

Econ 675 Assignment 1

Nathan Mather*

October 27, 2018

Contents

1	Question 1: Non-linear Least Squares	1
1.1	Q1 Part 1	1
1.2	Q1 Part 2	2
1.3	Q1 Part 3	3
1.4	Q1 Part 4	3
1.5	Q1 Part 5	3
1.6	Q1 Part 6	4
1.7	Q1 Part 7	4
1.8	Q1 Part 9	4
2	Question 2: Semiparametric GMM with Missing Data	5
2.1	Q2 part 1	5
2.2	Q2 part 2	7
3	Question 3: When Bootstrap Fails	7
3.1	Q3 part 1	7
3.2	Q3 part 2	8

1 Question 1: Non-linear Least Squares

1.1 Q1 Part 1

The general non-linear least squares estimator is

$$\hat{\beta}_n = \arg \min_{\beta \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^n (y_i - \mu(\mathbf{x}'_i \beta))^2$$

Now for $\beta_0 = \arg \min_{\beta \in \mathbb{R}^d} E[(y_i - \mu(\mathbf{x}'_i \beta))^2]$ to be identifiable we need:

$$\beta_0 = \beta_0^*$$

$$\iff \beta_0^* = \arg \min_{\beta \in \mathbb{R}^d} E[(y_i - \mu(\mathbf{x}'_i \beta))^2]$$

*Shouts out to Ani, Paul, Tyler, Erin, Caitlin and others for all the help with this

To find this start by noting that

$$\begin{aligned}
E[(y_i - \mu(\mathbf{x}'_i \boldsymbol{\beta}))^2] &= E[(y_i - \mu(\mathbf{x}'_i \boldsymbol{\beta}_0) + \mu(\mathbf{x}'_i \boldsymbol{\beta}_0) - \mu(\mathbf{x}'_i \boldsymbol{\beta}))^2] \\
&= E[(y_i - \mu(\mathbf{x}'_i \boldsymbol{\beta}_0))^2] + E[(\mu(\mathbf{x}'_i \boldsymbol{\beta}_0) - \mu(\mathbf{x}'_i \boldsymbol{\beta}))^2] + 2E[(y_i - \mu(\mathbf{x}'_i \boldsymbol{\beta}_0))(\mu(\mathbf{x}'_i \boldsymbol{\beta}_0) - \mu(\mathbf{x}'_i \boldsymbol{\beta}))] \\
&= E[(y_i - \mu(\mathbf{x}'_i \boldsymbol{\beta}_0))^2] + E[(\mu(\mathbf{x}'_i \boldsymbol{\beta}_0) - \mu(\mathbf{x}'_i \boldsymbol{\beta}))^2]
\end{aligned}$$

The last equality comes from the last term being zero by iterated expectations. I show this below.

$$\begin{aligned}
E[(y_i - \mu(\mathbf{x}'_i \boldsymbol{\beta}_0))(\mu(\mathbf{x}'_i \boldsymbol{\beta}_0) - \mu(\mathbf{x}'_i \boldsymbol{\beta}))] &= E[E[(y_i - \mu(\mathbf{x}'_i \boldsymbol{\beta}_0))(\mu(\mathbf{x}'_i \boldsymbol{\beta}_0) - \mu(\mathbf{x}'_i \boldsymbol{\beta})) | \mathbf{x}_i]] \\
&= E[(E[y_i | \mathbf{x}_i] - \mu(\mathbf{x}'_i \boldsymbol{\beta}_0))(\mu(\mathbf{x}'_i \boldsymbol{\beta}_0) - \mu(\mathbf{x}'_i \boldsymbol{\beta}))] = 0
\end{aligned}$$

