ECO 519 (First Half). Spring 2008

Homework # 1

1. (a) Let $\widehat{\gamma} = \gamma_0 + \frac{1}{n}\sum_{i=1}^n \psi_i + o_p(n^{-1/2})$, where $\frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_i \xrightarrow{d} \mathcal{N}(0,\Sigma)$, $\gamma \in \mathbb{R}^p$ and Σ is a $p \times p$ matrix. Now consider an M-estimator $\widehat{\theta}$ given by:

$$\widehat{\theta} = \underset{\theta\in\Theta}{\text{argmin}}\ \frac{1}{n}\sum_{i=1}^n q(z_i, \theta, \widehat{\gamma})$$

Assume that $E[z,\theta,\gamma]$ is uniquely minimized at (θ_0, γ_0). Provide conditions under which $\widehat{\theta}$ is $\sqrt{n}$-consistent, asymptotically normal. Characterize its asymptotic variance (assume q is $\mathcal{C}^2$).

   (b) Consider the Classical Minimum Distance (CMD) estimator (see Newey and McFadden's chapter in the Handbook of Econometrics, 1994) $\widehat{\theta}$ given by:

$$\widehat{\theta} = \underset{\theta\in\Theta}{\text{argmin}}\ \big(\widehat{\pi} - h(\theta)\big)'\widehat{W}\big(\widehat{\pi} - h(\theta)\big)$$

where $\widehat{\pi} = \pi_0 + \frac{1}{n}\sum_{i=1}^n \psi_i + o_p(n^{-1/2})$, and $\frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_i \xrightarrow{d} \mathcal{N}(0,\Sigma)$. Let $W = \text{plim}\ \widehat{W}$ and suppose that $\pi_0 = h(\theta_0)$ (by definition of CMD). Provide conditions under which $\widehat{\theta}$ is $\sqrt{n}$-consistent, asymptotically normal. Characterize its asymptotic variance (assume h is $\mathcal{C}^2$).

2. Suppose $y_i = x_i'\beta_0 + \varepsilon_i$ where ε_i is independent of x_i and ε_i is continuously distributed with strictly increasing CDF $F_\varepsilon(\epsilon)$ and corresponding density function $f_\varepsilon(\epsilon)$. Suppose also that the τ^{th} quantile of ε_i is zero, i.e: $F_\varepsilon(0) = \tau$ and $f_\varepsilon(0) > 0$.

   (a) For $\tau \in (0,1)$ define $\rho_\tau(z) = z(\tau - \mathbb{1}\{z < 0\})$. Show that if the previous assumptions are satisfied, then

$$\underset{\beta}{\text{argmin}}\ E\big[\rho_\tau(y_i - x_i'\beta)\big|x_i\big] = \beta_0$$

   (b) Suppose we observe an iid sample $(y_i, x_i)_{i=1}^n$ (no truncation or censoring). Based on the previous part, propose an estimator for β and sketch any additional conditions needed for consistency and asymptotic normality. Characterize precisely the asymptotic distribution of the proposed estimator if all your assumptions are satisfied. Show that the Least Absolute Deviation (LAD) estimator is a special case. Hint: for asymptotic normality, focus on Huber (1967).

3. Consider a sequence of random vectors $X_n \in \mathbb{R}^k$. We say that $X_n = O_p(n^r)$ if $\forall\ \epsilon > 0$, there exists M_ϵ and n_ϵ such that $\Pr\{\|n^{-r}X_n\| > M_\epsilon\} < \epsilon$ for all $n \geq n_\epsilon$. We say that $X_n = o_p(n^r)$ if $n^{-r}X_n \xrightarrow{p} 0$. With this in mind, let X_n and Y_n be two scalar random variables. Show the following:

   (a) If $X_n = O_p(n^{r_1})$ and $Y_n = O_p(n^{r_2})$, then $X_nY_n = O_p(n^{r_1+r_2})$. Let $\overline{r} = \max\{r_1, r_2\}$, then $X_n + Y_n = O_p(n^{\overline{r}})$.

   (b) If $X_n = o_p(n^{r_1})$ and $Y_n = o_p(n^{r_2})$, then $X_nY_n = o_p(n^{r_1+r_2})$. Let $\overline{r} = \max\{r_1, r_2\}$, then $X_n + Y_n = o_p(n^{\overline{r}})$.

