ECON675 – Assignment 3

Anirudh Yadav

October 26, 2018

Contents

1 Non-linear least squares

1.1 Identifiability

This is a standard M-estimation problem. The parameter vector $\boldsymbol{\beta}_0$ is assumed to solve the population problem

$$\boldsymbol{\beta}_0 = \arg\min_{\beta \in \mathbb{R}^d} \mathbb{E}[(y_i - \mu(\boldsymbol{x}_i' \boldsymbol{\beta}))^2].$$

For $\boldsymbol{\beta}_0$ to be identified, it must be the *unique* solution to the above population problem (i.e. the unique minimizer). In math, this means for all $\epsilon > 0$ and for some $\delta > 0$:

$$\sup_{||\beta - \beta_0|| > \epsilon} M(\boldsymbol{\beta}) \geq M(\boldsymbol{\beta}_0) + \delta$$

where $M(\boldsymbol{\beta}) = \mathbb{E}[(y_i - \mu(\boldsymbol{x}_i' \boldsymbol{\beta}))^2]$. Of course $\boldsymbol{\beta}_0$ can be written in closed form if $\mu(\cdot)$ is linear. In this case, we know that

$$\boldsymbol{\beta}_0 = \mathbb{E}[\boldsymbol{x}_i \boldsymbol{x}_i']^{-1} \mathbb{E}[\boldsymbol{x}_i y_i].$$

1.2 Asymptotic normality

The M-estimator is asymptotically normal if:

1. $\hat{\boldsymbol{\beta}} \to_p \boldsymbol{\beta}_0$
2. $\boldsymbol{\beta}_0 \in int(B)$ and $m(\boldsymbol{x}_i, \boldsymbol{\beta}) \equiv (y_i - \mu(\boldsymbol{x}_i' \boldsymbol{\beta}))^2$ is 3 times continuously differentiable.
3. $\Sigma_0 = \mathbb{V}[\frac{\partial}{\partial \beta} m(\boldsymbol{x}_i; \beta_0)] < \infty$ and $H_0 = \mathbb{E}[\frac{\partial^2}{\partial \beta \partial \beta'} m(\boldsymbol{x}_i; \beta_0)]$ is full rank (and therefore invertible).

Now, the FOC for the M-estimation problem is

$$0 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \mu(\boldsymbol{x}_i' \boldsymbol{\beta})) \dot{\mu}(\boldsymbol{x}_i' \boldsymbol{\beta})) \boldsymbol{x}_i \tag{1}$$

where $\dot{\mu} = \frac{\partial}{\partial \beta} \mu(\boldsymbol{x}_i' \boldsymbol{\beta})$. So, we've converted the M-estimation problem into a Z-estimation problem. Then we can use the standard asymptotic normality result to arrive at a precise form of the asymptotic variance:

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \to_d \mathcal{N}(0, H_0^{-1} \Sigma_0 H_0^{-1}).$$

Now, taking the second derivative gives the Hessian

$$\begin{aligned}
H_0 &= \mathbb{E}[\frac{\partial^2}{\partial \beta \partial \beta'} m(\boldsymbol{x}_i; \beta_0)] \\
&= \mathbb{E}\left[-\dot{\mu}(\boldsymbol{x}_i' \boldsymbol{\beta}_0)) \dot{\mu}(\boldsymbol{x}_i' \boldsymbol{\beta}_0)) \boldsymbol{x}_i \boldsymbol{x}_i' + (y_i - \mu(\boldsymbol{x}_i' \boldsymbol{\beta}_0)) \ddot{\mu}(\boldsymbol{x}_i' \boldsymbol{\beta}_0)) \boldsymbol{x}_i \boldsymbol{x}_i'\right] \\
&= -\mathbb{E}[\dot{\mu}(\boldsymbol{x}_i' \boldsymbol{\beta}_0))^2 \boldsymbol{x}_i \boldsymbol{x}_i']
\end{aligned}$$