Using this fact we have that

$$E[(y_i - \mu(\mathbf{x}'_i \boldsymbol{\beta}))^2] = E[(y_i - \mu(\mathbf{x}'_i \boldsymbol{\beta}_0))^2] + E[(\mu(\mathbf{x}'_i \boldsymbol{\beta}_0) - \mu(\mathbf{x}'_i \boldsymbol{\beta}))^2] \geq E[(y_i - \mu(\mathbf{x}'_i \boldsymbol{\beta}_0))^2]$$

$$\forall \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$$

This is strictly greater than unless $\exists \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ such that $E[(\mu(\mathbf{x}'_i \boldsymbol{\beta}_0) - \mu(\mathbf{x}'_i \boldsymbol{\beta}))^2] = 0$. Thus this gives us an identification condition that $E[(\mu(\mathbf{x}'_i \boldsymbol{\beta}_0) - \mu(\mathbf{x}'_i \boldsymbol{\beta}))^2] \neq 0 \forall \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$. This means that $\boldsymbol{\beta}_0$ is the unique minimizer of $E[(y_i - \mu(\mathbf{x}'_i \boldsymbol{\beta}))^2]$.

Next note that if $\mu(\cdot)$ is a linear function, $\boldsymbol{\beta}_0$ is the coefficient of the best linear predictor and has the usual closed form $\boldsymbol{\beta}_0 = E[\mathbf{x}_i \mathbf{x}'_i]^{-1} E[\mathbf{x}_i y_i]$.

1.2 Q1 Part 2

In order to set this up as a Z estimator let's take a first order condition. This gives us the following condition.

$$E[(\mu(\mathbf{x}'_i \boldsymbol{\beta}_0) - \mu(\mathbf{x}'_i \boldsymbol{\beta})) \dot{\mu}(\mathbf{x}'_i \boldsymbol{\beta}) \mathbf{x}_i] = 0$$

Now take the sample analog and let $m(\mathbf{z}_i, \boldsymbol{\beta}) = (y_i - \mu(\mathbf{x}'_i \boldsymbol{\beta})) \dot{\mu}(\mathbf{x}'_i \boldsymbol{\beta}) \mathbf{x}_i$ where $\mathbf{z}_i = (y_i, \mathbf{x}'_i)'$. We can write $\hat{\boldsymbol{\beta}}_n$ as the Z-estimator that solves:

$$0 = \frac{1}{n} \sum_{i=1}^n m(\mathbf{z}_i, \hat{\boldsymbol{\beta}}_n)$$

Now assuming $\hat{\boldsymbol{\beta}}_n \rightarrow \boldsymbol{\beta}_0$ and regularity conditions we get the standard M estimation result.

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \rightarrow_d \mathcal{N}(0, \mathbf{H}_0^{-1} \boldsymbol{\Sigma}_0 \mathbf{H}_0^{-1})$$

Where

$$\mathbf{H}_0 = E \left[\frac{\partial}{\partial \boldsymbol{\beta}} m(\mathbf{z}_i, \boldsymbol{\beta}_0) \right] = E[\dot{\mu}(\mathbf{x}'_i \boldsymbol{\beta}_0)^2 \mathbf{x}_i \mathbf{x}'_i]$$

and

$$\boldsymbol{\Sigma}_0 = V[m(\mathbf{z}_i, \boldsymbol{\beta}_0)] = E[\sigma^2(\mathbf{x}_i) \dot{\mu}(\mathbf{x}'_i \boldsymbol{\beta}_0)^2 \mathbf{x}_i \mathbf{x}'_i]$$

1.3 Q1 Part 3

$$\hat{V}_n^{HC} = \left(\frac{1}{n} \sum_{i=1}^n \hat{m} \hat{m}' \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n \hat{m} \hat{m}' \hat{e}_i^2 \right) \left(\frac{1}{n} \sum_{i=1}^n \hat{m} \hat{m}' \right)^{-1}$$

where $\hat{m} = \mathbf{m}_\beta(\mathbf{z}_i, \hat{\beta})$ and $\hat{e}_i = y_i - \mathbf{m}(\mathbf{z}_i, \hat{\beta})$

