Problem Set 7

1. The quadratic objective function $q(\mathbf{w}, \boldsymbol{\theta}) = (y - \mathbf{x}\boldsymbol{\theta})^2$ leads to the sample minimization problem
$$\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{1}{N}\sum_{i=1}^{N}(y_i - \mathbf{x}_i\boldsymbol{\theta})^2 = \min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{1}{N}\mathbf{u}'\mathbf{u}.$$
This is identical to the OLS minimizazion problem and is even more obvious when comparing the FOC's. The score of the M-estimator is
$$\mathbf{s}(\mathbf{w}, \boldsymbol{\theta}) = -2\mathbf{x}'(y - \mathbf{x}\boldsymbol{\theta}) = -2\mathbf{x}'u$$
which leads to the sample FOC
$$-2\sum_{i=1}^{N}\mathbf{x}_i'\hat{u}_i = 0 \Leftrightarrow \sum_{i=1}^{N}\mathbf{x}_i'\hat{u}_i = 0.$$
This is exactly the FOC of the OLS estimator. Hence, the OLS estimator coincides with the M-estimator. But what about the asymptotic distribution? To this end, we need to find the Hessian of the M-estimator. Taking the first derivative of the score yields
$$\mathbf{H}(\mathbf{w}, \boldsymbol{\theta}) = 2\mathbf{x}'\mathbf{x}.$$
Based on these results, we find
$$\mathbf{B}_0 = E[\mathbf{s}(\mathbf{w}, \boldsymbol{\theta})\mathbf{s}(\mathbf{w}, \boldsymbol{\theta})'] = 4E[u^2\mathbf{x}'\mathbf{x}]$$
and
$$\mathbf{A}_0 = E[\mathbf{H}(\mathbf{w}, \boldsymbol{\theta})] = 2E[\mathbf{x}'\mathbf{x}]$$
The asymptotic distribution of the M-estimator is
$$N^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} \mathcal{N}(0, \mathbf{V}),$$
where
$$\mathbf{V} = \mathbf{A}_0^{-1}\mathbf{B}_0\mathbf{A}_0^{-1} = (E[\mathbf{x}'\mathbf{x}])^{-1}E[u^2\mathbf{x}'\mathbf{x}](E[\mathbf{x}'\mathbf{x}])^{-1}.$$
This is the same asymptotic distribution as the one found in previous lectures for the OLS estimator. Note that OLS is just a special case of M-estimation if one chooses the above defined objective function.

2. (a) Objective function:
$$q(\mathbf{w}, \boldsymbol{\theta}) = \frac{[y - m(\mathbf{x}, \boldsymbol{\theta}))]^2}{2} = \frac{u^2}{2} = \frac{[y - \exp(\mathbf{x}\boldsymbol{\theta})]^2}{2}$$
Score:
$$\mathbf{s}(\mathbf{w}, \boldsymbol{\theta}) = \frac{\partial q(\mathbf{w}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -[y - \exp(\mathbf{x}\boldsymbol{\theta})]\exp(\mathbf{x}\boldsymbol{\theta})\mathbf{x}' = -e^{\mathbf{x}\boldsymbol{\theta}}\mathbf{x}'u$$
Hessian:
$$\mathbf{H}(\mathbf{w}, \boldsymbol{\theta}) = \frac{\partial^2 q(\mathbf{w}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} = -e^{\mathbf{x}\boldsymbol{\theta}}\mathbf{x}'\mathbf{x}u + e^{2\mathbf{x}\boldsymbol{\theta}}\mathbf{x}'\mathbf{x}$$

(b) Note that $E[e^{\mathbf{x}\boldsymbol{\theta}}\mathbf{x}'\mathbf{x}u] = 0$ by the LIE, given $E[u|\mathbf{x}] = 0$. Hence,

$$\mathbf{A}_0 = E[\mathbf{H}(\mathbf{w}, \boldsymbol{\theta})] = E[e^{2\mathbf{x}\boldsymbol{\theta}}\mathbf{x}'\mathbf{x}]$$

and

$$\mathbf{B}_0 = E[\mathbf{s}(\mathbf{w}, \boldsymbol{\theta})\mathbf{s}(\mathbf{w}, \boldsymbol{\theta})'] = E[e^{2\mathbf{x}\boldsymbol{\theta}}u^2\mathbf{x}'\mathbf{x}]$$

