This exam has two questions and a total 100 possible points. Please answer all the questions in a clear and concise manner.

Use the blue book for your answers. Indicate any extra assumptions you may require and make explicit all the steps in your derivations.

GOOD LUCK!
Spring Prelim Exam

Proposition 1 (Consistency with compact parameter space) Suppose that (i) $\Theta$ is a compact subset of $\mathbb{R}^p$, (ii) $Q_n(\theta)$ is continuous in $\theta$ for any data $(w_1, \ldots, w_n)$, and (iii) $Q_n(\theta)$ is a measurable function of the data for all $\theta$ in $\Theta$. If there is a function $Q_0(\theta)$ such that

a. (identification) $Q_0(\theta)$ is uniquely maximized in $\Theta$ at $\theta_0 \in \Theta$,

b. (uniform convergence) $Q_n(\cdot)$ converges uniformly in probability to $Q_0(\cdot)$,

then $\hat{\theta} \to_p \theta_0$.

Proposition 2 (Consistency without compactness) Suppose that (i) the true parameter vector $\theta_0$ is an element of the interior of a convex parameter space $\Theta \subset \mathbb{R}^p$, (ii) $Q_n(\theta)$ is concave over the parameter space for any data $(w_1, \ldots, w_n)$, and (iii) $Q_n(\theta)$ is a measurable function of the data for all $\theta$ in $\Theta$. If there is a function $Q_0(\theta)$ such that

a. (identification) $Q_0(\theta)$ is uniquely maximized in $\Theta$ at $\theta_0 \in \Theta$,

b. (pointwise convergence) $Q_n(\theta)$ converges in probability to $Q_0(\theta)$, $\forall \theta \in \Theta$,

then $\hat{\theta} \to_p \theta_0$.

Proposition 3 (Uniform Law of Large Numbers) Let $\{w_t\}$ be an ergodic stationary process. Suppose that (i) the set $\Theta$ is compact, (ii) $m(w_t; \theta)$ is continuous in $\theta$ for all $w_t$, and (iii) $m(w_t, \theta)$ is measurable in $w_t$ for all $\theta$ in $\Theta$. Suppose in addition, the

**dominance condition:** there exists a function $d(w_t)$ such that $|m(w_t, \theta)| \leq d(w_t)$ for all $\theta \in \Theta$ and $\mathbb{E}(d(w_t)) < \infty$.

then $\frac{1}{n} \sum_{i=1}^{n} m(w_t, \cdot)$ converges uniformly in probability to $\mathbb{E}(m(w_t, \cdot))$ over $\Theta$. Moreover, $\mathbb{E}(m(w_t, \cdot))$ is a continuous function of $\theta$. 

2
Proposition 4 (Asymptotic normality of M-estimators) Suppose that the conditions for consistency are satisfied. Assume also that

1. $\theta_0$ is in the interior of $\Theta$
2. $m(\mathbf{w}_t, \theta)$ is twice continuously differentiable in $\theta$ for any $(\mathbf{w}_t, \text{and})$
3. $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} s(\mathbf{w}_t, \theta_0) \rightarrow_p \mathcal{N}(0, \Sigma)$, $\Sigma$ is positive definite, where $s(\mathbf{w}_t, \theta)$ is the score associated with $m(\mathbf{w}_t, \theta)$.
4. (local dominance condition on the Hessian) for some neighborhood $\mathcal{N}$ of $\theta_0$,
   $$\mathbb{E}[\sup_{\theta \in \mathcal{N}} ||H(\mathbf{w}_t, \theta)||] < \infty$$
   so that for any consistent estimator $\hat{\theta}$, $\frac{1}{n} \sum_{i=1}^{n} H(\mathbf{w}_t, \hat{\theta}) \rightarrow_p \mathbb{E}[H(\mathbf{w}_t, \theta_0)]$
5. $\mathbb{E}[H(\mathbf{w}_t, \theta_0)]$ is nonsingular.

then $\hat{\theta}$ is asymptotically normal with

$$\text{Avar}(\hat{\theta}) = (\mathbb{E}[H(\mathbf{w}_t, \theta_0)])^{-1}\Sigma(\mathbb{E}[H(\mathbf{w}_t, \theta_0)])^{-1}.$$
Proposition 5 (Asymptotic normality of GMM) Suppose that the conditions for consistency are satisfied. Assume also that

1. $\theta_0$ is in the interior of $\Theta$

2. $g(w_t, \theta)$ is twice continuously differentiable in $\theta$ for any $w_t$, and

3. $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(w_t, \theta_0) \rightarrow_p N(0, S)$, $S$ is positive definite.

4. (local dominance condition on $\frac{\partial g(w_t, \theta)}{\partial \theta'}$) for some neighborhood $N$ of $\theta_0$,

$$E[\sup_{\theta \in N} ||\frac{\partial g(w_t, \theta)}{\partial \theta'}||] < \infty$$

so that for any consistent estimator $\hat{\theta}$, $\frac{1}{n} \sum_{i=1}^{n} \frac{\partial g(w_t, \hat{\theta})}{\partial \theta'} \rightarrow_p E[\frac{\partial g(w_t, \theta_0)}{\partial \theta'}]$

5. $E[\frac{\partial g(w_t, \theta_0)}{\partial \theta'}]$ is of full column rank

then $\hat{\theta}$ is asymptotically normal with

$$\text{Avar}(\hat{\theta}) = (G'WG)^{-1}G'WSWG(G'WG)^{-1}$$

where $G = E[\frac{\partial g(w_t, \theta_0)}{\partial \theta'}]$
1. (40 points) Let \((Y_1, \ldots, Y_n)\) be a random sample of \(n\) observations (i.e. \(Y_i\)'s are independent and identically distributed) from a uniform distribution with support \([0, \theta]\).