by LIE. And, the variance of the score is

$$
\begin{aligned}
\Sigma_0 &= \mathbb{V}[\frac{\partial}{\partial \beta} m(\boldsymbol{x}_i; \beta_0)] \\
&= \mathbb{E}\left[(y_i - \mu(\boldsymbol{x}_i'\boldsymbol{\beta}_0))^2 \dot{\mu}(\boldsymbol{x}_i'\boldsymbol{\beta}_0))^2 \boldsymbol{x}_i \boldsymbol{x}_i'\right] \\
&= \mathbb{E}[\sigma^2(\boldsymbol{x}_i)\dot{\mu}(\boldsymbol{x}_i'\boldsymbol{\beta}_0))^2 \boldsymbol{x}_i \boldsymbol{x}_i']
\end{aligned}
$$

again by LIE. Then we have the asymptotic variance

$$
\boldsymbol{V}_0 = H_0^{-1}\Sigma_0 H_0^{-1}.
$$

1.3 Variance estimator under heteroskedasticity

Under heteroskedasticity we can use the sandwich variance estimator

$$
\widehat{\boldsymbol{V}}_{HC} = \hat{H}^{-1}\hat{\Sigma}\hat{H}^{-1},
$$

where

$$
\begin{aligned}
\hat{H} &= \frac{1}{n}\sum_{i=1}^{n} \dot{\mu}(\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}))^2 \boldsymbol{x}_i \boldsymbol{x}_i' \\
\hat{\Sigma} &= \frac{1}{n}\sum_{i=1}^{n} \hat{e}_i^2 \dot{\mu}(\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}))^2 \boldsymbol{x}_i \boldsymbol{x}_i'
\end{aligned}
$$

Now, to get an asymptotically valid CI for $||\boldsymbol{\beta}_0||^2$ we need to use the Delta Method. First, note that:

$$
\begin{aligned}
&||\boldsymbol{\beta}_0||^2 = \boldsymbol{\beta}_0'\boldsymbol{\beta}_0 \\
\implies &\frac{\partial}{\partial \beta}||\boldsymbol{\beta}_0||^2 = 2\boldsymbol{\beta}_0
\end{aligned}
$$

Then, using the Delta Method

$$
\begin{aligned}
\sqrt{n}(||\hat{\boldsymbol{\beta}}||^2 - ||\boldsymbol{\beta}_0||^2) &\to_d 2\boldsymbol{\beta}_0 \mathcal{N}(0, \boldsymbol{V}_0) \\
&= \mathcal{N}(0, 4\boldsymbol{\beta}_0' \boldsymbol{V}_0 \boldsymbol{\beta}_0)
\end{aligned}
$$

Thus, an asymptotically valid 95% CI for $||\boldsymbol{\beta}_0||^2$ is

$$
CI_{95} = \left[\hat{\boldsymbol{\beta}} - 1.96\sqrt{\frac{4\hat{\boldsymbol{\beta}}'\widehat{\boldsymbol{V}}_{HC}\hat{\boldsymbol{\beta}}}{n}}, \hat{\boldsymbol{\beta}} + 1.96\sqrt{\frac{4\hat{\boldsymbol{\beta}}'\widehat{\boldsymbol{V}}_{HC}\hat{\boldsymbol{\beta}}}{n}}\right]
$$

1.4 Variance estimator under homoskedasticity

Using the above results, under homoskedasticity, the asymptotic variance collapses to

$$\begin{aligned}\boldsymbol{V}_0 &= \mathbb{E}[\dot{\mu}(\boldsymbol{x}_i'\boldsymbol{\beta}_0))^2\boldsymbol{x}_i\boldsymbol{x}_i']^{-1}\sigma^2\mathbb{E}[\dot{\mu}(\boldsymbol{x}_i'\boldsymbol{\beta}))^2\boldsymbol{x}_i\boldsymbol{x}_i']\mathbb{E}[\dot{\mu}(\boldsymbol{x}_i'\boldsymbol{\beta}_0))^2\boldsymbol{x}_i\boldsymbol{x}_i']^{-1}\\ &= \sigma^2\mathbb{E}[\dot{\mu}(\boldsymbol{x}_i'\boldsymbol{\beta}_0))^2\boldsymbol{x}_i\boldsymbol{x}_i']^{-1}\end{aligned}$$

