n \end{bmatrix}$ be an $n \times 1$ vector. The first result involves taking the derivative of the vector $t'A$ with respect to the vector t. What does it mean to take the derivative of an expression with respect to a vector? Without going into too much of the formalities associated with creating precise definitions, the derivative of an expression with respect to a vector is equal to the vector of the derivatives of the expression with respect to the individual components. In establishing the first proposition, one will see how to make this "definition" operational.

Proposition 1 *Let A and t be defined as above. Then* $\frac{\partial t'A}{\partial t} = A$.

Proof. Here, we have

$$t'A = \begin{bmatrix} t_1 & \cdots & t_n \end{bmatrix} \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n t_i a_{i1} & \cdots & \sum_{i=1}^n t_i a_{in} \end{bmatrix}$$

The derivative with respect to t, then, is given by

$$\begin{aligned} \frac{\partial t'A}{\partial t} &= \frac{\partial \begin{bmatrix} \sum_{i=1}^n t_i a_{i1} & \cdots & \sum_{i=1}^n t_i a_{in} \end{bmatrix}}{\partial \begin{bmatrix} t_1 \\ \vdots \\ t_n \end{bmatrix}} = \begin{bmatrix} \frac{\partial\left(\sum_{i=1}^n t_i a_{i1}\right)}{\partial t_1} & \cdots & \frac{\partial\left(\sum_{i=1}^n t_i a_{in}\right)}{\partial t_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial\left(\sum_{i=1}^n t_i a_{i1}\right)}{\partial t_n} & \cdots & \frac{\partial\left(\sum_{i=1}^n t_i a_{in}\right)}{\partial t_n} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial(t_1 a_{11} + \cdots + t_n a_{n1})}{\partial t_1} & \cdots & \frac{\partial(t_1 a_{1n} + \cdots + t_n a_{nn})}{\partial t_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial(t_1 a_{11} + \cdots + t_n a_{n1})}{\partial t_n} & \cdots & \frac{\partial(t_1 a_{1n} + \cdots + t_n a_{nn})}{\partial t_n} \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} = A \end{aligned}$$

Which is what we sought to prove. ■

Proposition 2 *Let A and t be defined as above. Then* $\frac{\partial t'At}{\partial t} = (A + A')\,t$.

Proof. Here,

$$\begin{aligned} t'At &= \begin{bmatrix} t_1 & \cdots & t_n \end{bmatrix} \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} t_1 \\ \vdots \\ t_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n t_i a_{i1} & \cdots & \sum_{i=1}^n t_i a_{in} \end{bmatrix} \begin{bmatrix} t_1 \\ \vdots \\ t_n \end{bmatrix} \\ &= t_1 \left(\sum_{i=1}^n t_i a_{i1} \right) + \cdots + t_n \left(\sum_{i=1}^n t_i a_{in} \right) = \left(t_1^2 a_{11} + \cdots + t_1 t_n a_{n1} \right) + \cdots + \left(t_n t_1 a_{1n} + \cdots + t_n^2 a_{nn} \right) \end{aligned}$$

