Problem Set 4

Artschil Okropiridse*

February 23, 2024

I try to be persistent, but can't promise no typos (check how its printed in Hanson when you get confused): When I write $\mathbf{X}$, I mean sample, thus a $n \times k$ matrix. When I write X, I mean population, thus a $k \times 1$ vector. Feel free to skip section 1 "General Procedures", I just added this after some question came about the topics more generally.

1 General Procedures

1.1 Asymptotics

For asymptotics we use a couple of theorems, those are mainly the Weak Law of Large Numbers (WLLN) and the Central Limit theorem (CLT). Then we have a few theorems that tell us how we can apply both WLLN and CLT through other functions. Those theorems are the Continuous Mapping theorem (CMT) and the Delta-method (DM). These theorems we just have to know (for the exam) or know how to look them up quickly (the rest of your life). If you have time it is obviously much better to go through the proofs of why they work and understand them instead of just remembering them. The book Mitch send does a good job at explaining those proofs. More formally, you will find them in any rigorous statistics text book.

Once we have the theorems at hand the trick is usually to take what ever is thrown at us and turn it into something more familiar such that we can apply the theorems.

So as an example, to show what OLS converges to we use:

*many thanks to Jakob Beuschlein

We can show consistency of the OLS estimator by using:

$$\begin{aligned}
\hat{\beta} &= (\mathbf{X'X})^{-1}(\mathbf{X'Y}) \\
&= \left(\sum^{n} X_i X_i'\right)^{-1} \left(\sum^{n} X_i Y_i\right) \\
&= \left(\frac{1}{n}\sum^{n} X_i X_i'\right)^{-1} \left(\frac{1}{n}\sum^{n} X_i Y_i\right) \\
&\equiv \hat{Q}_{XX}^{-1}\hat{Q}_{XY}
\end{aligned}$$

Note that both $\hat{Q}$'s are just sample means. So we know we can apply the WLLN (assuming positive definiteness, finite 2nd moments and i.i.d. observations):

$$\begin{aligned}
\hat{Q}_{XX} &\to^{p} Q_{XX} \equiv \mathbb{E}\left[XX'\right] \\
\hat{Q}_{XY} &\to^{p} Q_{XX} \equiv \mathbb{E}\left[XY\right]
\end{aligned}$$

Recall that the WLLN goes through functions, so if it holds for $\hat{Q}$ it also holds for $\hat{Q}^{-1}$. We have now shown that each separate part of the OLS estimator converges to where we want it to converge to. To be able to apply both these terms we need the continuous mapping theorem.

So let's define a function $g(a, b) = a^{-1} * b$ and write:

$$\hat{\beta} = g(\hat{Q}_{xx}, \hat{Q}_{XY})$$

So we have a function of two things that converge by the WLLN to their respective population counterparts. The CMT tells us (given some assumptions) that this function will converge to the function of those population things, i.e.:

$$\hat{\beta} = g(\hat{Q}_{XX}, \hat{Q}_{XY}) \to^{P} g(Q_{XX}, Q_{XY})$$

To show asymptotic normality of the OLS estimator we can do the following:

$$\begin{aligned}
\sqrt{n}(\hat{\beta} - \beta) &= \sqrt{n}((\mathbf{X'X})^{-1}(\mathbf{X'Y})) - \beta) \\
&= \sqrt{n}((\mathbf{X'X})^{-1}(\mathbf{X'}(\mathbf{X}\beta + e))) - \beta) \\
&= \sqrt{n}((\mathbf{X'X})^{-1}(\mathbf{X'X}\beta) + (\mathbf{X'X})^{-1}(\mathbf{X'}e) - \beta) \\
&= \sqrt{n}(\beta + (\mathbf{X'X})^{-1}(\mathbf{X'}e) - \beta) \\
&= \sqrt{n}(\left(\frac{1}{n}\sum^{n} XX'\right)^{-1} \left(\frac{1}{n}\sum^{n} Xe\right)) \\
&\equiv \sqrt{n}(\hat{Q}_{XX}^{-1}\hat{Q}_{Xe})
\end{aligned}$$

Now let's looks just at $\sqrt{n}\hat{Q}_{Xe}$. Note that the population equivalent is $\mathbb{E}\left[Xe\right]=0$. So we can write $\sqrt{n}\hat{Q}_{Xe}=\sqrt{n}(\hat{Q}_{Xe}-\mathbb{E}\left[Xe\right])$. This already looks familiar. We can apply the CLT:

$$
\begin{aligned}
\sqrt{n}(\hat{Q}_{Xe}-\mathbb{E}\left[Xe\right]) &= \sqrt{n}(\frac{1}{n}\sum^{n}Xe-\mathbb{E}\left[Xe\right]) \rightarrow^{d} \mathcal{N}\left(0,\Omega\right), \\
\text{where } \Omega &= \mathbb{E}\left[(Xe-\mathbb{E}\left[Xe\right])(Xe-\mathbb{E}\left[Xe\right])'\right] \\
&= \mathbb{E}\left[(Xe)(Xe)'\right] \\
&= \mathbb{E}\left[XX'e^2\right]
\end{aligned}
$$

So we know $\sqrt{n}\hat{Q}_{Xe}\rightarrow^{d}\mathcal{N}\left(0,\Omega\right)$. Now we can re-write:

$$
\begin{aligned}
\sqrt{n}(\hat{\beta}-\beta) &= \sqrt{n}(\hat{Q}_{XX}^{-1}\hat{Q}_{Xe}) \\
&= \sqrt{n}g(\hat{Q}_{Xe}), \\
\text{where we define } \mathbf{G} &\equiv \frac{\partial g(Q_{Xe})}{\partial Q_{Xe}} \\
&= Q_{XX}^{-1}
\end{aligned}
$$

So we can use the Delta method to show:

$$
\begin{aligned}
\sqrt{n}(g(\hat{Q}_{Xe})-g(Q_{Xe})) &\rightarrow^{d} \mathbf{G}' * \mathcal{N}\left(0,\Omega\right) \\
&= \mathcal{N}\left(0,\mathbf{G}'\Omega\mathbf{G}\right) \\
&= \mathcal{N}\left(0,\mathbb{E}\left[XX'\right]^{-1}\mathbb{E}\left[XX'e^2\right]\mathbb{E}\left[XX'\right]^{-1}\right)
\end{aligned}
$$

It is of course not necessary to redefine the OLS estimator as the $g()$ function. I am just doing this so it looks similar to the Delta-method from the books, and it becomes clearer why we can use it.

As a transition to the next topic let's derive the t-statistic. Let's define

$$
\mathbb{E}\left[XX'\right]^{-1}\mathbb{E}\left[XX'e^2\right]\mathbb{E}\left[XX'\right]^{-1} \equiv \mathbf{V}_{\beta}
$$

So we have $\sqrt{n}(\hat{\beta}-\beta)\rightarrow^{d}\mathcal{N}\left(0,\mathbf{V}_{\beta}\right)$. Let's assume there is only one coefficient, i.e. $\beta\in\mathbb{R}$.