Now by the delta method and letting $g(\beta) = \|\beta\| = \sum_{k=1}^d \beta_k^2$ we get

$$\sqrt{n}(g(\hat{\beta}_n) - g(\beta_0)) \rightarrow_d \mathcal{N}(0, \dot{g}(\beta_0) \mathbf{V}_0 \dot{g}(\beta_0)')$$

where $\dot{g}(\beta_0) = \frac{d}{d\beta'} g(\beta) = 2\beta'$ Hence the confidence interval is given by

$$CI_{0.95} = \left[\|\hat{\beta}_n\|^2 - 1.96 \sqrt{\frac{4\hat{\beta}_n' \hat{V} \hat{\beta}_n}{n}}, \|\hat{\beta}_n\|^2 + 1.96 \sqrt{\frac{4\hat{\beta}_n' \hat{V} \hat{\beta}_n}{n}} \right]$$

1.4 Q1 Part 4

In this case we get $\Sigma_0 = \sigma^2 \mathbf{H}_0$ and the asymptotic variance reduces to

$$\mathbf{V}_0 = \sigma^2 \mathbf{H}_0^{-1} = \sigma^2 \mathbb{E}[\dot{\mu}(\mathbf{x}_i' \beta_0)^2 \mathbf{x}_i \mathbf{x}_i']^{-1}$$

We can estimate variance using $\hat{\mathbf{V}} = \hat{\sigma}^2 \hat{\mathbf{H}}^{-1}$ where

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \mu(\mathbf{x}_i' \hat{\beta}_n))^2$$

and

$$\hat{\mathbf{H}} = \frac{a}{n} \sum_{i=1}^n \dot{\mu}(\mathbf{x}_i' \beta_0)^2 \mathbf{x}_i \mathbf{x}_i'$$

Which is consistent by the continuous mapping theorem. Now by the delta method and letting $g(\beta) = \|\beta\| = \sum_{k=1}^d \beta_k^2$ we get

$$\sqrt{n}(g(\hat{\beta}_n) - g(\beta_0)) \rightarrow_d \mathcal{N}(0, \dot{g}(\beta_0) \mathbf{V}_0 \dot{g}(\beta_0)')$$

where $\dot{g}(\beta_0) = \frac{d}{d\beta'} g(\beta) = 2\beta'$ Hence the confidence interval is given by

$$CI_{0.95} = \left[\|\hat{\beta}_n\|^2 - 1.96 \sqrt{\frac{4\hat{\beta}_n' \hat{V} \hat{\beta}_n}{n}}, \|\hat{\beta}_n\|^2 + 1.96 \sqrt{\frac{4\hat{\beta}_n' \hat{V} \hat{\beta}_n}{n}} \right]$$

1.5 Q1 Part 5

The conditional likelihood function is

$$f_{y|x}(y_i | \mathbf{x}_i) = \frac{1}{(2\pi)^{n/2} \sigma^2} \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu(\mathbf{x}_i' \beta))^2 \right)$$

with log likelihood

$$\ell_n(\beta, \sigma^2) = -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu(\mathbf{x}'_i \beta))^2 - \frac{n}{2} \log(\sigma^2)$$

This gives us the following first order conditions

$$\frac{\partial}{\partial \beta} \ell_n(\beta, \sigma^2) = \frac{1}{\hat{\sigma}_{ML}^2} \sum_{i=1}^n (y_i - \mu(\mathbf{x}'_i \beta_{ML})) \dot{\mu}(\mathbf{x}'_i \beta_{ML}) \mathbf{x}_i = 0$$

$$\frac{\partial}{\partial \sigma^2} \ell_n(\beta, \sigma^2) = \frac{1}{2\hat{\sigma}_{ML}^4} \sum_{i=1}^n (y_i - \mu(\mathbf{x}'_i \beta_{ML}))^2 - \frac{n}{2\hat{\sigma}_{ML}^2} = 0$$

These conditions are equivalent to those found above.

1.6 Q1 Part 6

If the link function is unknown, β_0 is not identified. To see this, consider two pairs of parameters $(\mu(\cdot), \beta_0)$ and $(\tilde{\mu}(\cdot), \tilde{\beta}_0)$ where $\tilde{\mu}(z) = \mu(z/k)$ and $\tilde{\beta}_0 = k\beta_0$ for some $k \neq 0$. Then the parameters are clearly different, but $(\mu(\cdot), \beta_0) = (\tilde{\mu}(\cdot), \tilde{\beta}_0)$. A common normalization is $\|\beta_0\| = 1$, but more conditions are needed to regain identification.