The derivative with respect to t is

$$
\begin{aligned}
\frac{\partial t'At}{\partial t} &= \frac{\partial\left[\left(t_1^2 a_{11}+\cdots+t_1 t_n a_{n1}\right)+\cdots+\left(t_n t_1 a_{1n}+\cdots+t_n^2 a_{nn}\right)\right]}{\partial\begin{bmatrix} t_1 \\ \vdots \\ t_n \end{bmatrix}} \\
&= \begin{bmatrix} \frac{\partial\left[\left(t_1^2 a_{11}+\cdots+t_1 t_n a_{n1}\right)+\cdots+\left(t_n t_1 a_{1n}+\cdots+t_n^2 a_{nn}\right)\right]}{\partial t_1} \\ \vdots \\ \frac{\partial\left[\left(t_1^2 a_{11}+\cdots+t_1 t_n a_{n1}\right)+\cdots+\left(t_n t_1 a_{1n}+\cdots+t_n^2 a_{nn}\right)\right]}{\partial t_n} \end{bmatrix} \\
&= \begin{bmatrix} \left(t_1 a_{11}+t_1 a_{11}+t_2 a_{21}+\cdots+t_n a_{n1}\right)+t_2 a_{12}+\cdots+t_n a_{1n} \\ \vdots \\ t_1 a_{n1}+\cdots+t_{n-1}a_{n,n-1}+\left(t_1 a_{1n}+\cdots+t_{n-1}a_{n-1,n}+t_n a_{nn}+t_n a_{nn}\right) \end{bmatrix} \\
&= \begin{bmatrix} \left(a_{11}+a_{11}\right)t_1+\left(a_{21}+a_{12}\right)t_2+\cdots+\left(a_{n1}+a_{1n}\right)t_n \\ \vdots \\ \left(a_{n1}+a_{1n}\right)t_1+\left(a_{n2}+a_{2n}\right)t_2+\cdots+\left(a_{nn}+a_{nn}\right)t_n \end{bmatrix} \\
&= \begin{bmatrix} a_{11}t_1+a_{12}t_2+\cdots+a_{1n}t_n \\ \vdots \\ a_{n1}t_1+a_{n2}t_2+\cdots+a_{nn}t_n \end{bmatrix} + \begin{bmatrix} a_{11}t_1+a_{21}t_2+\cdots+a_{n1}t_n \\ \vdots \\ a_{1n}t_1+a_{2n}t_2+\cdots+a_{nn}t_n \end{bmatrix} \\
&= At+A't=\left(A+A'\right)t
\end{aligned}
$$

And we have established the proposition. ■

2 The Basic Model

In undergraduate econometrics, we typically assume that there is some linear relationship between the left-hand variable (dependent variable) Y_i and the right-hand variables (independent variables) $X_{1i}, X_{2i}, \ldots, X_{ki}$. That is, we assume that

$$Y_i = X_{1i}\beta_1 + X_{2i}\beta_2 + \cdots + X_{ki}\beta_{ki} + \varepsilon_i,\ i = 1, 2, \ldots, n$$

Where ε_i is some disturbance term that captures the idea that there are probably many other variables that go into determining Y_i. The standard presentation in a basic econometrics class supposes that $k = 2$ and that $X_{1i} = 1$ for all i. That is,

$$Y_i = \beta_1 + X_{2i}\beta_2 + \varepsilon_i,\ i = 1, 2, \ldots, n$$

The fledgling researcher who is interested in estimating β_1 and β_2 will use OLS to derive some $\hat{\beta}_1$ and $\hat{\beta}_2$ that generate a line $\hat{Y}_i = \hat{\beta}_1 + X_{2i}\hat{\beta}_2$ which "best" fits the data. By "best," one means, of course, that $\hat{\beta}_1$ and $\hat{\beta}_2$ minimize the sum of the squared distances between the predicted values and the actual values. That is, $\hat{\beta}_1$ and $\hat{\beta}_2$ are chosen to

$$\min_{b_1,b_2}\sum_{i=1}^{n}\left(Y_i-\hat{Y}_i\right)^2 = \min_{b_1,b_2}\sum_{i=1}^{n}\left(Y_i - b_1 - X_{2i}b_2\right)^2$$

The familiar solution to this problem from undergraduate econometrics is

$$
\begin{aligned}
\hat{\beta}_1 &= \bar{Y} - \bar{X}_2\hat{\beta}_2 \\
\hat{\beta}_2 &= \frac{\sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)\left(X_{2i}-\bar{X}_2\right)}{\sum_{i=1}^{n}\left(X_{2i}-\bar{X}_2\right)^2}
\end{aligned}
$$

Where $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ and $\bar{X}_2 = \frac{1}{n}\sum_{i=1}^{n} X_{2i}$ are the sample mean, respectively, of the $Y_i's$ and the $X_{2i}'s$.