Some algebra:

$$\sqrt{n}(\hat{\beta}-\beta) \to^d \mathcal{N}(0, \mathbf{V}_\beta)$$
$$\frac{\sqrt{n}(\hat{\beta}-\beta)}{\sqrt{\hat{\mathbf{V}}_\beta}} \to^d \frac{\mathcal{N}(0, \mathbf{V}_\beta)}{\sqrt{\mathbf{V}_\beta}}$$
$$T - stat \equiv \frac{\sqrt{n}(\hat{\beta}-\beta)}{\sqrt{\hat{\mathbf{V}}_\beta}} \to^d \mathcal{N}(0,1)$$
$$\frac{(\hat{\beta}-\beta)}{\sqrt{\hat{\mathbf{V}}_{\hat{\beta}}}} \to^d \mathcal{N}(0,1)$$

To see why we can just do algebra with the RHS see below. The last step uses the exact instead of the asymptotic variance (see below what this means and Hanson 7.8 for clarification). The important point here is that **hypothesis testing builds on the distributions we can derive with asymptotic theory** (for a more formal derivation check Hanson 7.12).

Exact vs. Asymptotic variance

In the seminar there was some confusion about the exact vs the asymptotic variance. This is wholly justified, this is confusing stuff. Above I showed how to derive the asymptotic variance. So let's stick to Hanson's notation and write:

$$\begin{aligned} avar(\sqrt{n}(\hat{\beta}-\beta)) &= \mathbf{V}_\beta \\ &= \mathbb{E}[XX']^{-1}\,\mathbb{E}\left[XX'e^2\right]\mathbb{E}[XX']^{-1} \\ &\equiv Q_{XX}^{-1}\Omega Q_{XX}^{-1} \end{aligned}$$

Note that none of these terms are observed. So we use

$$\hat{Q}_{XX}^{-1} = \frac{1}{n}\sum_{i=1}^{n} X_i X_i'$$
$$\hat{\Omega} = \frac{1}{n}\sum_{i=1}^{n} X_i X_i' \hat{e}_i^2$$

to estimate those to get $\hat{\mathbf{V}}_\beta$.

And for the exact variance we do:

$$
\begin{aligned}
\mathrm{Var}\left[\hat{\beta}|\mathbf{X}\right] &= \mathbf{V}_{\hat{\beta}} \\
&= \mathbb{E}\left[\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} - \mathbb{E}\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\right|\mathbf{X}]\right)^2\right|\mathbf{X}\right] \\
&= \mathbb{E}\left[\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\mathbf{Y} - \mathbb{E}\left[\mathbf{Y}\right|\mathbf{X}]\right)\right)^2\right|\mathbf{X}\right] \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbb{E}\left[\mathbf{e}\mathbf{e}'\right|\mathbf{X}]\,\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&\equiv (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}
\end{aligned}
$$

But the $\mathbf{D}$ matrix is unobserved (same issue as above). So we have to estimate it somehow, in the case of heteroskedasticity we can use $\sum_n X_i X_i' \hat{e}_i^2$ as an estimator for $\mathbf{XDX}$. This gives us $\hat{\mathbf{V}}_{\hat{\beta}}$.

So both exact and asymptotic variance are very similar in essence, but we have:

$$
\begin{aligned}
\mathbf{V}_{\beta} &= n\mathbf{V}_{\hat{\beta}} \\
\hat{\mathbf{V}}_{\beta} &= n\hat{\mathbf{V}}_{\hat{\beta}}
\end{aligned}
$$

So the asymptotic variance is always larger than the exact variance (by factor n).

Algebra with the RHS

In the seminar a question came up about why we can just divide $\mathcal{N}()$ by something. For me, the intuition follows from thinking of this as some random variable with that distribution, i.e.:

$$
\begin{aligned}
Z &= \mathcal{N}\left(0, \mathbf{V}_{\beta}\right) \\
\mathbb{E}\left[Z\right] &= 0 \\
\mathrm{Var}\left[Z\right] &= \mathbb{E}\left[(Z - \mathbb{E}\left[Z\right])^2\right] = \mathbb{E}\left[Z^2\right] = \mathbf{V}_{\beta} \\
\implies \mathbb{E}\left[\frac{Z}{\sqrt{\mathbf{V}_{\beta}}}\right] &= (\sqrt{\mathbf{V}_{\beta}})^{-1}\mathbb{E}\left[Z\right] = 0 \\
\implies \mathrm{Var}\left[\frac{Z}{\sqrt{\mathbf{V}_{\beta}}}\right] &= \mathbb{E}\left[\left(\frac{Z}{\sqrt{\mathbf{V}_{\beta}}} - 0\right)^2\right] \\
&= ((\sqrt{\mathbf{V}_{\beta}})^2)^{-1}\mathbb{E}\left[Z^2\right] \\
&= \mathbf{V}_{\beta}^{-1}\mathbf{V}_{\beta} = 1
\end{aligned}
$$

BUT this is just how I justify this to myself. This might not be the perfectly formal way of doing it. So if you want something more rigorous, you will have to look at a textbook.

1.2 Hypothesis testing

The cook-book for hypothesis testing is the following. Given some estimator $\hat{\theta}$ and some hypothesized value θ_0 we construct our two hypothesis:

$$H_0 : \hat{\theta} = \theta_0$$
$$H_A : \hat{\theta} \neq \theta_0$$

1. First set a significance level α (we commonly set 0.05, but this is just a convention)

2. Formulate the hypothesis as a restriction, I.e.

$$r(\hat{\theta}) = \theta_0$$
$$\mathbf{R} \equiv \frac{\partial r(\hat{\theta})}{\partial \hat{\theta}}$$

3. Decide on a test statistics T. Here it is important to consider whether the restriction is linear (e.g. $H_0 : \beta_1 = 0 \implies r(\beta_1) = \beta_1 = 0 \implies \mathbf{R} = 1$) and whether it is uni- or multi-dimensional. If it is uni-dimensional and linear we commonly use a T-test, if it is multi dimensional and linear we commonly use a Wald test. If it is non-linear we better look for another statistic.

4. Set asymptotic critical value such that $1 - G(c) = \alpha$, where G is the CDF of what T converges to in distribution, given that the H_0 is true. To clarify this part a bit. All the tests we commonly use are statistics which's asymptotic distributions we know. E.g. for the Wald test we know $W \to^d \chi^2_q$. So that means given our sample and given that H_0 is true, we know that W will have that distribution (if we wouldn't know the distribution we couldn't interpret the test). So when we get a $\hat{W}$ (i.e. the statistic from our sample) we can say something about how likely this particular observation would have been under the null.

