

# ECON675 – Assignment 3

Anirudh Yadav

October 26, 2018

## Contents

<b>1</b>	<b>Non-linear least squares</b>	<b>2</b>
1.1	Identifiability . . . . .	2
1.2	Asymptotic normality . . . . .	2
1.3	Variance estimator under heteroskedasticity . . . . .	3
1.4	Variance estimator under homoskedasticity . . . . .	4
1.5	MLE . . . . .	4
1.6	When the link function is unknown . . . . .	5
1.7	Logistic link function . . . . .	5
1.8	Logistic link function, MLE . . . . .	6
1.9	Some data work . . . . .	6
<b>2</b>	<b>Semiparametric GMM with missing data</b>	<b>7</b>
<b>3</b>	<b>When bootstrap fails</b>	<b>8</b>
<b>4</b>	<b>Appendix</b>	<b>9</b>
4.1	R code . . . . .	9
4.2	STATA code . . . . .	9

# 1 Non-linear least squares

## 1.1 Identifiability

This is a standard M-estimation problem. The parameter vector  $\beta_0$  is assumed to solve the population problem

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

For  $\beta_0$  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  $\delta > 0$ :

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

where  $M(\beta) = \mathbb{E}[(y_i - \mu(\mathbf{x}_i' \beta))^2]$ . Of course  $\beta_0$  can be written in closed form if  $\mu(\cdot)$  is linear. In this case, we know that

$$\beta_0 = \mathbb{E}[\mathbf{x}_i \mathbf{x}_i']^{-1} \mathbb{E}[\mathbf{x}_i y_i].$$

## 1.2 Asymptotic normality

The M-estimator is asymptotically normal if:

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

Now, the FOC for the M-estimation problem is

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

where  $\dot{\mu} = \frac{\partial}{\partial \beta} \mu(\mathbf{x}_i' \beta)$ . 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:

$$\sqrt{n}(\hat{\beta} - \beta_0) \rightarrow_d \mathcal{N}(0, H_0^{-1} \Sigma_0 H_0^{-1}).$$

Now, taking the second derivative gives the Hessian

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

by LIE. And, the variance of the score is

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

again by LIE. Then we have the asymptotic variance

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

### 1.3 Variance estimator under heteroskedasticity

Under heteroskedasticity we can use the sandwich variance estimator

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

where

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

Now, to get an asymptotically valid CI for  $||\beta_0||^2$  we need to use the Delta Method. First, note that:

$$\begin{aligned}||\beta_0||^2 &= \beta'_0 \beta_0 \\ \implies \frac{\partial}{\partial\beta} ||\beta_0||^2 &= 2\beta_0\end{aligned}$$

Then, using the Delta Method

$$\begin{aligned}\sqrt{n}(|\hat{\beta}|^2 - ||\beta_0||^2) &\rightarrow_d 2\beta_0 \mathcal{N}(0, \mathbf{V}_0) \\ &= \mathcal{N}(0, 4\beta'_0 \mathbf{V}_0 \beta_0)\end{aligned}$$

Thus, an asymptotically valid 95% CI for  $||\beta_0||^2$  is

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

## 1.4 Variance estimator under homoskedasticity

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

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

The variance estimator is now takes a simpler form

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

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

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

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

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

## 1.5 MLE

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

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

Then, the sample log-likelihood function is

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

Dividing by  $n$  gives

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

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

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

which is equivalent to the FOC for the M-estimation problem (1) (since  $\sigma^2$  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  $\sigma^2$  is

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

Solving for  $\sigma^2$  gives the MLE:

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

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  $\mu_1(\mathbf{x}'_i \boldsymbol{\beta}_1) = \mu_2(\mathbf{x}'_i \boldsymbol{\beta}_2)$ .

## 1.7 Logistic link function

The link function is

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

The conditional variance of  $y_i$  is

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

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

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

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

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

where

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

and

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

## 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  $\mathbf{V}_0^{ML} \leq \mathbf{V}_0^{NLS}$ .

## 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  $s = 1 - \text{dmissing}$**

	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(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

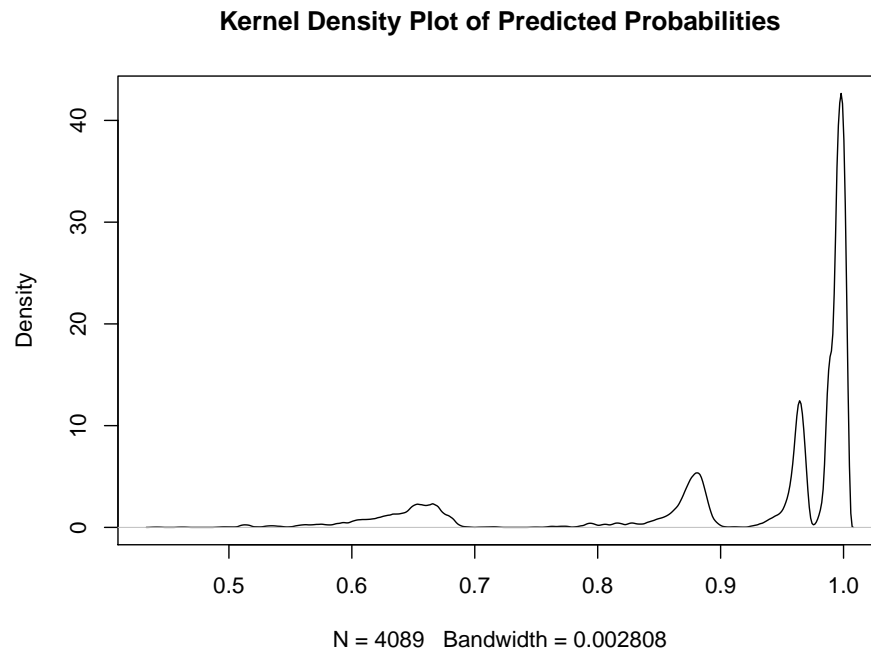
$$p^{boot} = \frac{1}{M} \sum_{i=1}^M \mathbf{1}[t^* \geq t_{obs}]$$

where  $M$  is the number of bootstrap replications.

Table 2: **Bootstrap Statistics for the Logistic Model of  $s = 1 - \text{dmissing}$**

	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(inc + 1)	-0.119	-0.216	-0.042	0.001

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



## 2 Semiparametric GMM with missing data

### 3 When bootstrap fails



## 4 Appendix

### 4.1 R code

### 4.2 STATA code