However, what does one do if he/she wants to include more than one right-hand variable in the regression? For example, a basic labor economics model trying to explain wages will posit that

$$Wage_i = \beta_1 + Education_i \cdot \beta_2 + Experience_i \cdot \beta_3 + \varepsilon_i$$

In order to do so, we will make use of the language of linear algebra to express compactly all the relevant information in such a model.

3 Matrix Representation of the Basic Model

Suppose we have that

$$Y_i = \beta_1 + X_{2i}\beta_2 + \cdots + X_{ki}\beta_k + \varepsilon_i,\ i = 1, \ldots, n$$

For each equation we can express this relationship using vector notation by defining the vectors $X_i = \begin{bmatrix} 1 \\ X_{2i} \\ \vdots \\ X_{ki} \end{bmatrix}$ and $\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$. Then,

$$\begin{aligned} Y_i &= X_i'\beta + \varepsilon_i = \begin{bmatrix} 1 & X_{2i} & \cdots & X_{ki} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \varepsilon_i,\ i = 1, \ldots, n \\ &= \beta_1 + X_{2i}\beta_2 + \cdots + X_{ki}\beta_k + \varepsilon_i,\ i = 1, \ldots, n \end{aligned}$$

Since this is actually a collection of n equations, we can express this model in an even more compact form.

$$\begin{gathered} Y_1 = X_1'\beta + \varepsilon_1 \\ \vdots \\ Y_n = X_n'\beta + \varepsilon_n \end{gathered}$$

Can be written as

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} X_1'\beta + \varepsilon_1 \\ \vdots \\ X_n'\beta + \varepsilon_n \end{bmatrix} = \begin{bmatrix} X_1'\beta \\ \vdots \\ X_n'\beta \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} = \begin{bmatrix} X_1' \\ \vdots \\ X_n' \end{bmatrix}\beta + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Define the matrices $Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$, $X = \begin{bmatrix} X_1' \\ \vdots \\ X_{n'} \end{bmatrix}$, and $\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$. Then this model can be expressed as

$$Y = X\beta + \varepsilon$$

4 Ordinary Least Squares Estimation

Without providing much motivation for why ordinary least squares is an important starting point for estimating the parameters $\beta_1, \ldots, \beta_k$, I will jump into the analysis. Assuming it is a good idea to minimize the sum of the squared distances between observations of Y_i and their predicted values, $\hat{Y}_i = X_i\hat{\beta}$, we have that we want to choose $\hat{\beta}$ to

$$\min_b \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2 = \min_b \sum_{i=1}^{n} \left(Y_i - X_i b\right)^2$$

Note that we can write this equivalently as

$$\min_b (Y - Xb)' (Y - Xb) = \min_b \begin{bmatrix} Y_1 - X_1 b & \cdots & Y_n - X_n b \end{bmatrix} \begin{bmatrix} Y_1 - X_1 b \\ \vdots \\ Y_n - X_n b \end{bmatrix}$$

Thus, our problem becomes choosing $\hat{\beta}$ to

$$\begin{aligned} \min_b (Y - Xb)' (Y - Xb) &= \min_b (Y' - b'X') (Y - Xb) \\ &= \min_b (Y'Y - Y'Xb - b'X'y + b'X'Xb) \end{aligned}$$

Since Y' is $1 \times n$, X is $n \times k$, and b is $k \times 1$, we have that $Y'Xb$ is 1×1. A scalar is equal to its transpose, so $Y'Xb = (Y'Xb)' = b'X'Y$. Thus, we want to

$$\min_b (Y'Y - 2b'X'y + b'X'Xb)$$

Taking first order conditions with respect to the vector b, using the results from section 1, gives us

$$(b) : -2X'y + \left(X'X + (X'X)'\right) \hat{\beta} = 0$$

Since $(X'X)' = (X')' X' = X'X$, we have that $X'X + (X'X)' = 2X'X$ and thus

$$\begin{aligned} 2X'X\hat{\beta} &= 2X'y \\ X'X\hat{\beta} &= X'y \end{aligned}$$

Assume that $X'X$, which is $k \times k$, is invertible. (This can be shown to be equivalent to assuming that there is no perfect multicollinearity.) Then we can premultiply both sides of this equation by $(X'X)^{-1}$ to get

$$\hat{\beta} = (X'X)^{-1} X'y$$

This is a very important result in basic econometrics. It will be pounded into your head time and time again, though, so don't worry if it is difficult to remember.