5. Calculate the p-value $p = 1 - G(T)$, note that this means p is only an interpretation of your coefficient given the null is true, i.e. it doesn't tell you anything if the null is not true.

6. Reject H_0 if $p < \alpha$, otherwise accept H_0 (or fail to reject).

2 Hypothesis Testing

Exercise 1

(Ch. 9, ex. 9.2) You have two independent samples $(y_{1i}, \mathbf{X}_{1i})$ and $(y_{2i}, \mathbf{X}_{2i})$ both with sample sizes n which satisfy $y_1 = X_1'\beta_1 + e_1$ and $y_2 = X_2'\beta_2 + e_2$, where $E[X_1 e_1] = 0$ and $E[X_2 e_2] = 0$. Let $\hat{\beta}_1$ and $\hat{\beta}_2$ be the OLS estimators of $\beta_1 \in R^k$ and $\beta_2 \in R^k$.

(a) Find the asymptotic distribution of $\sqrt{n}((\hat{\beta}_2 - \hat{\beta}_1) - (\beta_2 - \beta_1))$ as $n \to \infty$.

(b) Find an appropriate test statistic for $H_0 : \beta_2 = \beta_1$.

(c) Find the asymptotic distribution of this statistic under H_0.

Solution

(a) First note that we have to assume that the variables are i.i.d., $E[Y^4] < \infty$, $E||Y^4|| < \infty$ and $E[X_i X_i']$ is positive definite (assumptions 7.2 in Hanson). In the exercises Hanson tends to be a little more hand-wavy so quite often he will not state stuff like this. Most of the time it will be fine if you just impose those assumptions to arrive at a solution.

Looking at the question itself, note that the stated models are structural equations, i.e. something we assume to give an interesting interpretation of the real world. Only given the fact that $E[X_i e_i] = 0$ makes it a Linear Projection model.

We can rewrite the problem as follows:

$$\sqrt{n}\left((\hat{\beta}_2 - \hat{\beta}_1) - (\beta_2 - \beta_1)\right) = \sqrt{n}\left(\hat{\beta}_2 - \beta_2\right) - \sqrt{n}\left(\hat{\beta}_1 - \beta_1\right).$$

For a quick solution: We know that

$$\sqrt{n}\left(\hat{\beta}_1 - \beta_1\right) \to_d N(0, V_{\beta_1})$$
$$\sqrt{n}\left(\hat{\beta}_2 - \beta_2\right) \to_d N(0, V_{\beta_2})$$

as $n \to \infty$.

This implies:

$$\sqrt{n}\left(\hat{\beta}_1 - \beta_1\right) - \sqrt{n}\left(\hat{\beta}_2 - \beta_2\right) \to_d N(0, V_{\beta_2-\beta_1})$$

where $V_{\hat{\beta}_2-\hat{\beta}_1} = V_{\hat{\beta}_2} + V_{\hat{\beta}_1}$. Follows from both samples being independent thus $\hat{\beta}_1$ and $\hat{\beta}_2$ are independent (thus have zero covariance) and we can use $Var[X+Y] = Var[X] + Var[Y] + 2Cov[X,Y]$ (where in this case the last term is zero).

For a more detailed solution:

Note that we can rewrite either term as:

$$\sqrt{n}(\hat{\beta}_i - \beta_i) = (\frac{1}{n}\mathbf{X}_i'\mathbf{X}_i)^{-1}(\sqrt{n}\frac{1}{n}\mathbf{X}_i\mathbf{e}_i),\ i \in \{1,2\}$$

And note that

$$\text{Var}\left[\mathbf{X}_i\mathbf{e}_i\right] = \mathbb{E}[X_iX_i'e_i^2] =^{def} \Omega$$

Using the CLT we can show:

$$\begin{aligned}\sqrt{n}\frac{1}{n}\mathbf{X}_i\mathbf{e}_i &= \sqrt{n}\left(\frac{1}{n}\mathbf{X}'\mathbf{e} - \mathbb{E}[Xe]\right)\\ &\to^d \mathcal{N}\left(0,\Omega\right)\end{aligned}$$

Using the delta method we get:

$$\begin{aligned}\left(\frac{1}{n}\mathbf{X}'\mathbf{X}\right)^{-1}\sqrt{n}\left(\frac{1}{n}\mathbf{X}\mathbf{e}\right) &= \sqrt{n}\left[g(\left(\frac{1}{n}\mathbf{X}\mathbf{e}\right)) - g(\mathbb{E}\left[Xe\right])\right]\\ &\to^d \mathbb{E}\left[XX'\right]^{-1}\mathcal{N}\left(0,\Omega\right)\\ &= \mathcal{N}\left(0, \mathbb{E}\left[XX'\right]^{-1}\Omega\mathbb{E}\left[XX'\right]^{-1}\right)\end{aligned}$$

Where we have used

$$\begin{aligned}\frac{\partial g(\mathbb{E}\left[Xe\right])}{\partial\mathbb{E}\left[Xe\right]} &= \frac{\partial\mathbb{E}\left[XX'\right]^{-1}\mathbb{E}\left[Xe\right]}{\partial\mathbb{E}\left[Xe\right]}\\ &= \mathbb{E}\left[XX'\right]^{-1}\end{aligned}$$

This shows that either term converges as stated above.

You can view the addition of both terms as a new parameter θ and use theorem 7.8 and 7.9 in Hanson. Or just follow the same steps as above with both terms in each of the expressions (it works exactly the same).

(b) Since this is a two-sided test, we can construct the Wald test statistic[1]

We can rephrase $H_0 : \beta_2 - \beta_1 = 0$ and construct:

$$W = (\hat{\beta}_2 - \hat{\beta}_1)'\hat{V}^{-1}_{\beta_2 \hat{-} \beta_1}(\hat{\beta}_2 - \hat{\beta}_1)$$

where $\hat{V}_{\beta_2 \hat{-} \beta_1}$ is an estimator for $V_{\beta_2 \hat{-} \beta_1}$ and takes for example the form (i.e. we allow for Heteroskedasticity):

$$\hat{V}_{\beta_2 \hat{-} \beta_1} = \left(\hat{V}^{HC1}_{\hat{\beta}_2} + \hat{V}^{HC1}_{\hat{\beta}_1}\right)$$

(c) We know that $W \to_d \chi^2_k$ as $n \to \infty$ where k is the dimension of the restriction, i.e. here β_1 and β_2.

Exercise 2

(Ch. 9, ex. 9.7) Take the model $y = X\beta_1 + X^2\beta_2 + e$ with $\mathbb{E}[\mathbf{X}e] = 0$ where y is wage (dollars per hour) and X is age. Describe how you would test the hypothesis that the expected wage for a 40-year old worker is \$20 an hour.