1.7 Q1 Part 7

1.8 Q1 Part 9

(a)

R table

term	estimate	std.error	statistic	p.value	CI.L	CI.H
(Intercept)	1.755	0.335	5.245	0.000	1.099	2.411
S_age	1.333	0.123	10.826	0.000	1.092	1.575
S_HHpeople	-0.067	0.023	-2.871	0.004	-0.112	-0.021
log_inc	-0.119	0.044	-2.707	0.007	-0.205	-0.033

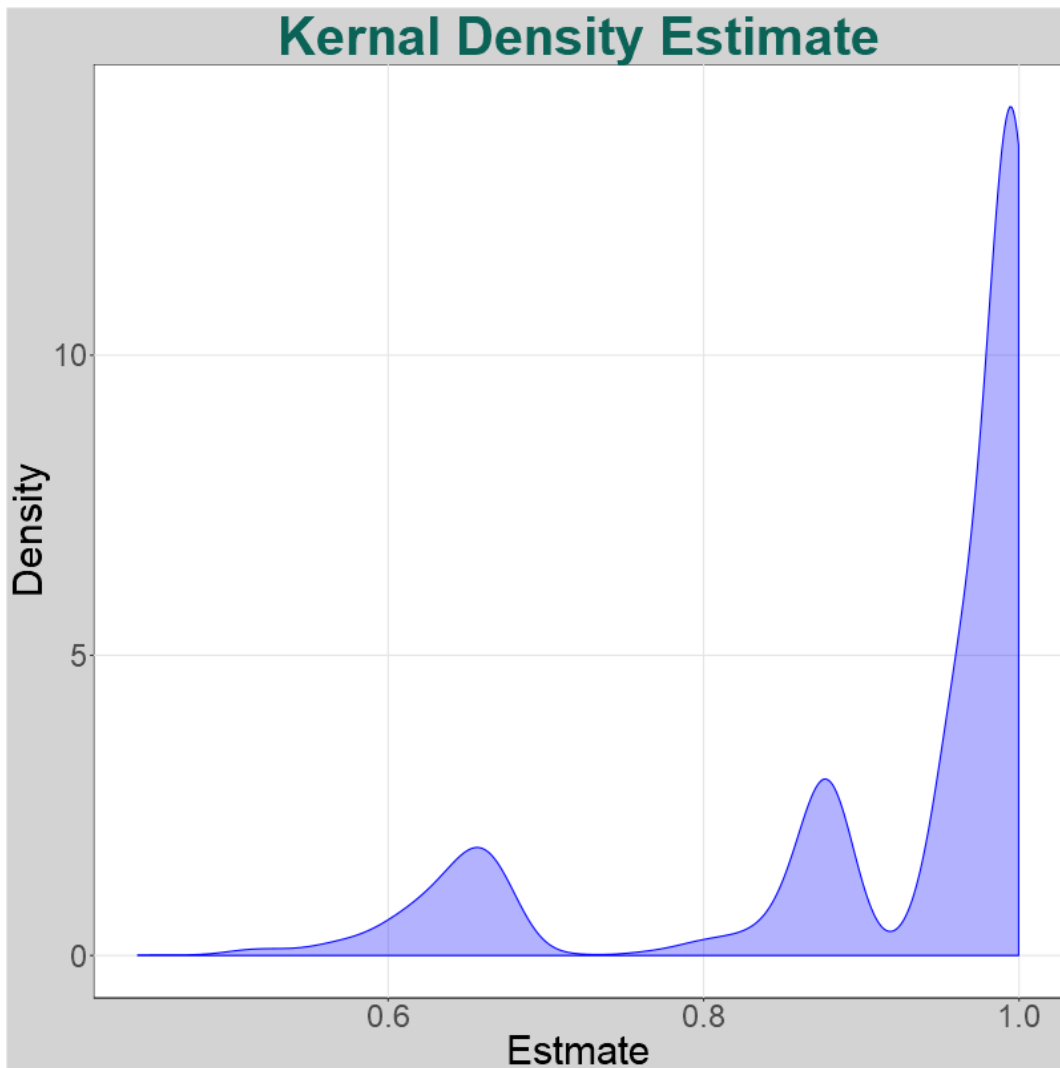
(b)