   (c) If $X_n = O_p(n^{r_1})$ and $Y_n = o_p(n^{r_2})$, then $X_nY_n = o_p(n^{r_1+r_2})$ Let $\overline{r} = \max\{r_1, r_2\}$, then $X_n + Y_n = O_p(n^{\overline{r}})$.

   Hint: The following probability inequalities (which come from the Axioms of Probability): $\Pr(A \cup B) \leq \Pr(A) + \Pr(B)$ and $\Pr(A \cap B) \leq \Pr(A) + \Pr(B)$ might be useful.

4. Suppose θ_n is a minimizer of $G_n(\theta)$, and θ_0 is a minimizer of $G(\theta)$ (you may think of $G(\theta)$ as the probability limit of $G_n(\theta)$, although that is not crucial here). Suppose $\theta_n \xrightarrow{p} \theta_0$ and also that

   (a) There exists a neighborhood $\mathcal{N}$ of θ_0 and a constant $\kappa > 0$ for which

   $$G(\theta) \geq G(\theta_0) + \kappa\|\theta - \theta_0\|^2 \text{ for all } \theta \in \mathcal{N}.$$

   (b) Uniformly over $o_p(1)$ neighborhoods of θ_0 (that is, uniformly over any neighborhood of the form $\{\theta : \|\theta - \theta_0\| = o_p(1)\}$),

   $$G_n(\theta) - G_n(\theta_0) = G(\theta) - G(\theta_0) + O_p\left(\frac{\|\theta - \theta_0\|}{\sqrt{n}}\right) + o_p\big(\|\theta - \theta_0\|^2\big) + O_p(b_n),$$

   where $\{b_n\}$ is a sequence of nonnegative real numbers that satisfies $b_n \to 0$ as $n \to \infty$.

Show that if these conditions are satisfied, we must have

$$\|\theta_n - \theta_0\| = O_p\big(\max\{b_n^{1/2}, n^{-1/2}\}\big).$$

Given this, provide sufficient conditions that guarantee $\|\theta_n - \theta_0\| = O_p(n^{-1/2})$ (in this case, we say that θ_n is a "root-n consistent" estimator of θ_0).

5. (This problem can be seen as a follow-up of the previous problem, which provided sufficient conditions for root-n consistency. Now we provide sufficient conditions for root-n asymptotic normality.) Let Θ be a subset of $\mathbb{R}^k$. Suppose θ_n minimizes $G_n(\theta)$ over Θ and suppose θ_n is root-n consistent for θ_0, where θ_0 is in the interior of Θ. Suppose also that uniformly over $O_p(n^{-1/2})$ neighborhoods of θ_0 (that is, uniformly over any set of the form $\big\{\theta : \|\theta - \theta_0\| = O_p(n^{-1/2})\big\}$),

$$G_n(\theta) - G_n(\theta_0) = \frac{1}{2}(\theta - \theta_0)' V(\theta - \theta_0) + \frac{1}{\sqrt{n}}(\theta - \theta_0)' W_n + o_p(1/n)$$

where V is a positive definite matrix and W_n converges in distribution to a $\mathcal{N}(0, \Delta)$ random vector. Show that in this case, $\sqrt{n}(\theta_n - \theta_0) \Rightarrow \mathcal{N}(0, V^{-1}\Delta V^{-1})$. Explain why problems 1 and 2 are a generalization of the condition $\frac{1}{\sqrt{n}}\sum_{i=1}^n \psi(x_i, \theta_0) + \sqrt{n}\lambda(\theta_n) = o_p(1)$, which is the starting point to prove asymptotic normality in Huber's framework.

6. Let $X_1, \ldots, X_n$ be an iid sample with probability measure $P \in \mathcal{P}$ (the probability measure belongs to a family of probability measures denoted by $\mathcal{P}$). Throughout, we will let P_n denote the empirical distribution of our sample. Take a real-valued function $\psi : \mathbb{R} \to \mathbb{R}$ that satisfies the following conditions:

   (i) ψ is monotone nondecreasing.

   (ii) $\psi(-\infty) < 0 < \psi(\infty)$

   (iii) $\big|\psi(x)\big| \leq M < \infty$ for all x

   We will denote the parameter of interest θ by $\theta(P)$ just to remind you that it depends on the underlying true distribution.