5 Distribution of $\hat{\beta}$ Under Normality Assumptions

In order to perform statistical inference (i.e. constructing confidence intervals or testing hypotheses) using $\hat{\beta}$, we need to make some assumptions about the model with which we are working.

Linearity: $Y_i = X_i'\beta + \varepsilon_i, \ i = 1, \ldots, n$
Zero Mean: $E[\varepsilon_i] = 0, \ i = 1, \ldots, n$
Homoskedasticity: $Var(\varepsilon_i) = E[\varepsilon_i^2] = \sigma^2, \ i = 1, \ldots, n$
No Serial Correlation: $E[\varepsilon_i \varepsilon_j] = 0$ for all $i \neq j$
Normality: $\varepsilon_i \sim N(0, \sigma^2)$
No Perfect Multicollinearity: $rank(X) = k$

It is useful to remark here that a quick glance at a graduate level econometrics textbook will reveal that these assumptions are made using the notation of the conditional expectations. That is, there might be something that looks like $E[\varepsilon_i | X_i] = 0$ or $E[\varepsilon_i \varepsilon_j | X] = 0$. For now, we will ignore this. In order to do so (and still get somewhere), though, we need to make some assumptions about the nature of the independent variables, the $X_i's$. For what follows, we will assume that the $X_i's$ are non-random. However, we will assume this with the caveat that this is a bad assumption to make in the social sciences, where the independent variables are typically random. In the hard sciences, controlled experimentation allows the researcher to, well, control the values of X_i that are used. In this case, these $X_i's$ are non-random. However, in economics, one typically does not have such control over the values of the independent variables which are observed. (A labor economist cannot assign a level of schooling to a particular individual just to see how it affects his/her wages!)

Using matrix notation, one can compress this list of six assumptions into only four

Linear Expectations: $E[Y] = X\beta$
Spherical Errors: $Var(\varepsilon) = \sigma^2 I_n$
Normality: $\varepsilon \sim N(0, \sigma^2 I_n)$
No Perfect Multicollinearity: $rank(X) = k$

To see why these three conditions are implied by the previous assumptions, suppose $E[\varepsilon] = 0$. Then

$$E[Y] = E[X\beta + \varepsilon] = X\beta + E[\varepsilon] = X\beta$$

In addition,

$$\begin{aligned}
Var(\varepsilon) &= E[\varepsilon\varepsilon'] = E\left[\begin{bmatrix}\varepsilon_1 \\ \vdots \\ \varepsilon_n\end{bmatrix}\begin{bmatrix}\varepsilon_1 & \cdots & \varepsilon_n\end{bmatrix}\right] \\
&= E\begin{bmatrix}\varepsilon_1^2 & \cdots & \varepsilon_1\varepsilon_n \\ \vdots & \ddots & \vdots \\ \varepsilon_n\varepsilon_1 & \cdots & \varepsilon_n^2\end{bmatrix} = \begin{bmatrix}E\left[\varepsilon_1^2\right] & \cdots & E[\varepsilon_1\varepsilon_n] \\ \vdots & \ddots & \vdots \\ E[\varepsilon_n\varepsilon_1] & \cdots & E\left[\varepsilon_n^2\right]\end{bmatrix} \\
&= \begin{bmatrix}\sigma^2 & \cdots & 0 \\ \vdots & \ddots & 0 \\ 0 & \cdots & \sigma^2\end{bmatrix} = \sigma^2\begin{bmatrix}1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1\end{bmatrix} = \sigma^2 I_n
\end{aligned}$$

Given these assumptions, one can show a few characteristics of the distribution of the OLS estimator.