R table

term	estimate	std.error	t_q.975	t_q.025	CI.L	CI.H
(Intercept)	1.755	0.335	2.167	-1.774	1.161	2.480
S_HHpeople	-0.067	0.023	2.040	-1.958	-0.112	-0.019
S_age	1.333	0.123	2.210	-1.558	1.142	1.606
log_inc	-0.119	0.044	1.817	-2.155	-0.213	-0.039

(c)

R Graph



2 Question 2: Semiparametric GMM with Missing Data

2.1 Q2 part 1

Start with the moment condition we are given

$$0 = \mathbb{E}[m(y_i^*, t_i, x_i; \beta_0) | t_i, x_i]$$

Now by law of iterated expectations we can multiply by a function g and still get zero

$$0 = \mathbb{E}[g(t_i, x_i) \mathbb{E}[m(y_i^*, t_i, x_i; \beta_0) | t_i, x_i]] \text{ for any } g(t_i, x_i)$$

Next we can put g inside the first expectation

$$0 = \mathbb{E}[\mathbb{E}[g(t_i, x_i) m(y_i^*, t_i, x_i; \beta_0) | t_i, x_i]]$$

now removing the inside expectation

$$0 = \mathbb{E}[g(t_i, x_i) m(y_i^*, t_i, x_i; \beta_0)]$$

We want to find g_0 , the optimal g that minimizes $\text{AsyVar}(\hat{\beta})$. Let $z_i = (t_i, x_i)$, $w_i = (y_i^*, t_i, x_i)$, and $\theta = \beta$.

The first thing we need to do is determine the asymptotic variance V associated with

$$\hat{\theta} = \arg \min \left(\frac{1}{n} \sum_i g(z_i) m(w_i, \theta) \right)' W \left(\frac{1}{n} \sum_i g(z_i) m(w_i, \theta) \right)$$

Taking first order conditions and setting equal to zero we get

$$\text{FOC: } 0 = \left[\frac{1}{n} \sum_i \frac{\partial}{\partial \theta} g(z_i) m(w_i, \theta) \right]' W \left[\frac{1}{n} \sum_i \frac{\partial}{\partial \theta} g(z_i) m(w_i, \theta) \right]$$

or

$$0 = \left[\frac{1}{n} \sum_i \frac{\partial}{\partial \theta} g(z_i) m(w_i, \hat{\theta}) \right]' W \left[\frac{1}{n} \sum_i \frac{\partial}{\partial \theta} g(z_i) m(w_i, \hat{\theta}) \right]$$

Since it is euqual to zero we can add another of the same term and multiply by $(\hat{\theta} - \theta_0)$ giving

$$0 = \left[\frac{1}{n} \sum_i \frac{\partial}{\partial \theta} g(z_i) m(w_i, \theta_0) \right]' W \left[\frac{1}{n} \sum_i \frac{\partial}{\partial \theta} g(z_i) m(w_i, \theta_0) \right] + \left[\frac{1}{n} \sum_i \frac{\partial}{\partial \theta} g(z_i) m(w_i, \hat{\theta}) \right]' W \left[\frac{1}{n} \sum_i \frac{\partial}{\partial \theta} g(z_i) m(w_i, \hat{\theta}) \right] (\hat{\theta} - \theta_0)$$

Which can be rearranged to give

$$\sqrt{n}(\hat{\theta} - \theta_0) = (\Omega'_0 W_0 \Omega_0)^{-1} \Omega_0 W_0 \frac{1}{\sqrt{n}} \sum_i g(z_i) m(w_i, \theta) + o_p(1)$$