(a) Show that any $\theta(P)$ that satisfies

$$E_P\big[\psi(X_i - \theta(P))\big] \geq 0 \geq E_P\big[\psi(X_i - \theta')\big] \text{ for all } \theta' > \theta(P)$$

must be finite.

(b) Suppose that for all $P \in \mathcal{P}$, $\theta(P)$ is the unique solution to $E_P\big[\psi(X_i - \theta)\big] = 0$. Let $\widehat{\theta}_n = \theta(P_n)$ (i.e, the solution to $0 = E_{P_n}\big[\psi(X_i - \theta)\big] = \frac{1}{n}\sum_{i=1}^n \psi(X_i - \theta)$). Show that for any $P \in \mathcal{P}$, $\widehat{\theta}_n \xrightarrow{p} \theta(P)$ (Hint: Show first that $E_P\big[\psi(X_i - \theta)\big]$ is nonincreasing in θ, then use a dominated convergence argument and the fact that $\psi(X, \theta) \to \psi(-\infty) < 0$ as $\theta \to \infty$ and $\psi(X, \theta) \to \psi(\infty) > 0$ as $\theta \to -\infty$. To see how to use these results to establish consistency formally, see for example Lemma 2 and Theorem 2 in Huber, pp. 225-226.)

(c) Assume the conditions of the previous part hold. Let $\lambda(\theta) = E_P\psi(X_i - \theta)$ and $\tau^2(\theta) = \text{Var}_P\psi(X_i - \theta)$. Suppose $\lambda'(\theta) < 0$ exists and suppose that for any sequence $\theta_n \to \theta$ that satisfies $\theta_n = \theta + t/\sqrt{N}$ for some t the following is true:

$$\frac{1}{\sqrt{n}\tau(\theta)}\sum_{i=1}^n\big[\psi(X_i - \theta_n) - \lambda(\theta_n)\big] \Rightarrow \mathcal{N}(0, 1).$$

Given this, let $\widehat{\theta}_n$ be the estimator defined in the previous part. Show that

$$\sqrt{n}(\widehat{\theta}_n - \theta) \Rightarrow \mathcal{N}\left(0, \frac{\tau^2(\theta)}{[\lambda'(\theta)]^2}\right).$$

Hint: The key is to note that for any t and any θ_n such that $\theta_n = \theta + t/\sqrt{N}$, we have: $\Pr\big(\sqrt{n}(\widehat{\theta}_n - \theta) < t\big) = \Pr(\widehat{\theta}_n < \theta_n) = \Pr\big(\sum_{i=1}^n \psi(X_i - \theta_n) > 0\big)$.

(d) Suppose the cdf of X, $F(x)$ is continuous. Let θ denote the median of X and suppose that $F'(\theta) = f(\theta) > 0$. Use the previous two parts to find the asymptotic distribution of the sample median, $\widehat{\theta}_n$. Compare your result with Example 24 in Pollard (1984). (Hint: Use $\psi(x) = \text{sign}(x)$).

7. Consider the censored regression model

$$y_i = \max\big\{0, x_i'\beta_0 + u_i\big\}, \quad i = 1, \ldots, n.$$

In this context, Powell's Censored LAD estimator was defined to be the value β that minimized $S_n(\beta) = \frac{1}{n}\sum_{i=1}^n\big|y_i - \max\{0, x_i'\beta\}\big|$.

(a) Fix $\tau \in (0,1)$ and let $F_Y^{-1}(\tau|x_i, \beta_0)$ denote the τth quantile of y_i. Suppose that the cumulative distribution function $F_u(\cdot)$ of the error terms $\{u_i\}$ is continuously differentiable with positive density at $F_u^{-1}(\tau)$, the τth quantile of u_i. Show that

$$F_Y^{-1}(\tau|x_i, \beta_0) = \max\big\{0, x_i'\beta_0 + F_u^{-1}(\tau)\big\}$$

(b) Let $\rho_\tau(s) = \big[\tau - \mathbb{1}\{s < 0\}\big] \cdot s$. Show that the τth quantile of a random variable Z minimizes $E\big[\rho_\tau(Z-b) - \rho_\tau(Z)\big]$ over b.