The variance estimator is now takes a simpler form

$$\widehat{\boldsymbol{V}}_{HO} = \hat{\sigma}^2\hat{H}^{-1}$$

where $\hat{H}$ is the same as above and

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

Then, as above, the asymptotically valid 95% CI for $||\boldsymbol{\beta}_0||^2$ is

$$CI_{95} = \left[\hat{\boldsymbol{\beta}} - 1.96\sqrt{\frac{4\hat{\boldsymbol{\beta}}'\widehat{\boldsymbol{V}}_{HO}\hat{\boldsymbol{\beta}}}{n}}, \hat{\boldsymbol{\beta}} + 1.96\sqrt{\frac{4\hat{\boldsymbol{\beta}}'\widehat{\boldsymbol{V}}_{HO}\hat{\boldsymbol{\beta}}}{n}}\right].$$

1.5 MLE

Given the assumption of a normal DGP we have the conditional density

$$f(y_i|\boldsymbol{x}_i) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\Big(-\frac{1}{2\sigma^2}(y_i - \mu(\boldsymbol{x}_i'\boldsymbol{\beta}_0))^2\Big).$$

Then, the sample log-likelihood function is

$$\ell(\boldsymbol{\beta}, \sigma^2; \boldsymbol{X}) = n\log\Big(\frac{1}{\sqrt{2\pi\sigma^2}}\Big) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu(\boldsymbol{x}_i'\boldsymbol{\beta}))^2$$

Dividing by n gives

$$\ell(\boldsymbol{\beta}, \sigma^2; \boldsymbol{X}) = \log\Big(\frac{1}{\sqrt{2\pi\sigma^2}}\Big) - \frac{1}{n2\sigma^2}\sum_{i=1}^{n}(y_i - \mu(\boldsymbol{x}_i'\boldsymbol{\beta}))^2$$

The FOC wrt $\boldsymbol{\beta}$ is

$$0 = \frac{1}{n\sigma^2}\sum_{i=1}^{n}(y_i - \mu(\boldsymbol{x}_i'\boldsymbol{\beta}))\dot{\mu}(\boldsymbol{x}_i'\boldsymbol{\beta}))\boldsymbol{x}_i,$$

which is equivalent to the FOC for the M-estimation problem (1) (since σ^2 just scales the FOC, it does not affect the solution). Thus,

$$\hat{\boldsymbol{\beta}}_{MLE} = \hat{\boldsymbol{\beta}}_{M.est}.$$

Now, the FOC of the log-likelihood wrt σ^2 is

$$0 = -\frac{1}{2}(2\pi\sigma^2)^{-1}2\pi + \frac{1}{2n}(\sigma^2)^{-2}\sum_{i=1}^{n}(y_i - \mu(\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}))^2$$

Solving for σ^2 gives the MLE:

$$\hat{\sigma}^2_{ML} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mu(\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}))^2,$$

which is not the same as the estimator proposed in [4], since it does not adjust for the number of regressors.

1.6 When the link function is unknown

Suppose the link function is unknown, and consider two pairs of true parameters, $(\mu_1, \boldsymbol{\beta}_1)$ and $(\mu_2, \boldsymbol{\beta}_2)$ where $\mu_2(u) = \mu_1(u/c)$ and $\boldsymbol{\beta}_2 = c\boldsymbol{\beta}_1$ for some $c \neq 0$. Then the parameters are clearly different, but $\mu_1(\boldsymbol{x}_i'\boldsymbol{\beta}_1) = \mu_2(\boldsymbol{x}_i'\boldsymbol{\beta}_2)$.