And then By the CLT, $\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow_d N(0, V_0)$

where $V_0 = (\Omega'_0 W_0 \Omega_0)^{-1} \Omega'_0 W_0 \Sigma_0 W_0 \Omega_0 (\Omega'_0 W_0 \Omega_0)^{-1}$

and where $\Sigma_0 = \mathbb{V}[g(z_i) m(w_i, \theta)]$

setting values Optimally to minimize V gives us the following conditions.

$$W_0^* = \Sigma_0^{-1} \text{ and } V_0^* = \Omega'_0 \Sigma_0 \Omega_0$$

$$g^*(z_i) = \frac{\partial m_i}{\partial \theta} \mathbb{V}[m(w_i, \theta_0) | z_i]^{-1}$$

Now applying this specifically to a probit model gives

$$\mathbb{V}[m(y_i^*, t_i, x_i, \beta_0) | t_i, x_i] = F(t_i \cdot \theta_0 + x_i \gamma_0) (1 - F(t_i \cdot \theta_0 + x_i' \gamma_0))$$

$$\mathbb{E} \left[\frac{\partial}{\partial \beta} m(y_i^*, t_i, x_i, \beta_0) | t_i, x_i \right] = \mathbb{E} [f(t_i \cdot \theta_0 + x_i \gamma_0) (t_i, x_i) | t_i, x_i] = f(t_i \cdot \theta_0 + x_i \gamma_0) [t_i, x_i']'$$

$$\text{Therefore, } g_0(t_i, x_i) = \frac{f(t_i \cdot \theta_0 + x_i \gamma_0)}{F(t_i \cdot \theta_0 + x_i \gamma_0) (1 - F(t_i \cdot \theta_0 + x_i' \gamma_0))} [t_i, x_i']'$$

If F is the logistic cdf we instead get

$$F(x) = \frac{1}{1 + e^{-x}}$$

$$f(x) = \frac{\partial}{\partial x} F(x) = \frac{-e^{-x}}{(1 + e^{-x})^2} = -e^{-x} F(x)^2$$

$$\frac{f(x)}{F(x)(1 - F(x))} = \frac{-e^{-x} F(x)^2}{F(x)(1 - F(x))} = \frac{-e^{-x} F(x)}{1 - F(x)} = 1$$

$$g_0(t_i, x_i) = [t_i, x_i]'$$

2.2 Q2 part 2

(a) The optimal unconditional moment condition is:

$$0 = \mathbb{E}[g(t_i, x_i)m(y_i^*, t_i, x_i; \beta_0)]$$

In part 2.1 we showed that, setting $g = g_0$ this is equivalent to:

$$0 = \mathbb{E}[m(y_i^*, t_i, x_i; \beta_0)|t_i, x_i]$$

Since $s_i \perp (y_i^*, t_i, x_i)$:

$$0 = \mathbb{E}[m(y_i^*, t_i, x_i; \beta_0)|t_i, x_i]$$

$$0 = \mathbb{E}[g_0(t_i, x_i)m(y_i^*, t_i, x_i; \beta_0)]$$

$$0 = \mathbb{E}[g_0(t_i, x_i)m(y_i^*, t_i, x_i; \beta_0)|s_i = 1]$$

Thus, $\hat{\beta}_{MCAR}$ solving $0 \approx \hat{\mathbb{E}}[g_0(t_i, x_i)m(y_i, t_i, x_i; \hat{\beta}_{MCAR})|s_i = 1]$ is consistent for β_0
 $\hat{\beta}_{MCAR, feasible}$ solves $0 \approx \hat{\mathbb{E}}[\hat{g}(t_i, x_i)m(y_i, t_i, x_i; \hat{\beta}_{MCAR})|s_i = 1]$