Estimate using sample counterparts

$$\hat{\mathbf{A}}_0 = \frac{1}{N}\sum_{i=1}^{N} e^{2\mathbf{x}_i\hat{\boldsymbol{\theta}}}\mathbf{x}_i'\mathbf{x}_i$$

and

$$\hat{\mathbf{B}}_0 = \frac{1}{N}\sum_{i=1}^{N} e^{2\mathbf{x}_i\hat{\boldsymbol{\theta}}}\hat{u}_i^2\mathbf{x}_i'\mathbf{x}_i$$

(c) Marginal effect:

$$\psi(\mathbf{x}, \boldsymbol{\theta}) = \frac{\partial E(y|\mathbf{x})}{\partial x_1} = \frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial x_1} = e^{\mathbf{x}\boldsymbol{\theta}}\theta_1$$

Elasticity:

$$\rho(\mathbf{x}, \boldsymbol{\theta}) = \frac{\partial E(y|\mathbf{x})}{\partial x_1}\frac{x_1}{E(y|\mathbf{x})} = \frac{\partial m(\mathbf{x}, \boldsymbol{\theta})}{\partial x_1}\frac{x_1}{m(\mathbf{x}, \boldsymbol{\theta})} = e^{\mathbf{x}\boldsymbol{\theta}}\theta_1\frac{x_1}{e^{\mathbf{x}\boldsymbol{\theta}}} = \theta_1 x_1$$

(d) The point estimate is

$$\psi(\mathbf{x}, \hat{\boldsymbol{\theta}}) = exp(0.693147 + 1 \cdot 0) = 2.$$

To find the distribution of the sample function $\psi(\mathbf{x}, \hat{\boldsymbol{\theta}})$, which is (given on $x_1 = 0$) a function of $\hat{\boldsymbol{\theta}}$, use the delta method:

$$\sqrt{N}\big(\psi(\mathbf{x}, \hat{\boldsymbol{\theta}}) - \psi(\mathbf{x}, \boldsymbol{\theta})\big) \xrightarrow{d} \mathcal{N}(0, \nabla_{\hat{\boldsymbol{\theta}}}\psi\big(\mathbf{x}, \hat{\boldsymbol{\theta}})\mathbf{V}\nabla_{\hat{\boldsymbol{\theta}}}'\psi(\mathbf{x}, \hat{\boldsymbol{\theta}})\big)$$

The gradient is

$$\begin{aligned}\nabla_{\boldsymbol{\theta}}\psi(\mathbf{x}, \hat{\boldsymbol{\theta}}) &= \left(\frac{\partial\psi(\mathbf{x}, \hat{\boldsymbol{\theta}})}{\partial\theta_0}, \ \frac{\partial\psi(\mathbf{x}, \hat{\boldsymbol{\theta}})}{\partial\theta_1}\right) = \left(e^{\mathbf{x}\boldsymbol{\theta}}\theta_1, \ e^{\mathbf{x}\boldsymbol{\theta}}\theta_1 x_1 + e^{\mathbf{x}\boldsymbol{\theta}}\right) \\ &= e^{\mathbf{x}\boldsymbol{\theta}}(\theta_1, \ \theta_1 x_1 + 1).\end{aligned}$$

Hence, the required variance is

$$\mathbf{V}_\psi = e^{2\mathbf{x}\boldsymbol{\theta}}(\theta_1, \theta_1 x_1 + 1)\mathbf{V}\begin{pmatrix}\theta_1 \\ \theta_1 x_1 + 1\end{pmatrix}$$

which is estimated as

$$\hat{\mathbf{V}}_\psi = e^{2ln(2)}(1, 1)\begin{pmatrix}0.15 & 0.07 \\ 0.07 & 0.2\end{pmatrix}\begin{pmatrix}1 \\ 1\end{pmatrix} = 4 \cdot 0.49 = 1.96.$$

The approximate distribution of $\psi(\mathbf{x}, \hat{\boldsymbol{\theta}})$ (given on $x_1 = 0$) is thus

$$\psi(\mathbf{x}, \hat{\boldsymbol{\theta}}) \sim \mathcal{N}\left(\psi(\mathbf{x}, \boldsymbol{\theta}), \frac{1.96}{N}\right),$$

from which we obtain the standard error $\sqrt{1.96/N} = \sqrt{1.96/49} = 0.2$. Now we construct a 95 percent confidence interval using the 97.5 percent quantile of the standard normal distribution, which yields

$$\begin{aligned}\mathrm{CI}_\psi &= \left[\psi(\mathbf{x}, \hat{\boldsymbol{\theta}}) \pm z_{1-\frac{\alpha}{2}} \cdot se(\psi(\mathbf{x}, \hat{\boldsymbol{\theta}}))\right] = [2 \pm 1.96 \cdot 0.2] \\ &= [1.608, 2.392].\end{aligned}$$

