

ECO 519 (Second Half). Spring 2008

Homework # 2

1. 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,

$$\text{Min}_{\theta \in \Theta} \frac{1}{N} \sum_{i=1}^N \text{Min} \left\{ |y_i - \theta_1' x_i|, |y_i - \theta_2' x_i| \right\}.$$

Denote the resulting estimators by $\hat{\theta} = (\hat{\theta}_1', \hat{\theta}_2')'$. 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 $\hat{\theta}$. Answer the following¹:

(a) Show that there exists a function $\psi(\cdot)$ of the same dimension as $\hat{\theta}$ such that

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \psi(y_i, x_i, \hat{\theta}) = o_p(1).$$

Using our assumptions, deduce from this condition the probability limit of $\hat{\theta}$, denote it by θ^* . What is the relationship between θ^* and β_0 ?

2. Suppose we have an iid sample $(X_i, Y_i)_{i=1}^N$, with $X_i \in \mathbb{R}^m$ and $Y_i \in \mathbb{R}$. We are interested in a specification test for a particular functional form for $E[Y_i|X_i]$. Specifically, we want to test if $E[Y_i|X_i] = f(X_i, \theta_0)$, where $f(\cdot, \theta_0)$ is a known function, and $\theta_0 \in \mathbb{R}^k$ is an unknown (but estimable) finite-dimensional parameter. The null hypothesis is

$$H_0 : \Pr[E[Y_i|X_i] = f(X_i, \theta_0)] = 1$$

Suppose $\theta_0 \in \Theta \subset \mathbb{R}^k$. If H_0 is satisfied, then $E[(Y_i - f(X_i, \theta))^2]$ is uniquely minimized at θ_0 . The alternative hypothesis is that the model is not correct, and

$$H_1 : \Pr[E[Y_i|X_i] = f(X_i, \theta)] < 1 \quad \text{for all } \theta \in \Theta.$$

Suppose we have available an estimator $\hat{\theta}$ with the property that, if the model is correct and therefore H_0 is true, $\hat{\theta}$ is consistent and asymptotically linear:

$$\hat{\theta} = \theta_0 + \frac{1}{N} \sum_{i=1}^N \psi_i + o_p(N^{-1/2}), \quad \text{where } E[\psi] = 0, \text{ and } E[\psi\psi'] \text{ is finite, nonsingular.}$$

If the model is not true, then $\hat{\theta} \xrightarrow{p} \theta^* \in \Theta$. Suppose the following conditions hold: Θ is compact, θ_0 belongs in the interior of Θ , $f(x, \theta)$ is twice continuously differentiable

¹At any point, feel free add assumptions you *think* are needed. However, unnecessary additional assumptions, or assumptions that invalidate the results you are asked to obtain, will be penalized.

on Θ for each $x \in \mathbb{R}^m$, and

$$\begin{aligned} E \left[\sup_{\theta \in \Theta} |f^2(X_i, \theta)| \right] < \infty, \quad E \left[\sup_{\theta \in \Theta} \left\| (Y_i - f(X_i, \theta))^2 \cdot \frac{\partial f(X_i, \theta)}{\partial \theta} \frac{\partial f(X_i, \theta)'}{\partial \theta} \right\| \right] < \infty, \\ E \left[\sup_{\theta \in \Theta} \left\| \frac{\partial f(X_i, \theta)}{\partial \theta} \frac{\partial f(X_i, \theta)'}{\partial \theta} \right\| \right] < \infty, \quad E \left[\sup_{\theta \in \Theta} \left\| (Y_i - f(X_i, \theta))^2 \cdot \frac{\partial^2 f(X_i, \theta)}{\partial \theta \partial \theta'} \right\| \right] < \infty \\ E \left[\frac{\partial f(X_i, \theta_0)}{\partial \theta} \frac{\partial f(X_i, \theta_0)'}{\partial \theta} \right] \text{ is invertible.} \end{aligned}$$