Proposition 3 $E\left[\hat{\beta}\right] = \beta$. *That is, $\hat{\beta}$ is unbiased.*

Proof. Since $\hat{\beta} = (X'X)^{-1} X'Y$, if we take expectations,

$$\begin{aligned}
E\left[\hat{\beta}\right] &= E\left[(X'X)^{-1} X'Y\right] = (X'X)^{-1} X'E[Y] \\
&= (X'X)^{-1} X'X\beta = \beta
\end{aligned}$$

Which is the desired result. ■

Proposition 4 $Var\left(\hat{\beta}\right) = \sigma^2 (X'X)^{-1}$

Proof. First note that

$$\begin{aligned}
Var\left(\hat{\beta}\right) &= Var\left((X'X)^{-1} X'Y\right) \\
&= (X'X)^{-1} X'Var(Y) X (X'X)^{-1}
\end{aligned}$$

Since $Y = X\beta + \varepsilon$ and X and β are nonrandom, we have that $Var(Y) = Var(\varepsilon) = \sigma^2 I_n$. Thus,

$$\begin{aligned}
Var\left(\hat{\beta}\right) &= (X'X)^{-1} X'\sigma^2 I_n X (X'X)^{-1} \\
&= \sigma^2 (X'X)^{-1} X'X (X'X)^{-1} \\
&= \sigma^2 (X'X)^{-1}
\end{aligned}$$

Which establishes the proposition. ■

It turns out that $\hat{\beta}$ is also normally distributed. Intuitively, this makes sense, since $\hat{\beta}$ is just a linear transformation of ε, which we have assumed to be normally distributed. The following proposition will help in establishing this fact.

Proposition 5 *Let $Z \sim N(\mu, \Sigma)$ be a multivariate normal random variable. Then $CZ \sim N(C\mu, C\Sigma C')$, where C is a non-random matrix of conformable dimension.*

Proof. The proof of this proposition follows by the use of moment generating functions. ■

Proposition 6 $\hat{\beta} \sim N\left(\beta, \sigma^2 (X'X)^{-1}\right)$.

Proof. Since $\hat{\beta} = (X'X)^{-1} X'Y = (X'X)^{-1} X'(X\beta + \varepsilon) = \beta + (X'X)^{-1} X'\varepsilon$, we have that

$$\hat{\beta} - \beta = (X'X)^{-1} X'\varepsilon$$

Since $\varepsilon \sim N\left(0, \sigma^2 I_n\right)$, we have that by the previous proposition,

$$\begin{aligned}\hat{\beta} - \beta &\sim N\left(0, (X'X)^{-1} X'\sigma^2 X (X'X)^{-1}\right) \\ &\stackrel{d}{=} N\left(0, \sigma^2 (X'X)^{-1}\right)\end{aligned}$$

A quick moment generating function argument will then give us that

$$\hat{\beta} \sim N\left(\beta, \sigma^2 (X'X)^{-1}\right)$$

Which is what we set out to prove. ■

6 Confidence Intervals

Let $\Gamma_i = \begin{bmatrix} 0 & \cdots & 1 & \cdots & 0 \end{bmatrix}$ be the row vector with a 1 in the ith position and zeros elsewhere. Given that $\hat{\beta} \sim N\left(\beta, \sigma^2 (X'X)^{-1}\right)$, it follows from proposition 5 that

$$\Gamma_i \hat{\beta} \sim N\left(\Gamma_i \beta, \Gamma_i \sigma^2 (X'X)^{-1} \Gamma_i'\right),$$

but $\Gamma_i \hat{\beta} = \begin{bmatrix} 0 & \cdots & 1 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_n \end{bmatrix} = \hat{\beta}_i$ and $\Gamma_i \beta = \begin{bmatrix} 0 & \cdots & 1 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} = \beta_i$. Thus,

$$\hat{\beta}_i \sim N\left(\beta_i, \sigma^2 \Gamma_i (X'X)^{-1} \Gamma_i'\right)$$

