# Problem Set 1

MIT 14.385, Fall 2007

Due: Friday, 21 September 2007, in class

This problem set emphasizes theory. Next week's problem set will consist of empirical applications of these results.

Prepare very brief, precise answers. Irrelevant, lengthy explanations will be penalized by the means of negative points. State clearly any additional assumptions if needed. You are strongly encouraged to discuss the pset in groups but the final write-up should be individual.

The page numbers give a rough indication how detailed your answer should be. For each problem, you can get either the full number of points for a good answer, half the points for an incomplete or only partly correct answer, or no points at all. If you receive less than the full number of points on a given problem, you may hand in a revised answer for that problem one week after you got back the problem set, and get up to 90 percent of the original score for a correct answer.

See the last page for some hints.

1. Basic Consistency Exercise #1 [5 points, 1/2 page]:

   Consider the model

$$y_t = \beta_0 + u_t$$

   where $\beta_0$ is an unknown scalar parameter and the $u_t$ are i.i.d. disturbances with

$$\begin{aligned} E[u_t] &= 0 \\ E\left[u_t^2\right] &= \beta_0^2 \end{aligned}$$

   *True, false or uncertain:* computing $\hat{\beta}$ using the minimum distance function

$$\sum_{t=1}^{T} \left( \frac{y_t - \beta}{\beta} \right)^2$$

   yields a consistent estimator.

2. Basic Consistency Exercise #2 [5 points, 1/2 page]:

   Suppose

   $$\log y_t = \beta_0 + x_t\beta_1 + u_t$$

   where $u_t \sim N\left(0, \sigma^2\right)$ i.i.d. with $u_t$ independent of $x_t$.

   *True, false or uncertain:* if one implements NLLS (nonlinear least squares) to estimate the model

   $$y_t = \exp\left(\gamma_0 + x_t\beta_1\right) + a_t$$

   the resulting estimate of $\beta_1$ is consistent.

3. Nonlinear Regression with a Transformation of the Dependent Variable [15 points, 1.5 pages]:

   Consider the empirical relation

   $$\left(y_t + \gamma\right)^3 = \delta + x_t\beta + \varepsilon_t$$

   where $\varepsilon_t \sim \left(0, \sigma^2\right)$ i.i.d., $x_t$ is a scalar exogenous variable and $\gamma, \delta, \beta$ are parameters. *Evaluate the following statements as true, false or uncertain:*

   (a) NLLS applied to this specification yields consistent estimates for $\delta$ and $\beta$ but not for $\gamma$.

   (b) Nonlinear IV estimation using $1, x_t, x_t^2$ as instruments produces estimates for $\gamma, \delta$ and $\beta$ that are consistent.

4. Least Absolute Deviation [20 points, 1 page]

   Consider the model

   $$y_t = \theta_{1/2} + u_t$$

   with $u_t$ i.i.d. and $P\left(u_t \leq 0\right) = \frac{1}{2}$. $u_t$ has a strictly positive density. Assume $E\left[|y_t - \theta|\right] < \infty$.

   (a) *True, false or uncertain:* $\theta_{1/2}$ is the median of $y_t$

(b) *True, false or uncertain:* The LAD (least absolute deviations) estimator

$$\begin{aligned}\hat{\theta} &\equiv \arg\min_{\theta} \frac{1}{T}\sum_{t=1}^{T}|y_t - \theta| \\ &\equiv \arg\min_{\theta} \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{2} - \mathbf{1}\{y_t < \theta\}\right)(y_t - \theta)\end{aligned}$$

is a consistent estimator of $\theta_{1/2}$.

Hint: use convexity and the fact that

$$\begin{aligned}\frac{\partial}{\partial\theta}E[|y_t - \theta|] &= E\left[\frac{\partial}{\partial\theta}|y_t - \theta|\right] \\ &= E\left[\mathbf{1}\{y_t < \theta\} - \frac{1}{2}\right]\end{aligned}$$

to check if $\theta_{1/2}$ is the unique optimum of the limit objective function. (Note: $\frac{\partial}{\partial\theta}|y_t - \theta|$ is not defined when $y_t = \theta$, but for a given $\theta$, $P(y_t = \theta) = 0$, so this is unimportant.)