We assume the vector X_i to be jointly continuously distributed, with density $f_X(x)$, which is M -times continuously with bounded derivatives up to order M (where $M \geq 1$). Define $\varepsilon_i \equiv Y_i - f(X_i, \theta_0)$ and $\widehat{\varepsilon}_i \equiv Y_i - f(X_i, \widehat{\theta})$. Let h_N be a bandwidth sequence that satisfies $h_N \rightarrow 0$ and $Nh_N^m \rightarrow \infty$. Let $K : \mathbb{R} \rightarrow \mathbb{R}$ be a nonnegative, bounded, continuous and symmetric kernel function that satisfies $\int K(u)du = 1$. Define

$$H_N(Z_i, Z_j) = \frac{1}{h_N^m} K\left(\frac{X_i - X_j}{h_N}\right) \widehat{\varepsilon}_i \widehat{\varepsilon}_j$$

and let

$$U_N = \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N H_N(Z_i, Z_j) = \binom{N}{2}^{-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^N N H_N(Z_i, Z_j) \equiv \binom{N}{2}^{-1} \sum_{i < j} H_N(Z_i, Z_j).$$

(a) Show that if H_0 is satisfied, then

$$Nh_N^{m/2} U_N \xrightarrow{d} \mathcal{N}(0, \Sigma)$$

and characterize Σ . Use this result to suggest a pivotal statistic. Add any further assumptions you think are needed. In particular: do you need additional smoothness or the use of a bias-reducing kernel (and additional conditions for the bandwidth sequence) for this result to hold?

(b) Show that under the alternative H_1 , your test-statistic explodes, and therefore it is a consistent test (it detects any deviation from the null hypothesis of the type displayed in H_1).

Hint: Use Hall (1984).

3. Suppose Z is a real-valued random variable with $E[Z] \equiv \mu$ and $\text{Var}[Z] \equiv \sigma^2 < \infty$. Suppose we have an iid sample $(Z_i)_{i=1}^N$ from this population and let

$$U_N = \binom{N}{2}^{-1} \sum_{i < j} X_i X_j - \mu^2$$

Find a value α such that $N^\alpha U_N$ has a nondegenerate limiting distribution. Characterize the asymptotic distribution, and how the asymptotic distribution depend on the value of μ (and σ^2).

4. Let $m(x) = E[Y|X = x]$, where $X \in \mathbb{R}$. Consider a nonparametric estimator

$$\widehat{m}(x) = \frac{1}{Nh\widehat{f}(x)} \sum_{i=1}^N Y_i K\left(\frac{X_i - x}{h}\right), \quad \text{where } x \in \mathbb{S}(X) \text{ (the support of } X)$$

Suppose $K(\cdot)$ is differentiable everywhere in \mathbb{R} . Now consider the object

$$\frac{d\widehat{m}(x)}{dx} \equiv \widehat{m'(x)}$$

- (a) Find conditions under which $\widehat{m'(x)}$ is a consistent estimator for $m'(x) \equiv \frac{dE[Y|X=x]}{dx}$.

Make whatever smoothness assumptions needed.

- (b) Find conditions –if they exist– under which

$$\frac{1}{\sqrt{N}} \sum_{i=1}^n (\widehat{m'(X_i)} - m'(X_i)) \xrightarrow{d} \mathcal{N}(0, V)$$

What type of kernel would yield this result? Provide an example of such a kernel –if it exists–. If this result is valid, find the expression for the asymptotic variance V .

5. Let $\mu(W) = E[Z|W]$. Suppose the following model is true about Y, X, W (each of which is a real-valued random variable) and an unobservable ε :

$$Y = \beta_0 + \beta_1 X + \beta_2 \mu(W) + \varepsilon; \quad \text{where } E[\varepsilon|X] = 0 \text{ and } E[\varepsilon|\mu(W)] = 0.$$

Let

$$\widehat{\mu}(w) = \frac{1}{Nh} \sum_{i=1}^N Z_i K\left(\frac{W_i - w}{h}\right)$$

and suppose that $\mu(\cdot)$ is assumed to be smooth enough, and $K(\cdot)$ is chosen appropriately so that the following result is true for any $w \in \mathbb{S}(W)$:

$$\widehat{\mu}(w) - \mu(w) = \frac{1}{Nh} \sum_{i=1}^N \frac{[Z_i - \mu(w)]}{f_w(w)} K\left(\frac{W_i - w}{h}\right) + \xi_N(w),$$

where $\sup_{w \in \mathbb{S}(W)} (N^{1-\delta}h)|\xi_N(w)| = O_p(1)$ for any $\delta > 0$. With all this in mind, consider the estimator $\widehat{\beta} = (\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2)$ that solves:

$$\widehat{\beta} = \underset{\widehat{\beta}}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^N (Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_i - \widehat{\beta}_2 \mu(W_i))^2.$$

Find conditions under which $\widehat{\beta}$ is \sqrt{N} -consistent, and characterize its asymptotic distribution. Explain why in general, it is more likely that we would have to use a trimmed-sum of squared residuals of the form $\frac{1}{N} \sum_{i=1}^N (Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_i - \widehat{\beta}_2 \mu(W_i))^2 \phi(W_i)$, where $\phi(W_i)$ is different from zero inside some compact set in the interior of $\mathbb{S}(W)$ and zero elsewhere.