Solution

We want to know if $m(40) = E[Y|X = 40] = 20$. To turn this into a formal hypothesis, we can first define a restriction $r(\hat{\beta}) = 40\hat{\beta}_1 + 40^2\hat{\beta}_2 = 20$. Note that we can define

$$\mathbf{R} = \frac{\partial r(\hat{\beta})}{\partial \hat{\beta}} = \begin{pmatrix} 40 \\ 40^2 \end{pmatrix}$$

So that $H_0 : r(\hat{\beta}) = \mathbf{R}'\hat{\beta} = r(\hat{\beta}) = 20$. Now we can construct a Wald statistic as $\mathbf{R}$ is 2×1 we could have also used a t-test instead. The Wald test is just the square of the t-test in the uni-dimensional case. So tests are equivalent:

[1] note that in a Wald statistic we basically evaluate how significant the distance between parameter and estimate is. So we don't care if the estimate over-shoots or under-shoots. This implies it lends itself well for two-sided testing. Also note that we can't use a T-test as long as $k > 1$.

$$W = \left(\mathbf{R}'\hat{\beta} - 20\right)' \left(\mathbf{R}'\hat{\mathbf{V}}_{\hat{\beta}}\mathbf{R}\right)^{-1} \left(\mathbf{R}'\hat{\beta} - 20\right)$$

Under the null we know that $W \to^{d} \chi_1^2$. Thus, we can calculate a p-value: $p = 1 - G_1(W)$, where $G_1(u) = P[W \leq u]$ is the CDF of the χ_1^2 distribution. Lastly, we decide to reject or fail to reject the null (depending on the statistical significance level we decide on).

Exercise 3

(Ch. 9, ex. 9.10) In Exercise 7.8 you showed that $\sqrt{n}(\hat{\sigma}^2 - \sigma^2) \to_d N(0, V)$ as $n \to \infty$ for some V. Let $\hat{V}$ be an estimator of V.

(a) Using this result construct a t-statistic for $\mathrm{H}_0 : \sigma^2 = 1$ against $\mathrm{H}_1 : \sigma^2 \neq 1$.

(b) Using the Delta Method find the asymptotic distribution of $\sqrt{n}(\hat{\sigma} - \sigma)$.

(c) Use the previous result to construct a t-statistic for $\mathrm{H}_0 : \sigma = 1$ against $\mathrm{H}_1 : \sigma \neq 1$.

(d) Are the null hypothesis in a) and c) the same or are they different? Are the tests in a) and b) the same or are they different? If they are different, describe a context in which the two tests would give contradictory results.

Solution

(a) Note that we have an asymptotic variance given in the question. So we can construct the following T-test:

$$T = \frac{\hat{\sigma}^2 - 1}{\frac{1}{\sqrt{n}}\sqrt{\hat{V}_{asymptot}}} = \frac{\hat{\sigma}^2 - 1}{\sqrt{\hat{V}_{exact}}}$$

Note that there is two common ways of getting a variance estimate. There is the exact variance and the asymptotic variance.

(b) The Delta Method states that if $\sqrt{n}(\hat{\theta} - \theta_0) \to_d N(0, V)$ and g is a continuously differentiable function then $\sqrt{n}(g(\hat{\theta}) - g(\theta_0)) \to_d N(0, (g')^2 V)$ as $n \to \infty$.
Now let

$$g\left(\sigma^2\right) = \sqrt{\sigma^2} = \sigma.$$

Then

$$g'\left(\sigma^2\right)=\frac{1}{2}\sigma^{-1}.$$

Since g is continuous and differentiable the asymptotic distribution of $\sqrt{n}(\hat{\sigma}-\sigma)$ is given by $N(0,\frac{1}{4}\sigma^{-2}V)$.

(c) We can now construct the t-statistic quite similar to before as

$$T_2=\sqrt{n}\frac{\hat{\sigma}-1}{\sqrt{\frac{1}{4}\hat{\sigma}^{-2}\hat{V}}}=\frac{\hat{\sigma}-1}{\frac{1}{\sqrt{n}2\hat{\sigma}}\sqrt{\hat{V}}}$$

(d) Since $\sigma=1$ implies $\sigma^2=1$ and vice versa, these two null hypotheses are equivalent. However, the tests in a) and c) are generally not and can yield different results. This shows that Wald and T tests generally are ill-suited for non-linear hypothesis.

Exercise 4

(Ch. 9, ex. 9.26) In a paper in 1963, Marc Nerlove analyzed a cost function for 145 American electric companies. Nerlov was interested in estimating a cost function: $C=f(Q,PL,PF,PK)$ where the variables are listed in the table below. His data set `Nerlove1963` is on the textbook website.

C	Total Cost
Q	Output
PL	Unit price of labor
PK	Unit price of capital
PF	Unit price of fuel

First, estimate an unrestricted Cobb-Douglas specification

$$\log(C)=\beta_1+\beta_2\log Q+\beta_3\log PL+\beta_4\log PK+\beta_5\log PF+e.$$

Report parameter estimates and standard errors.

(a) What is the economic meaning of the restriction $H_0:\beta_3+\beta_4+\beta_5=1$?

(b) Estimate the model by constrained least squares imposing $\beta_3+\beta_4+\beta_5=1$. Report your parameter estimates and standard errors.

(c) Estimate the model by efficient minimum distance imposing $\beta_3+\beta_4+\beta_5=1$. Report your parameter estimates and standard errors.

(d) Test $H_0:\beta_3+\beta_4+\beta_5=1$ using a Wald statistic.

(e) Test $H_0:\beta_3+\beta_4+\beta_5=1$ using a minimum distance statistic.

Solution

R solution at the end of the document.

Exercise 5

(Ch. 9, ex. 9.28) Using the `cps09mar` dataset and the subsample of non-Hispanic Black individuals (race code = 2) test the hypothesis that marriage status does not affect mean wages.

(a) Take the regression reported in Table 4.2. Which variables will need to be omitted to estimate a regression for this subsample?

(b) Express the hypothesis"marriage status does not affect mean wages" as a restriction on the coefficients. How many restrictions is this?

(c) Find the Wald for this hypothesis. What is the appropriate distribution for the test statistic? Calculate the p-value of the test.

(d) What do you conclude

Solution

R solution at the end of the document.

3 Lecture 9 Resampling methods

Exercise 6

(Ch. 10, ex. 10.19) Take the model $y = \mathbf{X}'\beta + e$ with $\mathbb{E}[\mathbf{X}e] = 0$. Describe the bootstrap percentile confidence interval for $\sigma^2 = \mathbb{E}[e^2]$.

