Econ 675 Assignment 1

Nathan Mather*

October 25, 2018

Contents

1 Question 1: Non-linear Least Squares

1.1 Q1 Part 1

The general non-linear least squares estimator is

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

Now for $\boldsymbol{\beta}_0 = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^d} \mathrm{E}[(y_i - \mu(\boldsymbol{x_i'}\boldsymbol{\beta}))^2]$ to be identifiable we need:

$$\boldsymbol{\beta}_0 = \boldsymbol{\beta}_0^*$$

$$\iff \boldsymbol{\beta}_0^* = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^d} \mathrm{E}[(y_i - \mu(\boldsymbol{x_i'}\boldsymbol{\beta}))^2]$$

To find this start by noting that

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

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

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

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

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

$$\mathrm{E}[(y_i - \mu(\boldsymbol{x}_i'\boldsymbol{\beta}_0))(\mu(\boldsymbol{x}_i'\boldsymbol{\beta}_0) - \mu(\boldsymbol{x}_i'\boldsymbol{\beta}))] = \mathrm{E}[\mathrm{E}[(y_i - \mu(\boldsymbol{x}_i'\boldsymbol{\beta}_0))(\mu(\boldsymbol{x}_i'\boldsymbol{\beta}_0) - \mu(\boldsymbol{x}_i'\boldsymbol{\beta}))]|\boldsymbol{x}_i]$$

$$= \mathrm{E}[(\mathrm{E}[y_i|\boldsymbol{x}_i] - \mu(\boldsymbol{x}_i'\boldsymbol{\beta}_0))(\mu(\boldsymbol{x}_i'\boldsymbol{\beta}_0) - \mu(\boldsymbol{x}_i'\boldsymbol{\beta}))] = 0$$

Using this fact we have that

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

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

This is strictly greater than unless $\exists \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ such that $\mathrm{E}[(\mu(\boldsymbol{x}_i'\boldsymbol{\beta}_0) - \mu(\boldsymbol{x}_i'\boldsymbol{\beta}))^2] = 0$ Thus this give us an identification condition that $\mathrm{E}[(\mu(\boldsymbol{x}_i'\boldsymbol{\beta}_0) - \mu(\boldsymbol{x}_i'\boldsymbol{\beta}))^2] \neq 0 \ \forall \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$. This means that $\boldsymbol{\beta}_0$ is the unique minimizer of $\mathrm{E}[(y_i - \mu(\boldsymbol{x_i'\beta}))^2]$
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 $\boldsymbol{\beta}_0 = \mathrm{E}[\boldsymbol{x}_i\boldsymbol{x}_i']^{-1}\mathrm{E}[\boldsymbol{x}_i y_i]$

1.2 Q1 Part 2

In order to set this up as a Z estimator lets take a first order condition. This gives use the following condition.

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

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

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

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

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

Were

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

and

$$\boldsymbol{\Sigma_0} = \mathrm{V}[m(\boldsymbol{z}_i, \boldsymbol{\beta}_0)] = \mathrm{E}[\sigma^2(\boldsymbol{x_i})\dot{\mu}(\boldsymbol{x_i'\beta_o})^2 \boldsymbol{x_i x_i'}]$$

1.3 Q1 Part 3

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

where $\hat{\boldsymbol{m}} = \boldsymbol{m_\beta(z_i, \hat{\beta})}$ and $hate = y_i - \boldsymbol{m(z_i, \hat{\beta})}$
Now by the delta method and letting $g(\boldsymbol{\beta}) = \|\boldsymbol{\beta}\| = \sum_{k=1}^{d} \boldsymbol{\beta}_k^2$ we get

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

where $\dot{g}(\boldsymbol{\beta}_0) = \frac{d}{d\boldsymbol{\beta}'}g(\boldsymbol{\beta}) = 2\boldsymbol{\beta}'$ Hence the confidence interval is given by

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

1.4 Q1 Part 4

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

$$\boldsymbol{V}_0 = \sigma^2\boldsymbol{H}_0^{-1} = \sigma^2\mathrm{E}[\dot{\mu}(\boldsymbol{x}_i'\boldsymbol{\beta}_0)^2\boldsymbol{x_i}\boldsymbol{x_i'}]^{-1}$$

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

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

and

$$\hat{\boldsymbol{H}} = \frac{a}{n}\sum_{i=1}^{n}\dot{\mu}(\boldsymbol{x}_i'\boldsymbol{\beta}_0)^2\boldsymbol{x_i}\boldsymbol{x_i'}$$

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

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

where $\dot{g}(\boldsymbol{\beta}_0) = \frac{d}{d\boldsymbol{\beta}'}g(\boldsymbol{\beta}) = 2\boldsymbol{\beta}'$ Hence the confidence interval is given by

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

1.5 Q1 Part 5

The conditional likelihod funciton is

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

with log likelihood

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

This gives us the following first order conditions

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

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

These conditions are equivalent to those found above.

1.6 Q1 Part 6

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

1.7 Q1 Part 7

2 Q2

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 outside 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 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

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

Taking first order conditions and setting equal to zero we get

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

or

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

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

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

Which can be rearanged 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, $\sqrt{n}(\hat{\theta} - \theta_0) \to_d N(0, V_0)$

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

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

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

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

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

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

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

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)

3 Question 3: When Bootstrap Fails