(c) Without loss of generality, suppose x_i includes an intercept. For $\tau \in (0,1)$ define

$$\beta_0(\tau) = \beta_0 + F_u^{-1}(\tau) \cdot \boldsymbol{\iota}, \quad \text{where} \quad \boldsymbol{\iota} = (1, 0, \ldots, 0) \text{ —same dimension as } x_i\text{.—}$$

Then $F_Y^{-1}(\tau|x_i, \beta_0) = \max\big\{0, x_i'\beta_0(\tau)\big\}$. Using the analogy principle and the previous parts of this problem, we suggest to estimate $\beta_0(\tau)$ (denote the corresponding estimator as $\widehat{\beta}(\tau)$) as the value of β that minimizes

$$Q_n(\beta, \tau) = \frac{1}{n}\sum_{i=1}^{n} \rho_\tau\big(y_i - \max\{0, x_i'\beta\}\big).$$

Show how to generalize the assumptions of the Censored LAD estimator so that $\widehat{\beta}(\tau)$ is a $\sqrt{n}$-consistent estimator of $\beta_0(\tau)$, and find its asymptotic distribution. Assume for simplicity that you have an iid sample.

(d) Using your last answer, what would be the asymptotic joint distribution of $\big(\widehat{\beta}(\tau_1), \widehat{\beta}(\tau_2)\big)$ for $\tau_1 \in (0,1)$, $\tau_2 \in (0,1)$ and $\tau_1 \neq \tau_2$?

8. Consider the linear model $y_i = \beta_0' x_i + u_i$ and make the following assumptions:

(A1) x_i and u_i are independent. x_i contains a constant (i.e, an intercept) along with a K-dimensional vector of jointly continuously distributed random variables with compact support. In addition, $\nexists\ \delta \neq 0$ such that with probability one: $x_i'\delta = 0$. (i.e, it is not the case that the elements in x_i are linearly dependent w.p.1). We will also assume that the matrix $E[x_i x_i']$ is invertible.

(A2) u_i is a continuously distributed random variable with unbounded support and is symmetrically distributed around zero. Denote its cdf and density functions by

$F_u(\cdot)$ and $f_u(\cdot)$ respectively. The unbounded support assumption implies that $f_u(u) > 0$ for all $u \in \mathbb{R}$. Symmetry implies that the median of u_i is zero.

Take a constant $k \in \mathbb{R}_+$ and define the following function:

$$\rho_k(r) = \begin{cases} r^2 & \text{if } |r| < k \\ k \cdot (2 \cdot |r| - k) & \text{if } |r| \geq k \end{cases}$$

Now fix $k \in \mathbb{R}_+$ and let $\widehat{\beta}$ be the M-estimator that solves:

$$\operatorname*{Min}_{\beta} \ \frac{1}{N} \sum_{i=1}^{N} \rho_k\big(y_i - x_i'\beta\big).$$

(a) Find the probability limit (plim) of $\widehat{\beta}$, call it β^*. What is the relationship between β^* and β_0?

(b) Characterize the asymptotic distribution of $\sqrt{N}(\widehat{\beta} - \beta^*)$

9. Consider the linear model $y_i = \beta_0' x_i + u_i$ and make the following assumptions:

(A1) x_i and u_i are independent. x_i contains a constant (i.e, an intercept) along with a K-dimensional vector of jointly continuously distributed random variables with compact support. In addition, $\nexists\ \delta \neq 0$ such that with probability one: $x_i'\delta = 0$. (i.e, it is not the case that the elements in x_i are linearly dependent w.p.1). We will also assume that the matrix $E[x_i x_i']$ is invertible.

(A2) u_i is a continuously distributed random variable with unbounded support and is symmetrically distributed around zero. Denote its cdf and density functions by $F_u(\cdot)$ and $f_u(\cdot)$ respectively. The unbounded support assumption implies that $f_u(u) > 0$ for all $u \in \mathbb{R}$. Symmetry implies that the median of u_i is zero. We will also assume that $2f_u(u) > f_u(0)$ for all $u \in (0, \tau_{0.90})$, where $\tau_{0.90}$ is the 0.90-percentile of u_i.