Solution

First, we set a number of bootstrap draws B (e.g. 10.000) and a number α (e.g. 0.05) that denotes what percentiles we restrict the interval to. Then we draw B i.i.d. samples with replacement. In each sample we calculate the sample error variance $\frac{1}{n}\sum_{i=1}^{n}\hat{e}_{ib}^2 = \hat{\sigma}_b^2$. We then order B estimates by their size and pick the following quantiles $q_{\frac{\alpha}{2}}$ (e.g. $10.000 * 0.025 = 250$ so the 250th observation) and $q_{1-\frac{\alpha}{2}}$ (e.g. the 9750s observation). These denote the upper and lower bound of the percentile interval.

Exercise 7

(Ch 10, ex. 10.20) The model is $y = X_1\beta_1 + X_2\beta_2 + e$ with $\mathbb{E}[\mathbf{X}e] = 0$ and X_2 scalar. Describe how to test $\mathrm{H}_0 : \beta_2 = 0$ against $\mathrm{H}_1 : \beta_2 \neq 0$ using the nonparametric bootstrap.

Solution

As before we pick α and B. Note that we are asked to test a linear, one-dimensional hypothesis. This suggests a t-test (a BS t-test in this case). To do that, we must calculate $\hat{\beta}_2$ - the OLS estimate on the whole sample. Then we can draw our first i.i.d. BS sample and calculate our bootstrap t-statistic:

$$T^*(b) = \frac{\hat{\beta}_2^*(b) - \hat{\beta}_2}{s(\hat{\beta}_2^*(b))}$$

Here $\hat{\beta}_2^*(b)$ denotes the OLS estimator on the BS sample b. We repeat this procedure B times and order all B estimates $|T^*(b)|$ according to size. We now pick the $1-\alpha$ quintile $q^*_{1-\alpha}$ (e.g. the 95th percentile). Note that until now the process would be exactly the same if we interested in any other hypothesis (e.g. $\beta_2 = 100$). This changes in our last step. Here we calculate our specific t-statistic $T = \frac{\hat{\beta}_2 - 0}{s(\hat{\beta}_2)}$ and compare it with the rest. So we reject H_0 if $|T| > q^*_{1-\alpha}$.

Exercise 8

(Ch. 10, ex. 10.28) In Exercise 9.26 you estimated a cost function for 145 electric companies and tested the restriction $\theta = \beta_3 + \beta_4 + \beta_5 = 1$.

(a) Estimate the regression by unrestricted least squares and report standard errors calculated by asymptotic, jackknife and the bootstrap.

(b) Estimate $\theta = \beta_3 + \beta_4 + \beta_5$ and report standard errors calculated by asymptotic, jackknife and the bootstrap.

(c) Report confidence intervals for θ using the percentile and BCa methods.

Solution

R solution at the end of the document.

Exercise 9

(Ch. 10, ex. 10.30) In Exercise 7.28 you estimated a wage regression with the `cps09mar` dataset and the subsample of white Male Hispanics. Further restrict the sample to those never-married and live in the Midwest region. (This sample has 99 observations.) As in subquestion (b) let θ be the ratio of the return to one year of education to the return of one year of experience.

(a) Estimate θ and report standard errors calculated by asymptotic, jackknife and the bootstrap.

(b) Explain the discrepancy between the standard errors.

(c) Report confidence intervals for θ using the BC percentile method.

Solution

R solution at the end of the document.

Metrics1 PS4

Artschil Okropiridse

Housekeeping

```r
rm(list=ls())

library(readr)
library(tidyverse)
library(knitr)
library(estimatr)
library(sandwich)
library(matrixStats)
```

Question 4

```r
# Load data
Nerlove1963 <- read.table("Data/Nerlove1963.txt", header = TRUE, sep = "\t")

# Prepare data
data <- Nerlove1963 |>
  mutate(lC = log(Cost),
         lQ = log(output),
         lPL = log(Plabor),
         lPK = log(Pcapital),
         lPF = log(Pfuel))

# Create matrix y
y <- matrix(data$lC, ncol = 1)

# Create matrix x
```

```r
x <- as.matrix(cbind(
  matrix(1,
    nrow = nrow(
      matrix(data$lQ, ncol = 1)
    ),
    1
  ),
  data$lQ, data$lPL, data$lPK, data$lPF
))

# Calculate the inverse of x'x
inv.xx <- solve(t(x) %*% x)

# Get the number of observations (n) and the number of variables (k)
n <- nrow(x)
k <- ncol(x)

# Estimate OLS
beta.ols <- inv.xx%*%t(x)%*%y

# Standard errors
e.ols <- rep((y-x%*%beta.ols),times=k)
xe.ols <- x*e.ols
V.ols <- (n/(n-k))*inv.xx%*%(t(xe.ols)%*%xe.ols)%*%inv.xx
se.ols <- sqrt(diag(V.ols))

# Report coefficients
table <-
  rbind(t(beta.ols),se.ols) %>%
  as_tibble() %>%
  bind_cols(` ` = c('Beta', 'Se'), .)
names(table) <- c(' ', 'Constant', 'lQ', 'lPL', 'lPK', 'lPF')

# To double check let's also use one of R's packages
summary(lm_robust(data = data, lC ~ lQ + lPL + lPK + lPF, se_type = "HC1"))
```

```
Call:
lm_robust(formula = lC ~ lQ + lPL + lPK + lPF, data = data, se_type = "HC1")

Standard error type:  HC1
```

```
Coefficients:
             Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper  DF
(Intercept)  -3.5265    1.71860 -2.0520 4.204e-02 -6.92427  -0.1287 140
lQ            0.7204    0.03260 22.0997 1.717e-47  0.65595   0.7848 140
lPL           0.4363    0.24564  1.7764 7.784e-02 -0.04929   0.9220 140
lPK          -0.2199    0.32381 -0.6791 4.982e-01 -0.86008   0.4203 140
lPF           0.4265    0.07548  5.6505 8.626e-08  0.27728   0.5758 140

Multiple R-squared:  0.926 ,    Adjusted R-squared:  0.9238
F-statistic: 175.8 on 4 and 140 DF,  p-value: < 2.2e-16
```

```r
kable(table)
```

	Constant	lQ	lPL	lPK	lPF
Beta	-3.526503	0.7203941	0.4363412	-0.2198884	0.4265170
Se	1.718601	0.0325975	0.2456358	0.3238121	0.0754827

Note above, a few expressions are a bit confusing to wrap your head around.

`{r} e.ols <- rep((y-x%*%beta.ols),times=k)` this creates an $n \times k$ long vector repeating the $n \times 1$ error vector k times.

The next step is `{r} xe.ols <- x*e.ols` this multiplies each X_i of each observation by the residual of that observation. I.e.

$$\begin{pmatrix} 1 * e_1 & x_{11} * e_1 & x_{12} * e_1 & \dots \\ 1 * e_2 & x_{21} * e_2 & x_{22} * e_2 & \dots \\ \vdots & \dots & & \end{pmatrix}$$

Lastly, `{r} t(xe.ols)%\*%xe.ols` gives us $\sum_n X_i X_i' \hat{e}_i^2$.