(e) The point estimate is

$$\hat{\gamma} = \gamma(\hat{\boldsymbol{\theta}}) = 0.6931^2 - 1^2 = -0.5195.$$

To obtain the asymptotic distribution use the Delta Method. For doing so we need the gradient

$$\nabla_{\boldsymbol{\theta}}\gamma(\boldsymbol{\theta}) = (2\theta_0, -2\theta_1)$$

The asymptotic variance is

$$\mathbf{V}_\gamma = (2\theta_0, -2\theta_1)\mathbf{V}\begin{pmatrix} 2\theta_0 \\ -2\theta_1 \end{pmatrix}$$

which is estimated as

$$\hat{\mathbf{V}}_\gamma = (2 \cdot 0.6931, -2)\begin{pmatrix} 0.15 & 0.07 \\ 0.07 & 0.2 \end{pmatrix}\begin{pmatrix} 2 \cdot 0.6931 \\ -2 \end{pmatrix} = 0.7$$

The asymptotic distribution of $\gamma(\hat{\boldsymbol{\theta}})$ is

$$\gamma(\hat{\boldsymbol{\theta}}) \overset{a}{\sim} \mathcal{N}\left(\gamma(\boldsymbol{\theta}), \frac{0.7}{N}\right)$$

from which we obtain the estimated asymptotic standard error $\widehat{se(\gamma)} = \sqrt{0.7/N} = \sqrt{0.7/49} = 0.1195$.

The 95% confidence interval is

$$CI_\gamma = [-0.5195 \pm 1.96 \cdot 0.1195] = [-0.7537, -0.2853].$$

(f) Test the two-sided null hypothesis $\gamma = 0$ using both a t-test and a Wald test.

i. Start with a t-test.

Hypothesis:

$$H_0: \ \gamma = 0.$$

Test statistic:

$$t = \frac{\hat{\gamma}}{se(\gamma)} = \frac{-0.5195}{0.1195} = -4.35$$

Critical Value:

The critical value at the 5% level is $CV_t = 1.96$.
Test decision: Because $|t| > CV_t$ we reject H_0.

ii. Wald test:

Test statistic:

$$W = \gamma(\hat{\boldsymbol{\theta}})\left[\frac{\hat{V}}{N}\right]^{-1}\gamma(\hat{\boldsymbol{\theta}}) = (-0.5195)^2\left(\frac{0.7}{49}\right)^{-1} = 18.89$$

Critical Value:
The critical value is $CV_W = 3.84$.

Test decision:
Because $W > CV_W$ we reject H_0.

3.

$$q(\mathbf{w}, \boldsymbol{\theta}) = \frac{1}{2} \ln(1 + u^2/k), u = y - \mathbf{x}\boldsymbol{\theta}, k > 0$$

(a) Score:

$$\mathbf{s}(\mathbf{w}, \boldsymbol{\theta}) = \frac{\partial q(\mathbf{w}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{2} \frac{2u/k}{1 + u^2/k}(-\mathbf{x}') = \frac{1}{k + u^2}(-\mathbf{x}'u) = \frac{u}{k + u^2}(-\mathbf{x}')$$

Hessian:

$$\mathbf{H}(\mathbf{w}, \boldsymbol{\theta}) = \frac{\partial^2 q(\mathbf{w}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = (-\mathbf{x}') \frac{k + u^2 - u * (2u)}{(k + u^2)^2}(-\mathbf{x}) = \frac{k - u^2}{(k + u^2)^2}\mathbf{x}'\mathbf{x}$$

(b)

$$\mathbf{s}(\mathbf{w}, \boldsymbol{\theta}) = \frac{-\mathbf{x}'u}{k + u^2}$$

$\Rightarrow$ sample FOC: $-\sum_{i=1}^{N} \frac{\mathbf{x}_i'\hat{u}_i}{k+\hat{u}_i^2} \overset{!}{=} 0 \Leftrightarrow \sum_{i=1}^{N} \frac{\mathbf{x}_i'\hat{u}_i}{k+\hat{u}_i^2} = 0$

OLS: $\sum_{i=1}^{N} \mathbf{x}_i'\hat{u}_i \overset{!}{=} 0$

$\Rightarrow$ OLS: Every deviation $\hat{u}_i = y_i - \mathbf{x}_i\hat{\boldsymbol{\theta}}$ has the same effect on the deviation of the LHS from zero (conditional on $\mathbf{x}_i$).