By a simple moment generating function argument, one can show that

$$\hat{\beta}_i - \beta_i \sim N\left(0, \sigma^2 \Gamma_i (X'X)^{-1} \Gamma_i'\right)$$

And

$$\frac{\hat{\beta}_i - \beta_i}{\sqrt{\sigma^2 \Gamma_i (X'X)^{-1} \Gamma_i}} \sim N(0,1)$$

One can then use $\frac{\hat{\beta}_i-\beta_i}{\sqrt{\sigma^2\Gamma_i(X'X)^{-1}\Gamma_i}}$ to construct confidence intervals for β_i. For instance,

$$\begin{aligned}
0.95 &= \Pr\left[\left|\frac{\hat{\beta}_i-\beta_i}{\sqrt{\sigma^2\Gamma_i\left(X'X\right)^{-1}\Gamma_i}}\right|\leq 1.96\right] \\
&= \Pr\left[\left|\hat{\beta}_i-\beta_i\right|\leq 1.96\sigma\sqrt{\Gamma_i\left(X'X\right)^{-1}\Gamma_i}\right] \\
&= \Pr\left[\hat{\beta}_i-1.96\sigma\sqrt{\Gamma_i\left(X'X\right)^{-1}\Gamma_i}\leq\beta_i\leq\hat{\beta}_i+1.96\sigma\sqrt{\Gamma_i\left(X'X\right)^{-1}\Gamma_i}\right]
\end{aligned}$$

That is, $\left(\hat{\beta}_i-1.96\sigma\sqrt{\Gamma_i\left(X'X\right)^{-1}\Gamma_i},\hat{\beta}_i+1.96\sigma\sqrt{\Gamma_i\left(X'X\right)^{-1}\Gamma_i}\right)$ constitutes a 95% confidence interval for β_i. However, this quantity is only well-defined when σ^2 is known. If σ^2 is unknown, we typically estimate it using

$$\hat{\sigma}^2=\frac{1}{n-k}\sum_{i=1}^{n}\left(Y_i-X_i\hat{\beta}\right)^2=\frac{1}{n-k}\left(Y-X\hat{\beta}\right)'\left(Y-X\hat{\beta}\right)=\frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-k}$$

Where $\hat{\varepsilon}$ is defined to be the vector of OLS residuals: $\hat{\varepsilon}=Y-\hat{Y}=Y-X\hat{\beta}$. When we estimate σ^2 using $\hat{\sigma}^2$, it can be shown that

$$\frac{\hat{\beta}_i-\beta_i}{\sqrt{\hat{\sigma}^2\Gamma_i\left(X'X\right)^{-1}\Gamma_i}}\sim t\left(n-k\right)$$

We refer to the quantity $\sqrt{\hat{\sigma}\Gamma_i\left(X'X\right)^{-1}\Gamma_i}\equiv se\left(\hat{\beta}_i\right)$ as the standard error for $\hat{\beta}_i$. One can then similarly construct a confidence interval for β_i when $\hat{\sigma}^2$ is unknown.

$$\begin{aligned}
0.95 &= \Pr\left[\left|\frac{\hat{\beta}_i-\beta_i}{se\left(\hat{\beta}_i\right)}\right|\leq t_{0.95,t(n-k)}\right] \\
&= \Pr\left[\left|\hat{\beta}_i-\beta_i\right|\leq t_{0.95,t(n-k)}se\left(\hat{\beta}_i\right)\right] \\
&= \Pr\left[\hat{\beta}_i-t_{0.95,t(n-k)}se\left(\hat{\beta}_i\right)\leq\beta_i\leq\hat{\beta}_i+t_{0.95,t(n-k)}se\left(\hat{\beta}_i\right)\right]
\end{aligned}$$

Where $t_{0.975,t(n-k)}$ is the 97.5% critical value of the t distribution. Thus, our 95% confidence interval is given by $\left(\hat{\beta}_i-t_{0.95,t(n-k)}se\left(\hat{\beta}_i\right),\hat{\beta}_i+t_{0.95,t(n-k)}se\left(\hat{\beta}_i\right)\right)$.