(a)

(a) The restriction implies that the production function has constant returns to scale.

(b)

```
# Restriction matrix
R <- c(0,0,1,1,1)

# CLS estimation
beta.cls <- beta.ols - inv.xx%*%R%*%solve(t(R)%*%inv.xx%*%R)%*%(t(R)%*%beta.ols - 1)

# Standard errors
e.cls <- rep((y-x%*%beta.cls),times=k)
xe.cls <- x*e.cls
iR <- inv.xx%*%R%*%solve(t(R)%*%inv.xx%*%R)%*%(t(R))
V.tilde <- (n/(n-k+1))*inv.xx%*%(t(xe.cls)%*%xe.cls)%*%inv.xx
V.cls <- V.tilde - iR%*%V.tilde - V.tilde%*%t(iR) +iR%*%V.tilde%*%t(iR)
se.cls <- sqrt(diag(V.cls))

# Report coefficients
table <-
  rbind(t(beta.cls),se.cls) %>%
  as_tibble() %>%
  bind_cols(` ` = c('Beta', 'Se'), .)
names(table) <- c(' ', 'Constant', 'lQ', 'lPL', 'lPK', 'lPF')
kable(table)
```

	Constant	lQ	lPL	lPK	lPF
Beta	-4.6907891	0.7206875	0.5929096	-0.0073811	0.4144715
Se	0.8148579	0.0324593	0.1690685	0.1557913	0.0728673

(c)

```
# Efficient minimum distance
beta.emd <- beta.ols - V.ols%*%R%*%solve(t(R)%*%V.ols%*%R)%*%(t(R)%*%beta.ols-1)

# Standard errors
e.emd <- rep((y-x%*%beta.emd),times=k)
xe.emd <- x*e.emd
V2 <- (n/(n-k+1))*inv.xx%*%(t(xe.emd)%*%xe.emd)%*%inv.xx
V.emd <- V2 - V2%*%R%*%solve(t(R)%*%V2%*%R)%*%t(R)%*%V2
se.emd <- sqrt(diag(V.emd))
```

```r
# Report coefficients
table <-
  rbind(t(beta.emd),se.emd) %>%
  as_tibble() %>%
  bind_cols(` ` = c('Beta', 'Se'), .)
names(table) <- c(' ', 'Constant', 'lQ', 'lPL', 'lPK', 'lPF')
kable(table)
```

	Constant	lQ	lPL	lPK	lPF
Beta	-4.7446460	0.7201908	0.5805196	0.0092190	0.4102613
Se	0.8154166	0.0323057	0.1694646	0.1552476	0.0724407

(d)

```r
# Wald statistic
W <- t(t(R)%*%beta.ols - 1) %*% solve(t(R)%*%V.ols%*%R) %*% (t(R)%*%beta.ols - 1)

# Find critical value in chi2 distribution
c <- qchisq(.95, df=1)

print(W)
```

```
          [,1]
[1,] 0.6454737
```

```r
print(c)
```

```
[1] 3.841459
```

Since our Wald statistic is smaller than the critical value stemming from a χ^2_1 distribution. We cannot reject H_0.

(e)

```
# Efficient minimum distance statistic
J <- t(beta.ols-beta.emd) %*% solve(V.ols) %*% (beta.ols-beta.emd)

print(J)
```

```
          [,1]
[1,] 0.6454737
```

```
print(c)
```

```
[1] 3.841459
```

As we can see $J = W$ from (d) and we can therefore again not reject H_0.

Question 5

```
# Load data
data <- read.table("Data/cps09mar.txt", header = TRUE)

# Keep sub-sample
data <- data %>% filter(race == 2)

# Prepare data
data <- data %>%
  mutate(lwage = log(earnings / (hours * week))) %>%
  mutate(experience = age - education - 6) %>%
  mutate(experience2 = experience^2 / 100) %>%
  mutate(married = as.integer((marital <= 3))) %>%
  mutate(formerly_married = as.integer(marital >= 4 & marital <= 6)) %>%
  mutate(female_union = as.integer(female == 1 & union == 1)) %>%
  mutate(male_union = as.integer(female == 0 & union == 1)) %>%
  mutate(married_female = as.integer(female == 1 & married == 1)) %>%
  mutate(married_male = as.integer(female == 0 & married == 1)) %>%
  mutate(formerly_married_female = as.integer(female == 1 & formerly_married == 1)) %>%
  mutate(formerly_married_male = as.integer(female == 0 & formerly_married == 1))
```

```r
y <- matrix(data$lwage, ncol = 1)
x <- as.matrix(cbind(
  rep(1, nrow(y)), data$education, data$experience,
  data$experience2, data$female, data$female_union, data$male_union,
  data$married_female, data$married_male, data$formerly_married_female,
  data$formerly_married_male
))
inv.xx <- solve(t(x) %*% x)
n <- nrow(x)
k <- ncol(x)
```

(a)

Since we restrict our sample to only include non-Hispanic Black individuals, we have to drop all variables related to race to avoid multicollinearity.

(b)

We want to test the hypothesis that marital status does not affect wages. Given that we have specified the correct model this implies that all coefficients related to marital should equal 0. Put differently H_0 can be expressed as the four restrictions

$$
\begin{aligned}
\beta_{\text{married,female}} &= 0 \\
\beta_{\text{formerly married,female}} &= 0 \\
\beta_{\text{married,male}} &= 0 \\
\beta_{\text{formerly married,male}} &= 0.
\end{aligned}
$$

This gives us the restriction matrix

$$
\mathbf{R}' = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
$$

(c)

Note that the distribution of the Wald-statistic, under the H_0 is χ^2_k.

```r
# Restriction matrix
R <- cbind(c(0,0,0,0,0,0,0,1,0,0,0),
           c(0,0,0,0,0,0,0,0,1,0,0),
           c(0,0,0,0,0,0,0,0,0,1,0),
           c(0,0,0,0,0,0,0,0,0,0,1))

# Estimate OLS
beta.ols <- inv.xx%*%t(x)%*%y

# Standard errors
e.ols <- rep((y-x%*%beta.ols),times=k)
xe.ols <- x*e.ols
V.ols <- (n/(n-k))*inv.xx%*%(t(xe.ols)%*%xe.ols)%*%inv.xx

# Wald test statistic
W = t(t(R)%*%beta.ols)%*%solve(t(R)%*%V.ols%*%R)%*%(t(R)%*%beta.ols)
print(W)
```

```
         [,1]
[1,] 39.70686
```

```r
# Find critical value in chi2 distribution
c <- qchisq(.95, df=4)

print(c)
```

```
[1] 9.487729
```

```r
# p-values
pval <- 1-pchisq(W,df=4)
print(pval)
```

```
             [,1]
[1,] 4.976708e-08
```

(d)

Since $W > c$ and the p-value is effectively 0 we can reject the hypothesis that marital status has no effect on wage.