(b) The tables for this section are below

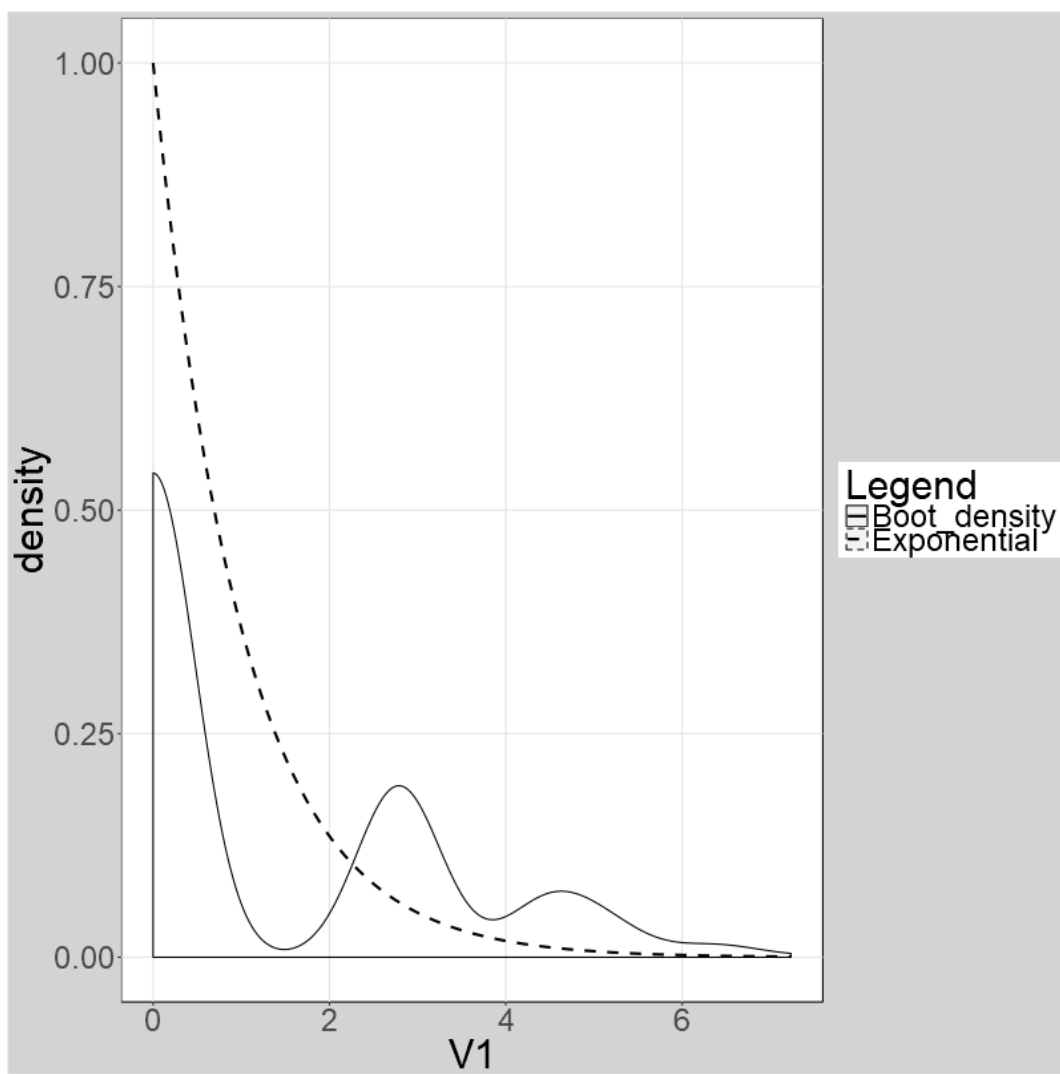
R table

term	Estimate	sd	p-value	t	CI.L	CI.H
dpisofirme	-0.317	0.073	0.008	-4.363	-0.453	-0.187
S.age	-0.244	0.020	0.000	-11.975	-0.284	-0.205
S.HHpeople	0.024	0.013	0.104	1.775	-0.002	0.049
log_inc	0.033	0.014	0.020	2.397	0.006	0.058

3 Question 3: When Bootstrap Fails

3.1 Q3 part 1

R Graph



3.2 Q3 part 2

R Graph