Suppose we observe an iid sample $(y_i, x_i)_{i=1}^N$ from the population described above. Let $\theta = (\theta_1', \theta_2')'$ and consider the following M-estimation problem,

$$\operatorname*{Min}_{\theta \in \Theta} \frac{1}{N} \sum_{i=1}^{N} \operatorname{Min}\Big\{ \big|y_i - \theta_1' x_i\big|, \big|y_i - \theta_2' x_i\big| \Big\}.$$

Denote the resulting estimators by $\widehat{\theta} = (\widehat{\theta}_1', \widehat{\theta}_2')'$. The space Θ satisfies the following restrictions:

(A3) For all $(\theta_1, \theta_2) \in \Theta$, $\theta_1' x_i \geq \theta_2' x_i$ with probability one. This is necessary for identification purposes.

(A4) Θ is compact, but big enough to include (at least) the entire interval $[\tau_{0.01}, \tau_{0.99}]$ in the co-ordinates of the constant terms in θ_1 and θ_2, where $\tau_{0.01}$ and $\tau_{0.99}$ denote the 0.01 and 0.99-percentiles of u_i respectively.

Note that the minimization problem *must* incorporate the condition (A3), which will impact the dimensionality of Θ and therefore, of θ and $\widehat{\theta}$. Answer the following:

(a) Find the probability limit of $\widehat{\theta}$, denote it by θ^*. What is the relationship between θ^* and β_0?

(b) Characterize the asymptotic distribution of $\sqrt{N}(\widehat{\theta} - \theta^*)$.

10. Assume the exact same setup of the Censored LAD estimator. Consider a real-valued function $g(\cdot)$ that satisfies: (i) $g(z)$ is strictly increasing for all $z \geq 0$, (ii) $g(0) = 0$, $g'(0) = 0$ and $g''(0) = 0$.

   (a) Show that under the censored-LAD assumptions,

   $$\text{median}\big(g(y_i)\big|x_i\big) = g\big(\max\{0, x_i'\beta_0\}\big)$$

   (b) Consider an estimator $\widetilde{\beta}$ such that

   $$\widetilde{\beta} = \underset{\beta}{\text{argmin}}\ \frac{1}{n}\sum_{i=1}^{n}\Big|g(y_i) - g\big(\max\{0, x_i'\beta\}\big)\Big|$$

   characterize the asymptotic properties of $\widetilde{\theta}$, making any necessary modification to Powell's censored LAD assumptions. *Which of Powell's assumptions can be relaxed now?* (Hint: It's a nontrivial one).

11. (a) Consider some abstract class of estimators indexed by $\tau \in \mathbb{T}$ such that each member of this class has an asymptotically linear representation of the form

   $$\widehat{\theta}(\tau) = \theta_0 + \frac{1}{n}\sum_{i=1}^{n}\psi_i(\tau) + o_p(n^{-1/2})$$

where θ_0 denotes the "true" parameter value (whatever this means in this abstract setting) and the influence function satisfies

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_i(\tau) \xrightarrow{d} \mathcal{N}(0, V(\tau)) \quad \text{for each } \tau \in \mathbb{T}$$

show that if there exists a $\overline{\tau} \in \mathbb{T}$ such that $E\big[(\psi_i(\tau) - \psi_i(\overline{\tau}))\psi_i(\overline{\tau})'\big] = 0$ for all $\tau \in \mathbb{T}$, then $\widehat{\theta}(\overline{\tau})$ is asymptotically efficient in the class $\{\widehat{\theta}(\tau) : \tau \in \mathbb{T}\}$.

(b) Suppose there exists a function ϕ such that $E\big[\phi(z, \theta_0)\big|x\big] = 0$. Assume that $\theta \in \mathbb{R}^k$ and $\phi \in \mathbb{R}^r$. Consider a family of estimators $\widehat{\theta}(\tau)$ all of which are asymptotically linear with influence functions of the form

$$\underbrace{\psi_i(\tau)}_{k\times 1} = \underbrace{E\big[\tau(x)\nabla_\theta\phi(z, \theta_0)\big]}_{k\times k}\,\tau(x_i)\phi(z_i, \theta_0)$$

where $\tau(x)$ is a $k \times r$ matrix. Apply Theorem 5.3 in Newey and McFadden to find the τ that yields the smallest variance within this class of estimators. What is its asymptotic variance?