Question 8

```
# Load data
data <- read.table("Data/Nerlove1963.txt",header=TRUE)

# Prepare data
data %>% mutate(lC = log(Cost)) -> data
data %>% mutate(lQ = log(output)) -> data
data %>% mutate(lPL = log(Plabor)) -> data
data %>% mutate(lPK = log(Pcapital)) -> data
data %>% mutate(lPF = log(Pfuel)) -> data

y <- matrix(data$lC,ncol=1)
x <- as.matrix(cbind(matrix(1,nrow(matrix(data$lQ,ncol=1)),1),data$lQ,data$lPL,data$lPK,da
inv.xx <- solve(t(x)%*%x)
n <- nrow(x)
k <- ncol(x)
```

(a)

```
## Unrestricted estimation
# Estimate OLS
beta.ols <- inv.xx%*%t(x)%*%y

# Standard errors
e.ols <- rep((y-x%*%beta.ols),times=k)
xe.ols <- x*e.ols
V.ols <- (n/(n-k))*inv.xx%*%(t(xe.ols)%*%xe.ols)%*%inv.xx
se.ols <- sqrt(diag(V.ols))

# Jackknife
beta_loo <- matrix(NA,nrow=k,ncol=n)
beta_bar_loo <- matrix(0,nrow=k,ncol=1)

for (i in 1:n){
  y_loo <- as.matrix(y[-i,], ncol=1)
  x_loo <- as.matrix(x[-i,])
  inv.xx_loo <- solve(t(x_loo) %*% x_loo)

  beta_loo[,i] <- inv.xx_loo%*%t(x_loo)%*%y_loo
```

```
  beta_bar_loo <- beta_bar_loo + beta_loo[,i]
}

beta_bar_loo <- beta_bar_loo/n

difference <- matrix(0,nrow=k,ncol=k)

for (i in 1:n){
  difference <- difference + (
    matrix(beta_loo[,i]) - matrix(beta_bar_loo)
    ) %*% t(matrix(beta_loo[,i]) - matrix(beta_bar_loo))
}

V.jack <- (n-1)/n * difference
se.jack <- sqrt(diag(V.jack))

# Bootstrap
set.seed(123)
B <- 1000
beta.boot <- matrix(NA,nrow=k,ncol=B)

for (i in 1:B) {
  sample <- data[sample(n,replace=TRUE),]
  beta.boot[,i] <- coef(lm(lC ~ lQ + lPL + lPK + lPF, data = sample))
}

se.boot <- sqrt(rowVars(beta.boot))

table <-
  rbind(t(beta.ols),se.ols,se.jack,se.boot) %>%
  as_tibble() %>%
  bind_cols(` ` = c('Beta', 'Asymptotic', 'Jackknife', 'Bootstrap'), .)
names(table) <- c(' ', 'Constant', 'lQ', 'lPL', 'lPK', 'lPF')
kable(table)
```

	Constant	lQ	lPL	lPK	lPF
Beta	-3.526503	0.7203941	0.4363412	-0.2198884	0.4265170
Asymptotic	1.718601	0.0325975	0.2456358	0.3238121	0.0754827
Jackknife	1.788028	0.0339337	0.2531660	0.3363424	0.0777519
Bootstrap	1.737121	0.0328970	0.2500843	0.3283541	0.0753286

(b)

We are going to us the plug-in estimator $\hat{\theta} = \hat{\beta}_3 + \hat{\beta}_4 + \hat{\beta}_5$.

```
# OLS
ols <- lm(lC ~ lQ + lPL + lPK + lPF, data = data)
beta.ols <- coef(ols)
theta.ols <- beta.ols[3] + beta.ols[4] + beta.ols[5]

# Standard errors
R <- c(0,1,1,1,0)
V1 <- vcovHC(ols, type="HC1")
V.theta.ols <- t(R)%*%V1%*%R
se.theta.ols <- sqrt(V.theta.ols)

# Jackknife
theta_loo <- rep(NA,n)
theta_bar_loo <- 0

for (i in 1:n){
  y_loo <- as.matrix(y[-i,], ncol=1)
  x_loo <- as.matrix(x[-i,])
  inv.xx_loo <- solve(t(x_loo) %*% x_loo)
  beta_loo <- inv.xx_loo%*%t(x_loo)%*%y_loo

  theta_loo[i] <- beta_loo[3,1] + beta_loo[4,1] + beta_loo[5,1]
  theta_bar_loo <- theta_bar_loo + theta_loo[i]
}

theta_bar_loo <-theta_bar_loo/n

difference <- 0
for (i in 1:n){
  difference <- difference + (theta_loo[i] - theta_bar_loo)^2
}

V.theta.jack <- (n-1)/n * difference
se.theta.jack <- sqrt(V.theta.jack)

# Bootstrap
set.seed(123)
B <- 1000
```

```r
theta.boot <- rep(NA,B)

for (i in 1:B) {

  sample <- data[sample(n,replace=TRUE),]
  beta.boot <- coef(lm(lC ~ lQ + lPL + lPK + lPF, data = sample))
  theta.boot[i] <- beta.boot[3] + beta.boot[4] + beta.boot[5]

}

se.theta.boot <- sqrt(var(theta.boot))

# Report results
table <-
  rbind(theta.ols,se.theta.ols,se.theta.jack,se.theta.boot) %>%
  as_tibble() %>%
  bind_cols(` ` = c('Theta', 'Asymptotic', 'Jackknife', 'Bootstrap'), .)
names(table) <- c('','')
kable(table)
```

Theta	0.6429698
Asymptotic	0.4716812
Jackknife	0.4626814
Bootstrap	0.4474232

(c)

```r
# Percentile interval

# Sort the distribution of bootstrap thetas from (b)
alpha <- 0.05
sorted.theta <- sort(theta.boot)
q.pc.lower <- sorted.theta[(alpha/2)*B]
q.pc.upper <- sorted.theta[(1-(alpha/2))*B]

q.pc <- c(q.pc.lower, q.pc.upper)

# BCa
# See Hansen Ch. 10.18 for details
```

```
a <- (sum(theta_bar_loo - theta_loo))^3/(6*(sum(theta_bar_loo - theta_loo)^2))^(3/2)
p <- mean(theta.boot <= theta.ols)

z.lower <- qnorm(alpha/2)
z.upper <- qnorm(1-(alpha/2))
z0 <- qnorm(p)

x.lower <- pnorm(z0 + (z.lower+z0)/(1-a*(z.lower + z0)))
x.upper <- pnorm(z0 + (z.upper+z0)/(1-a*(z.upper + z0)))

q.bca.lower <- sorted.theta[x.lower*B]
q.bca.upper <- sorted.theta[x.upper*B]

q.bca <- c(q.bca.lower,q.bca.upper)

# Report results
table <-
  rbind(q.pc,q.bca) %>%
  as_tibble() %>%
  bind_cols(` ` = c('95% Percentile Interval', '95% BCa Interval'), .)
names(table) <- c('', 'q lower','q upper')
kable(table)
```

	q lower	q upper
95% Percentile Interval	-0.2366988	1.527583
95% BCa Interval	-0.3039949	1.457843