(c) *True, false or uncertain:* OLS is a consistent estimator of $\theta_{1/2}$.

(d) *True, false or uncertain:* (Extra points) The condition $E[|y_t - \theta|] < \infty$ is needed for consistency in (b).

(e) Now consider a new the error term $\tilde{u}_t$ , with $P(\tilde{u}_t \leq 0) = \tau$ and the model is

$$y_t = \theta_\tau + \tilde{u}_t$$

*True, false or uncertain:* $\theta_\tau$ is the $\tau$-quantile of $y_t$.

Now let the $\tau$-quantile estimator be

$$\hat{\theta}_\tau \equiv \arg\min_{\theta} \frac{1}{T}\sum_{t=1}^{T}(\tau - \mathbf{1}\{y_t < \theta\})(y_t - \theta)$$

This is sometimes written as

$$\hat{\theta}_\tau \equiv \arg\min_\theta \frac{1}{T} \sum_{t=1}^{T} \left\{ \tau \left(y_t - \theta\right)^+ + (1-\tau)\left(y_t - \theta\right)^- \right\}$$

where $x^+ = \max(x, 0)$ and $x^- = -\min(x, 0)$. Be sure you understand why these are equivalent. Note that the median discussed above is a special case.

(f) *True, false or uncertain:* $\hat{\theta}_\tau$ is consistent for $\theta$.

(g) Why might we ever be interested in $\theta_\tau$ for $\tau \neq \frac{1}{2}$?

5. Censored Sample – Top-Coding [25 points, 2.5 pages]

Consider the model

$$y^* = x'\beta_0 + \varepsilon \; ; \; \varepsilon | x \sim N\left(0, \sigma_0^2\right)$$

where $y^*$ is the log of income and $x$ are the variables that should predict income, e.g. age, education, ability, etc. Suppose that high incomes are *censored* for confidentiality reasons. Rather than observing $(y_i^*, x_i)$, we observe

$$\begin{aligned} (y_i^*, x_i) \text{ if } y_i^* &< L \\ (L, x_i) \text{ if } y_i^* &\geq L \end{aligned}$$

$L$ is known. Another way to write this is that you only observe

$$y_i = \min(y_i^*, L)$$

(a) What is the conditional mean of $y$ given $x$, i.e.

$$E[\, y \,|\, x \,]$$

Note that this is *observed* $y$, not the true $y^*$.

(b) Suppose you run an OLS regression on the data you have. Will $\hat{\beta}_{LS}$ be consistent for $\beta_0$? Give intuition as to why or why not. (Hint: a sketch will help.)

(c) Explain how the expression for the conditional mean you derived in part (a) can be used for Nonlinear Least Squares (NLS) estimation of $\beta_0$. Is the NLS estimate consistent? A very brief answer suffices.

(d) What is the conditional mean of $y$ given $x$ and that $y < L$, i.e.

$$E\left[\, y \,|\, x \,;\, y^* < L \,\right]$$

(e) Suppose you run an OLS regression on the non-censored data, i.e. OLS on $(y_i^*, x_i)$ for all $i$ such that $y_i^* < L$. Will $\hat{\beta}_{LS}$ be consistent for $\beta_0$? Give intuition as to why or why not. (Hint: a sketch will help.)

(f) What is the (conditional) likelihood function of $y$ given the observed $x$? Is the MLE consistent (a brief answer suffices)? Give the large sample distribution of the MLE.

(g) (Truncated Sample) Now consider the model

$$y^* = x'\beta_0 + \varepsilon \;;\; \varepsilon|x \sim N\left(0, \sigma_0^2\right)$$

where $y^*$ is the log of income and $x$ are the variables that should predict income, e.g. age, education, ability, etc. Suppose that high incomes are excluded from the sample, i.e. if $y_i^* < L$, we observe $(y_i, x_i)$ with $y_i = y_i^*$, whereas if $y_i^* > L$ we do not observe anything. $L$ is known.

- What is the conditional mean of $y$ given $x$ and that $y$ is observed, i.e.

$$E\left[y \mid x; y^* < L\right]$$

- Suppose you run an OLS regression on the data you have. Will $\hat{\beta}_{LS}$ be consistent for $\beta_0$?