$\Rightarrow$ Here: Every deviation $\hat{u}_i$ from zero is weighted by $\frac{1}{k+\hat{u}_i^2}$. Hence, large deviations are down-weighted because they increase the denominator quadratically.

(c) Direct approach: If k is large such that u^2/k is small, the loss function is approximately $q(\mathbf{w}, \boldsymbol{\theta}) \approx \frac{1}{2}u^2/k$ which leads to the same estimator as the OLS loss function $q(\mathbf{w}, \boldsymbol{\theta}) = u^2$.

Taylor expansion at $u = 0$:

$f(u) = \frac{1}{2} \ln\left(1 + \frac{u^2}{k}\right), f'(u) = \frac{u}{k+u^2}, f''(u) = \frac{k-u^2}{(k+u^2)^2}$

$q(\mathbf{w}, \boldsymbol{\theta}) = f(0) + f'(0) * u + \frac{1}{2}f''(0) * u^2 = \frac{k}{2k^2}u^2 = \frac{1}{2}u^2/k$

Discussion as above.

4. (a) Piecewise differentiation

$$\frac{\partial q(\mathbf{w}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{cases} -2k\mathbf{x}', & \text{if } u > k \\ -2\mathbf{x}'u, & \text{if } -k \leq u \leq k \\ 2k\mathbf{x}', & \text{if } u < -k \end{cases}$$

What happens at $u = k$?

$$\lim_{u \downarrow k} \frac{\partial q(\mathbf{w}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -2k\mathbf{x}', \qquad \lim_{u \uparrow k} \frac{\partial q(\mathbf{w}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \lim_{u \uparrow k} -2\mathbf{x}'u = -2k\mathbf{x}'$$

$\Rightarrow$ continuously differentiable

What happens at $u = -k$?

$$\lim_{u \downarrow -k} \frac{\partial q(\mathbf{w}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \lim_{u \downarrow -k} -2\mathbf{x}'u = 2k\mathbf{x}', \qquad \lim_{u \uparrow -k} \frac{\partial q(\mathbf{w}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 2k\mathbf{x}'$$

$\Rightarrow$ continuously differentiable

(b)

$$\frac{\partial^2 q(\mathbf{w}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \begin{cases} 0, & \text{if } u > k \\ 2\mathbf{x}'\mathbf{x}, & \text{if } -k \leq u \leq k \\ 0, & \text{if } u < -k \end{cases}$$

In general (unless $\mathbf{x}'\mathbf{x} = 0$), the Hessian is discontinuous at k and $-k$.

(c) Simplified model:

$$\mathbf{s}(\mathbf{w}, \boldsymbol{\theta}) = \begin{cases} -2k, & \text{if } u > k \\ -2u, & \text{if } -k \leq u \leq k \\ 2k, & \text{if} u < -k \end{cases}, \mathbf{H}(\mathbf{w}, \boldsymbol{\theta}) = \begin{cases} 0, & \text{if } u > k \\ 2, & \text{if } -k \leq u \leq k \\ 0, & \text{if } u < -k. \end{cases}$$

Assume the starting value $\theta^{\{0\}}$ is chosen far away from $\bar{y}$. In this case, $|\hat{u}_1| = |y_1 - \theta^{\{0\}}| > k$. In this region, the score is flat. Hence, the Hessian is zero and the following parameter updating equation $\theta^{\{1\}} = \theta^{\{0\}} - \left[\mathbf{H}\left(\mathbf{w}_1, \theta^{\{0\}}\right)\right]^{-1} \cdot \mathbf{s}\left(\mathbf{w}_1, \theta^{\{0\}}\right)$ does not work because the inverse of the Hessian does not exist.

When $N > 1$, typically only few $|\hat{u}_i| > k$ and the Newton-Raphson algorithm may work. However, if $\boldsymbol{\theta}^{\{0\}}$ is chosen very badly, something similar may happen. In particular, the inverse $\left[\sum_{i=1}^{N} \mathbf{H}\left(\mathbf{w}_i, \boldsymbol{\theta}^{\{0\}}\right)\right]^{-1}$ may become quite inaccurate or near-singular if many $\mathbf{H}\left(\mathbf{w}_i, \boldsymbol{\theta}^{\{0\}}\right)$ are zero.