Question 9

```
# Load data
data <- read.table("Data/cps09mar.txt",header=TRUE)

# Keep sub sample
data <- data %>% filter(race == 1 & female == 0 & hisp == 1 & region == 2 & marital == 7 )

# Prepare data
data <- data %>% mutate(lwage = log(earnings/(hours*week))) %>%
  mutate(experience = age - education - 6)  %>%
  mutate(experience2 = experience^2/100)
```

```r
y <- matrix(data$lwage,ncol=1)
x <- as.matrix(cbind(rep(1,nrow(y)),data$education, data$experience,
                     data$experience2))
inv.xx <- solve(t(x)%*%x)
n <- nrow(x)
k <- ncol(x)
```

(a)

We will estimate the regression

$$\text{lwage} = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{experience} + \beta_3 \text{experience}^2/100 + e$$

and then use the plugin estimator

$$\hat{\theta} = \hat{\beta}_1 / \left(\hat{\beta}_2 + \frac{1}{5}\hat{\beta}_3 \right)$$

where $\left(\hat{\beta}_2 + \frac{1}{5}\hat{\beta}_3 \right)$ is the derivative of lwage wrt to experience evaluated at experience $= 10$.

```r
# OLS
ols <- lm(lwage ~ education + experience + experience2, data = data)
beta.ols <- coef(ols)
theta.ols <- beta.ols[2] / (beta.ols[3]  + beta.ols[4]/5)
```

We will then use the Delta method to get asymptotic standard errors. Note that

$$\theta = g(\beta) = \beta_1 / \left(\beta_2 + \frac{1}{5}\beta_3 \right)$$

and therefore

$$\mathbf{V}_\theta = \frac{\partial}{\partial \beta} g' \mathbf{V}_\beta \frac{\partial}{\partial \beta} g.$$

where

$$\frac{\partial}{\partial \beta} g = \begin{pmatrix} 0 \\ 1/\left(\beta_2 + \frac{1}{5}\beta_3\right) \\ -\beta_1/\left(\beta_2 + \frac{1}{5}\beta_3\right)^2 \\ -\frac{1}{5}\beta_1/\left(\beta_2 + \frac{1}{5}\beta_3\right)^2 \end{pmatrix}.$$

Bootstrap fun fact: Sometimes it can be better to repeatedly draw random weights for each observation instead of performing the classical bootstrap that we are using here. This is especially true in situations in which excluding certain observations can lead to multicollinearitry. See Shao and Tu (1995) Bayesian Bootstrap and Random Weighting.

```
# Standard errors
g.partial <- rbind(0,
                   1 / (beta.ols[3]  + beta.ols[4]/5),
                   -beta.ols[2] / (beta.ols[3]  + beta.ols[4]/5)^2,
                   -beta.ols[2] /( 5 * (beta.ols[3]  + beta.ols[4]/5)^2))

V.beta <- vcovHC(ols, type="HC1")
V.theta.ols <- t(g.partial)%*%V.beta%*%g.partial
se.theta.ols <- sqrt(V.theta.ols)

# Jackknife
# Note that we can just loop over all observations here since the sample is small
# In large samples you may want to compute the LOO coefficients using the leverage values

theta.loo <- rep(NA,n)
theta.bar.loo <- 0

for (i in 1:n){
  y_loo <- as.matrix(y[-i,], ncol=1)
  x_loo <- as.matrix(x[-i,])
  inv.xx_loo <- solve(t(x_loo) %*% x_loo)
  beta.loo <- inv.xx_loo%*%t(x_loo)%*%y_loo

  theta.loo[i] <- beta.loo[2,1]/(beta.loo[3,1] + beta.loo[4,1]/5)
  theta.bar.loo <- theta.bar.loo + theta.loo[i]
}

theta.bar.loo <-theta.bar.loo/n

difference <- 0
for (i in 1:n){
  difference <- difference + (theta.loo[i] - theta.bar.loo)^2
}

V.theta.jack <- (n-1)/n * difference
se.theta.jack <- sqrt(V.theta.jack)

# Bootstrap
set.seed(123)
B <- 1000
theta.boot <- rep(NA,B)
```

```r
for (i in 1:B) {

  sample <- data[sample(n,replace=TRUE),]
  ols.boot <- lm(lwage ~ education + experience + experience2, data = sample)
  beta.boot <- coef(ols.boot)
  theta.boot[i] <- beta.boot[2] / (beta.boot[3]  + beta.boot[4]/5)

}

se.theta.boot <- sqrt(var(theta.boot))

# Report results
table <-
  rbind(theta.ols,se.theta.ols,se.theta.jack,se.theta.boot) %>%
  as_tibble() %>%
  bind_cols(` ` = c('Theta', 'Asymptotic', 'Jackknife', 'Bootstrap'), .)
names(table) <- c('','')
kable(table)
```

Theta	2.8993231
Asymptotic	0.7603923
Jackknife	0.8229674
Bootstrap	0.9981012

(b)

Note that since g is non-linear, $\hat{\theta}$ is biased in small samples as $E[g(\hat{\beta})] \neq g\left(E\left[\hat{\beta}\right]\right)$ (our sample size is only 99!). Since we rely on $\hat{\theta}$ in estimating the asymptotic variance the corresponding standard errors will likely also not be correct. Furthermore, an untrimmed bootstrap estimator can also be unreliable for non-linear estimators (see Hansen p. 277). As a sanity check you can re-estimate the bootstrap standard errors using a different seed.

(c)

```r
# Percentile interval

# Sort the distribution of bootstrap thetas from (a)
alpha <- 0.05
sorted.theta <- sort(theta.boot)
```

```r
q.pc.lower <- sorted.theta[(alpha/2)*B]
q.pc.upper <- sorted.theta[(1-(alpha/2))*B]

q.pc <- c(q.pc.lower, q.pc.upper)


# Report results
table <-
  rbind(q.pc) %>%
  as_tibble() %>%
  bind_cols(` ` = c('95% Percentile Interval'), .)
names(table) <- c('', 'q lower','q upper')
kable(table)
```

	q lower	q upper
95% Percentile Interval	1.840826	5.444653