- What is the (conditional) likelihood function of $y$ given the observed $x$? Is the MLE consistent? A very brief answer suffices.

6. Consistency of QMLE / mis-specified MLE [15 points, 2 pages]

   Consider a linear *exponential density* of the form

   $$f(y|m) = \exp(A(m) + B(y) + C(m)y)$$

   where $\int y f(y|m)\,dy = m$.

   (a) Warmup question: does the Poisson distribution

   $$f(y \mid \lambda) = e^{-\lambda}\lambda^{y}/y!$$

   belong in this family?

   (b) What is $E[\ln f(y|m)]$ when $y \sim f(y|m_0)$? (Note: we are looking at the general linear exponential density now, the Poisson just applied to part (a).)

   (c) Using the *information inequality* show that for any $m_0$, $A(m) + C(m)m_0$ is maximized (as a function of $m$) at $m = m_0$.

   (d) What is $E[\ln f(y|m)]$ in general, i.e. when $y$ is not distributed $f(y|m)$?

   (e) Show that as long as $E[y] = m_0$, $E[\ln f(y|m)]$ is maximized at $m_0$, even if $y \nsim f(y|m)$.

   (f) Think about the result in (e). Think some more. What implications does this have for the consistency of the MLE of $m$ when $y$ does not have an exponential density?

   (g) Extend the previous analysis to show that if the conditional likelihood has the form

   $$f(y|x,\beta,\gamma) = \exp\{A(h(x,\beta),x,\gamma) + B(y,x,\gamma) + C(h(x,\beta),x,\gamma)y\}$$

   then

   $$E[\ln f(y|x,\beta,\gamma)]$$

will be maximized at $\beta = \beta_0$ so long as $E[y|x] = h(x, \beta_0)$

(h) Do the following models belong to the family described in part (f)?

i. Normal:

$$y = x'\beta_0 + \varepsilon \ ; \ \varepsilon \sim N(0, \sigma_0)$$

ii. Logit:

$$P(y = 1 \mid x, \beta_0) = \frac{\exp(x'\beta_0)}{1 + \exp(x'\beta_0)}$$

iii. Poisson regression

$$\begin{aligned} f(y \mid \lambda) &= e^{-\lambda}\lambda^y / y! \\ \lambda &= \exp(x'\beta_0) \end{aligned}$$

(i) Suppose you are interested in firms' applications for patents. You run a Poisson regression model (as in 6.g.iv above). However, the truth (unbeknownst to you) is that patents actually follow a negative binomial model (which permits the variance to differ from the mean), but the mean is correctly specified.

i. Will your estimate be consistent? Asymptotically normal? If yes, give the asymptotic variance of the MLE.

ii. (Optional) Will your estimated standard errors be consistent? (This gets beyond the consistency results we're emphasizing in this pset, but if you've studied the asymptotic distribution of extremum estimators, MLEs and QMLEs you should be able to answer this one. If not, don't worry about it.)

7. Consistency and asymptotic normality of logit MLE [20 points, 2 pages]

In the previous questions, we asked for informal arguments for consistency and normality. In this question, you will need to go through a formal proof of consistency and asymptotic normality. A good model is the proof for the probit case is in the lecture notes or in Newey and McFadden handbook chapter. Unlike in other problems try to be pedantic about details. The purpose of this exercise is for you to study and review the more technical material of the lecture notes. Be as concise as possible.

(a) Consider the usual binary choice logit model where $y \in \{0,1\}$ and $P(y=1 \mid x) = \Lambda(x'\beta_0)$, with $\Lambda(x'\beta) = \exp(x'\beta)/(1+\exp(x'\beta))$. Give conditions on the distribution of $x$ that are sufficient for the consistency and asymptotic normality (CAN) of the logit MLE $\hat{\beta}$ and prove that, under these conditions, $\hat{\beta}_{ML}$ is CAN. You may use without proof the result that $\log \Lambda(t)$ and $\log(1-\Lambda(t))$ are concave in $t$. (Bonus points for proving this.)

(b) For the same problem define the NLLS estimator and give conditions for its consistency and asymptotic normality. You may use "high level" assumptions if you can't prove the more basic ones hold.