1.7 Logistic link function

The link function is

$$\begin{aligned}
\mu(\boldsymbol{x}_i'\boldsymbol{\beta}_0) &= \mathbb{E}[y_i|\boldsymbol{x}_i] \\
&= \mathbb{E}[\mathbf{1}(\boldsymbol{x}_i'\boldsymbol{\beta}_0 \geq \epsilon_i)|\boldsymbol{x}_i] \\
&= \Pr[\boldsymbol{x}_i'\boldsymbol{\beta}_0 \geq \epsilon_i|\boldsymbol{x}_i] \\
&= F(\boldsymbol{x}_i'\boldsymbol{\beta}_0) \\
&= \frac{1}{1+\exp(-\boldsymbol{x}_i'\boldsymbol{\beta}_0)}, \text{ if } s_0 = 1.
\end{aligned}$$

The conditional variance of y_i is

$$\sigma^2(\boldsymbol{x}_i)\mathbb{V}[y_i|\boldsymbol{x}_i]$$

Now, note that $y_i|\boldsymbol{x}_i$ is a Bernoulli random variable, with $\Pr[y_i = 1|\boldsymbol{x}_i] = F(\boldsymbol{x}_i'\boldsymbol{\beta}_0)$. Then

$$\begin{aligned}
\sigma^2(\boldsymbol{x}_i) &= F(\boldsymbol{x}_i'\boldsymbol{\beta}_0)(1 - F(\boldsymbol{x}_i'\boldsymbol{\beta}_0)) \\
&= \mu(\boldsymbol{x}_i'\boldsymbol{\beta}_0)(1 - \mu(\boldsymbol{x}_i'\boldsymbol{\beta}_0))
\end{aligned}$$

To derive an expression for the asymptotic variance, first note that for the logistic cdf: $\dot{\mu}(u) = (1-\mu(u))\mu(u)$. Then, the asymptotic variance is

$$\boldsymbol{V}_0 = H_0^{-1}\Sigma_0 H_0^{-1}.$$

where

$$H_0 = \mathbb{E}[(1 - \mu(\boldsymbol{x}_i'\boldsymbol{\beta}_0))^2\mu(\boldsymbol{x}_i'\boldsymbol{\beta}_0)^2\boldsymbol{x}_i\boldsymbol{x}_i']$$

and

$$\Sigma_0 = \mathbb{E}[(1 - \mu(\boldsymbol{x}_i'\boldsymbol{\beta}_0))^3\mu(\boldsymbol{x}_i'\boldsymbol{\beta}_0)^3\boldsymbol{x}_i\boldsymbol{x}_i']$$

1.8 Logistic link function, MLE

MLE gives the same point estimator as NLS (i.e. they have the same FOC; we did this in 672), but MLE is asymptotically efficient, so $\boldsymbol{V}_0^{ML} \leq \boldsymbol{V}_0^{NLS}$.

1.9 Some data work

(a) I estimated the logistic model with robust (HC1) standard errors in both `R` and `Stata`. The results from `R` are presented in Table 1. The standard errors from `Stata` are very slightly different, but I'm not sure why.

Table 1: **Logistic Regression Estimates for** $s = 1-$`dmissing`

	Coef.	Std. Err.	t-stat	p-val	CI.lower	CI.upper
Const.	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
$\log(\text{inc}+1)$	-0.119	0.044	-2.707	0.007	-0.205	-0.033

(b) Table 2 presents the 95% confidence interval and p-values for each coefficient derived from 999 bootstrap replications of the t-statistic: $t^* = (\beta^* - \hat{\beta}_{obs})/se^*$. The statistics are very similar to those in Table 1, which rely on large sample approximations.