(a) (8 points) Find the pdf of \(\hat{\sigma}^2 \equiv \frac{(Y_1 - Y_2)^2}{2}\).

(b) (8 points) Show that \(\hat{\sigma}^2\) is an unbiased estimator for the variance of this uniform distribution.

(c) (8 points) Consider the following estimator for \(\theta\):

\[
\hat{\theta} = \sqrt{\frac{12}{n} \sum_{i=1}^{n} (Y_i - \bar{Y})^2}
\]

Show that \(\hat{\theta}\) is a consistent estimator for \(\theta\).

(d) (8 points) Obtain the asymptotic distribution of \(\hat{\theta}\).

(e) (8 points) Consider the representation

\[
Y_i = \theta \frac{1}{2} + \epsilon
\]

and let \(\tilde{\theta}\) denote the ordinary least squares estimator of \(\theta\). What is the formula for \(\tilde{\theta}\)?

Discuss: “By the Gauss-Markov Theorem, \(\tilde{\theta}\) is the best linear unbiased estimator for \(\theta\).”
2. (60 points) An economist is interested in the relation between labor supply and fertility. To study the subject she collects data on female job market status and fertility (let’s ignore other variables for simplicity). The analysis relies on the following model:

\[ Y_i = 1(\alpha + \beta X_i - \epsilon_i \geq 0) \]

where \( Y_i \) is a binary variable (i.e. \( Y_i \in \{0, 1\} \)) that marks whether individual \( i \) works in that period, \( X_i \) is also binary (i.e. \( X_i \in \{0, 1\} \)) and records whether a child is born to individual \( i \) in that period and \( \epsilon_i \) represents other factors related to labor supply. Observations are iid.

Denote by \( F_\epsilon \) the cdf for \( \epsilon_i \) and assume it is known and continuous with full support on the real line. Let \( p_1 \) be the probability that women who had a child participate in the labor force (i.e. \( p_1 = \mathbb{P}(Y = 1|X = 1) = F_\epsilon(\alpha + \beta) \)) and, analogously, \( p_0 \) denotes the probability that those who did not have a child participate in the labor force (i.e. \( p_0 = \mathbb{P}(Y = 1|X = 0) = F_\epsilon(\alpha) \)). Finally, let \( \pi = \mathbb{P}(X_i = 1) \).

(a) (6 points) Derive the expression for the slope and intercept coefficients of the Best Linear Projection (BLP) of \( Y_i \) onto \( X_i \).

(b) (10 points) The values of \( p_1 \) and \( p_0 \) can be easily estimated as:

\[ \hat{p}_1 = \frac{\sum_{i=1}^{n} Y_i X_i}{\sum_{i=1}^{n} X_i} \]

and

\[ \hat{p}_0 = \frac{\sum_{i=1}^{n} Y_i (1 - X_i)}{\sum_{i=1}^{n} (1 - X_i)} \].

These can then be used to estimate \( \alpha \) and \( \beta \) according to the following equations:

\[ \hat{\alpha} = F_\epsilon^{-1}(\hat{p}_0) \quad \text{and} \quad \hat{\beta} = F_\epsilon^{-1}(\hat{p}_1) - F_\epsilon^{-1}(\hat{p}_0). \]

Show that these estimators are consistent.

(c) (12 points) Obtain the asymptotic distribution for the estimators in (b).

*Hint: One way to obtain the asymptotic distribution is to stack the four estimating equations.*

(d) (5 points) Show that the likelihood function for the problem above is given by the
expression below and show that the MLE corresponds to the estimator in (b):

\[
\Pi_{i=1}^{n} \left\{ \left( F_\epsilon(a + b) \pi \right)^{Y_i X_i} \left( (1 - F_\epsilon(a + b)) \pi \right)^{X_i(1 - Y_i)} \times \\
\times \left( F_\epsilon(a)(1 - \pi) \right)^{(1 - X_i) Y_i} \left( (1 - F_\epsilon(a)) (1 - \pi) \right)^{(1 - X_i)(1 - Y_i)} \right\}
\]

(e) (5 points) Suggest a test for the null hypothesis that \((\alpha, \beta) = (0, 1)\) against the alternative hypothesis that \((\alpha, \beta) = (1, 0)\). Justify your answer.

(f) (12 points) Because

\[
\mathbb{E}(Y_i | X_i) = F_\epsilon(a + \beta X_i),
\]

the following estimator for \((\alpha, \beta)\) is suggested:

\[
(\bar{\alpha}, \bar{\beta}) \equiv \arg\min_{(a,b)} \sum_{i=1}^{n} (Y_i - F_\epsilon(a + b X_i))^4
\]

Is this estimator consistent? Is it more or less efficient than the estimator suggested in (b)?

(g) (10 points) Assume that instead of \(X_i\) you observe a noisy measure of whether a woman had a birth, \(\tilde{X}_i\). Whereas the record is accurate when there is no birth (i.e. \(\mathbb{P}(\tilde{X}_i = 0 | X_i = 0) = 1\)) there is a positive probability, say \(\psi\), that a birth is not recorded (i.e. \(\mathbb{P}(\tilde{X}_i = 0 | X_i = 1) = \psi\)). Consider the estimator in (b) with \(\tilde{X}_i\) instead of \(X_i\). Obtain the probability limit for the estimator. If you were given \(\psi\), how would you modify the estimator to reestablish consistency?
I used to think correlation implied causation.

Then I took a statistics class. Now I don't.

Sounds like the class helped. Well, maybe.