6. Define the following empirical process:

$$\nu_N(\tau) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \left[\psi(y_i, x_i, \tau) - E[\psi(y_i, x_i, \tau)] \right]$$

where $\psi(y_i, x_i, \cdot)$ is the function you found in the previous part. Show that there exists a pseudometric $\rho : \Theta \times \Theta \rightarrow \mathbb{R}_+$ such that $\nu_N(\tau)$ is stochastically equicontinuous in the pseudometric space (Θ, ρ) . That is, show that for all $\varepsilon > 0$ and $\eta > 0$, there exists $\delta > 0$ such that

$$\overline{\lim}_{N \rightarrow \infty} \Pr \left[\sup_{\tau_1, \tau_2 \in \Theta: \rho(\tau_1, \tau_2) < \delta} \|\nu_N(\tau_1) - \nu_N(\tau_2)\| > \eta \right] < \varepsilon.$$

You may invoke without proof any valid existing result that is pertinent to the case at hand. What is the pseudometric that supports your result?

7. Use the two previous results along with the assumptions (A1)-(A4) to characterize the asymptotic distribution of $\sqrt{N}(\widehat{\theta} - \theta^*)$.

8. Suppose we drop the assumption that u_i is symmetrically distributed around zero. Describe in qualitative terms how your results would change. Can we use $\hat{\theta}$ (perhaps along with another estimator) to learn something about how symmetric (around zero) the distribution of u_i is?
9. Consider the discrete choice model $y_i = \mathbb{1}\{\beta_{0,1} + \beta_{0,2}x_i - \varepsilon_i \geq 0\}$ and denote $\beta_0 \equiv (\beta_{0,1}, \beta_{0,2})$. For simplicity we assume $x_i \in \mathbb{R}$. Suppose that ε_i might be correlated with x_i , but there exists a real-valued random variable η_i such that for every $\epsilon \in \mathbb{R}$

$$\Pr(\varepsilon_i \leq \epsilon | x_i, \eta_i) = \frac{\exp(\epsilon + \varphi(\eta_i))}{1 + \exp(\epsilon + \varphi(\eta_i))},$$

where $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ is *unknown* but assumed to be one-to-one everywhere in \mathbb{R} . This model says that conditional on η_i , ε_i is independent of x_i . Suppose η_i is unobserved but is defined by the equation $x_i = \gamma_0'z_i + \eta_i$, where z_i is an L -dimensional random vector independent of ε_i and satisfies $E[\eta_i | z_i] = 0$. Assume throughout that x_i , z_i and η_i are all continuously distributed. Let $f_\eta(\eta)$ and $f_{x,\eta}(x, \eta)$ denote the marginal density of η and the joint density of x and η respectively. We will assume both of them to be smooth functions of η (you will determine how “smooth” they need to be below). Suppose we observe an iid sample (y_i, x_i, z_i) from the population described above. Answer the following²:

- (a) Given all these assumptions, find a population moment condition that identifies the parameter vector $\theta_0 = (\beta_0, \gamma_0)$.
- (b) Suppose γ_0 is known. Based on your previous answer, propose an estimation procedure to estimate β . Provide any additional assumptions needed for your proposed estimator to satisfy:
- (i) Consistency: $\hat{\beta} \xrightarrow{p} \beta_0$.
- (ii) Asymptotic normality: $\sqrt{N}(\hat{\beta} - \beta_0) \xrightarrow{d} \mathcal{N}(0, V)$, and characterize its asymptotic variance V .

²At any point, feel free add assumptions you *think* are needed. However, unnecessary additional assumptions, or assumptions that invalidate the results you are asked to obtain, will be penalized.

(c) Suppose now that γ_0 is unknown and to keep the algebra simple suppose z_i is a real-valued continuous random variable. Suppose we estimate γ in a first-stage using OLS. In this case, go as far as you can in repeating your answers to part (9b) assuming that you use the OLS estimate $\hat{\gamma}$ instead of the unknown γ_0 to estimate β_0 .