The idea for computing bootstrapped CIs is simple: for each bootstrap replication, compute t^* for each coefficient; this gives an empirical distribution for t^*; then extract the desired quantiles from the empirical distribution, and compute the confidence intervals as

$$CI_{95}^{boot}(\beta) = \left[\hat{\beta}_{obs} + q_{0.025}^* \times \hat{se}_{obs},\ \hat{\beta}_{obs} + q_{0.975}^* \times \hat{se}_{obs}\right]$$

I computed the bootstrapped p-values as

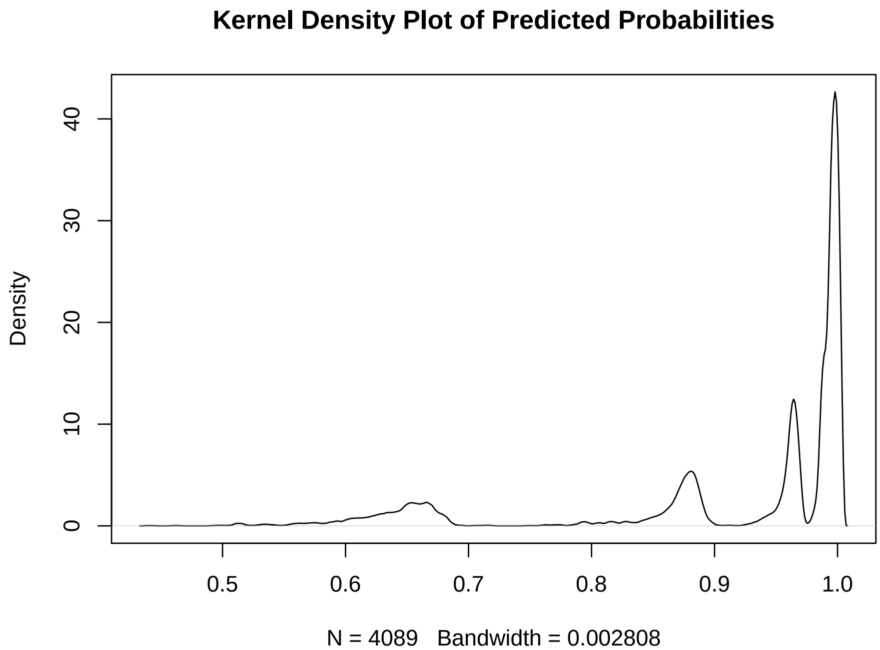
$$p^{boot} = \frac{1}{M}\sum_{i=1}^{M} \mathbf{1}[t^* \geq t_{obs}]$$

where M is the number of bootstrap replications.

Table 2: **Bootstrap Statistics for the Logistic Model of** $s = 1-$`dmissing`

	Coef.	CI.lower	CI.upper	p-val
Const.	1.755	1.157	2.471	0.000
S_age	1.333	1.142	1.609	0.000
S_HHpeople	-0.067	-0.112	-0.020	0.001
$\log(\text{inc}+1)$	-0.119	-0.216	-0.042	0.001

(c) I plot the kernel density estimate of the predicted probabilities of reporting data, $\hat{\mu}(\boldsymbol{x}_i'\hat{\boldsymbol{\beta}})$, using an Epanechnikov kernel with `R`'s unbiased cross-validation bandwidth.