(c) State how you would construct estimates of the variance-covariance matrix. Derive the Wald, Lagrange Multiplier and Likelihood Ratio Test statistics to test the hypothesis $\beta_1 = 1$. You do not have to be pedantic about details in this part.

8. Consistency and asymptotic normality of GMM [15 points, 2 pages].

Assume that agents allocate their investments and consuption across time such that the first order condition

$$\mathbb{E}\left[\beta R_{t+1}\left(\frac{C_{t+1}}{C_t}\right)^{\gamma} - 1 \,\middle|\, Z_t\right] = 0$$

holds for $\beta = \beta_0$ and $\gamma = \gamma_0$. Here, $\beta_0$ is the non-stochastic discount factor, $\gamma_0$ is the parameter of risk aversion, $C_t$ is consumption at time $t$, $R_{t+1}$ is the return on a risky asset with price $P_t$ such that $R_{t+1} = P_{t+1}/P_t$. A sample $\{C_t, P_t, Z_t\}_{t=1}^{T}$ is observed. $Z_t$ are variables that represent the information available at time $t$.

(a) Very briefly derive the above equation from the standard representative agent model with power utility. (See Hansen and Singleton (1982, Econometrica)).[1] Why is rational

[1] Hansen, Lars Peter, and Kenneth J. Singleton. "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models." *Econometrica* 50, no. 5 (Sep. 1982): 1269–1286.

expectation assumption crucial for this derivation?

(b) Assume that you have a set of random variables $Z_t$ where $Z_t$ takes values in $\mathbb{R}^d$ for some constant $d$. Also define the parameter space $\Theta \subset \mathbb{R}^2$. Define

$$\varrho_t(\beta, \gamma) = \beta R_{t+1} \left( \frac{C_{t+1}}{C_t} \right)^{\gamma} - 1$$

Assuming that $\{\varrho_t(\beta_0, \gamma_0) z_t\}_{t=1}^{T}$ are serially uncorrelated (extra credit: which additional assumptions would be needed for that in the economic model?), write down the criterion function for a GMM estimator of $(\beta, \gamma)$ that is based on instruments $z_t$. Pay particular attention to the weighting matrix. How would you estimate the weight matrix?

(c) Show that your GMM estimator is consistent and find the limiting distribution of your GMM estimator for $(\beta, \gamma)$. State any additional assumptions you may need.

## Useful Hints

Here are a few things that you should keep in mind when checking *consistency*.

1. Continuous mapping theorem: If $\hat{\theta}$ is consistent for $\theta$ and $g$ is a continuous function, then $g(\hat{\theta})$ is consistent for $g(\theta)$.

2. A good way to check consistency is to check whether the limit of your objective function is in fact minimized at the truth.

3. In extremum estimation problems, with smooth limit objective functions, we can check for inconsistency by looking whether the first-order condition is satisfied by the parameter value $\theta_0$we want to estimate:

$$\nabla_\theta Q(\theta_0) = 0$$

   Failure of this condition implies inconsistency. Note that for multivariate $\theta$, if even one equation in this system fails, it may imply inconsistency of *all* the components, unless equations are independent. For concave $g(z, \theta)$ functions, the FOC is a sufficient condition for identifiability, so we can go the other way and infer the identifiability condition if this is satisfied.

4. GMM/Nonlinear IV estimators: These are based on a moment condition (i) $E(g(z_t, \theta_0)) = 0$ under the true value $\theta_0$ and (ii) $E(g(z_t, \theta)) \neq 0$ under other values of $\theta$. Usually one checks the first part for consistency and assumes (this is wishful thinking) that the second part holds (particularly for non-linear $g$ functions). Many common estimators can be thought of as GMM estimators. For instance, consistency of OLS is based on the fact that the regressors are uncorrelated with the error terms, so it can be considered a GMM estimator with $g(z_t, \theta) = x_t \epsilon_t = x_t(y_t - x_t\beta)$. IV (2SLS) is based on $g(z_t, \theta) = w_t(y_t - x_t\beta)$ where $w_t$ are the instruments.

5. If $u \sim N(0, 1)$, then $E[u|u < L] = -\frac{\phi(L)}{\Phi(L)}$, called Mill's ratio.