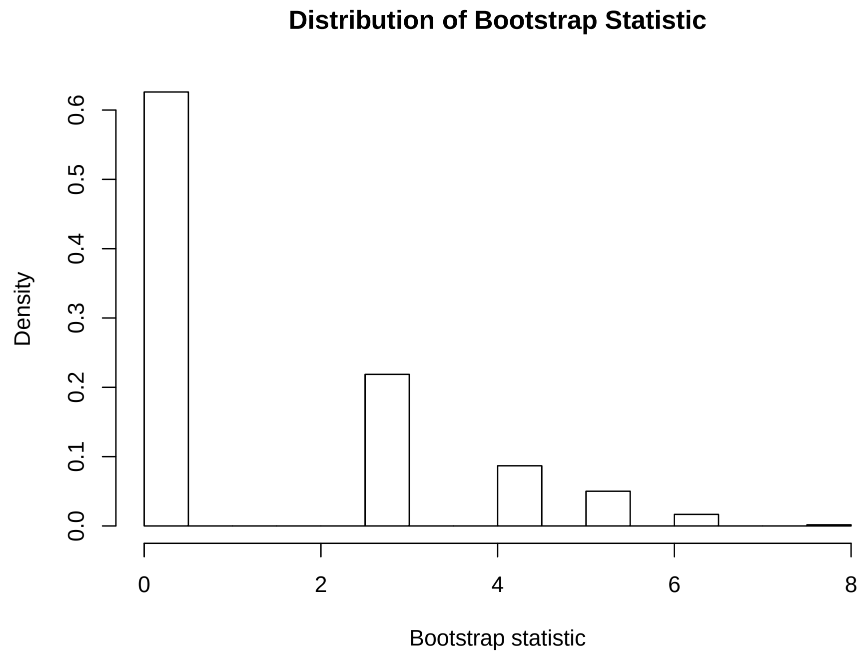


2 Semiparametric GMM with missing data

3 When bootstrap fails

3.1

I plot the empirical distribution of the bootstrap statistic, $n(\max\{x_i\} - \max\{x_i^*\})$, below. Clearly, the empirical distribution does not coincide with the theoretical Exponential (1) distribution.



3.2

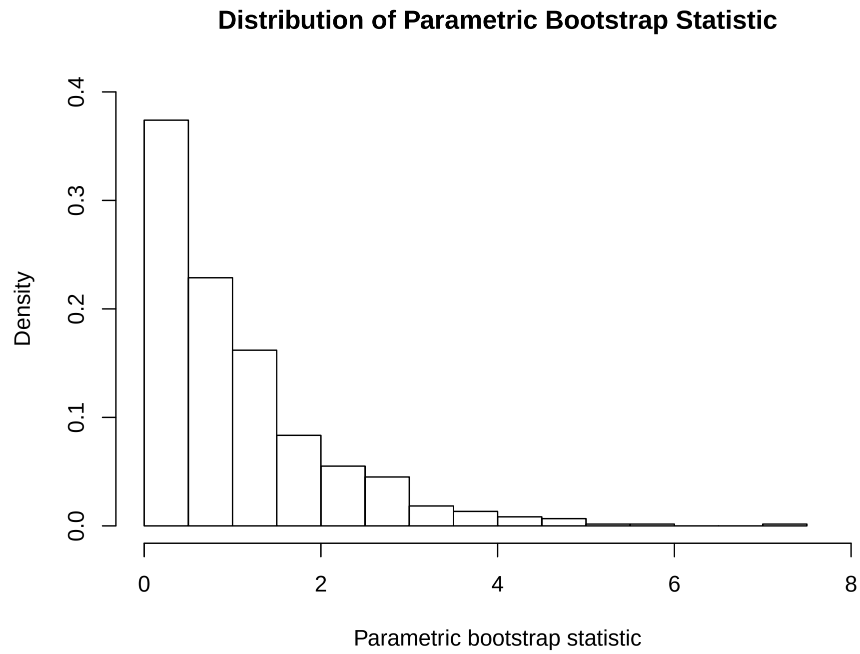
Now, consider the parametric bootstrap statistic, $t_{\mathtt{P}}^* = n(\max\{x_i\} - \max\{x_i^*\})$, where

$$x_i^* \sim_{iid} \mathtt{Uniform}[0, \max\{x_i\}].$$

I plot the empirical distribution of $t_{\mathtt{P}}^*$ below. Now, the empirical distribution *does* seem to coincide with the theoretical Exponential (1) distribution.

3.3

In the nonparametric case, the bootstrap statistic has a mass point at zero since $\Pr[\max\{x_i\} = \max\{x_i^*\}]$ converges to 1. However, the parametric bootstrap corrects for this "bias": in this case $\Pr[\max\{x_i\} = \max\{x_i^*\}] = 0$, since $x_i^* \sim_{iid} \mathtt{Uniform}[0, \max\{x_i\}]$.



4 Appendix

4.1 `R` code

4.2 `STATA` code