International School of Economics at TSU
Econometrics II
Lasha Chochua

Problem Set 6

Instructions: You are encouraged to solve the problems before the recitation. Additionally, you are encouraged to work in groups. It is **not mandatory** to submit solutions unless stated otherwise. However, if you would like to share your solution, I would be happy to review it.

Problem 1 (Truncated Normal Moment). Let $Y^* \sim N(\alpha, \sigma_Y^2)$. Prove that

$$\mathbb{E}[(Y^* - \alpha)\, Y^* \mid Y^* > 0] = \sigma_Y^2.$$

Hint: decompose $(Y^* - \alpha)Y^* = (Y^* - \alpha)^2 + \alpha(Y^* - \alpha)$, *then standardize* $Z = (Y^* - \alpha)/\sigma_Y$ *and evaluate* $\int_a^\infty z^2 \phi(z)\, dz$ *by parts using* $z\phi(z) = -\phi'(z)$.

Explain in one sentence why this identity matters for deriving OLS bias under censoring.

Solution.

1: Decompose.

$$\mathbb{E}[(Y^* - \alpha)Y^* \mid Y^* > 0] = \mathbb{E}[(Y^* - \alpha)^2 \mid Y^* > 0] + \alpha\, \mathbb{E}[(Y^* - \alpha) \mid Y^* > 0].$$

2: Standardize.

Let $Z = (Y^* - \alpha)/\sigma_Y \sim N(0, 1)$ and $a = -\alpha/\sigma_Y$, so $\{Y^* > 0\} = \{Z > a\}$. Since $(Y^* - \alpha)^2 = \sigma_Y^2 Z^2$ and $Y^* - \alpha = \sigma_Y Z$:

$$\mathbb{E}[(Y^* - \alpha)Y^* \mid Y^* > 0] = \sigma_Y^2\, \mathbb{E}[Z^2 \mid Z > a] + \alpha\sigma_Y\, \mathbb{E}[Z \mid Z > a].$$

3: Compute $\mathbb{E}[Z \mid Z > a]$.

$$\mathbb{E}[Z \mid Z > a] = \frac{\int_a^\infty z\, \phi(z)\, dz}{1 - \Phi(a)}.$$

Differentiating $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ gives $\phi'(z) = -z\phi(z)$, so $z\phi(z) = -\phi'(z)$. Therefore:

$$\int_a^\infty z\,\phi(z)\,dz = \Big[-\phi(z)\Big]_a^\infty = \phi(a).$$

Hence $\mathbb{E}[Z \mid Z > a] = \phi(a)/[1-\Phi(a)] \equiv \lambda_a$.

4: Compute $\mathbb{E}[Z^2 \mid Z > a]$.

$$\mathbb{E}[Z^2 \mid Z > a] = \frac{\int_a^\infty z^2\,\phi(z)\,dz}{1-\Phi(a)}.$$

Write $z^2\phi(z) = z \cdot (-\phi'(z))$ and integrate by parts with $u = z$, $dv = -\phi'(z)\,dz$, so $du = dz$ and $v = -\phi(z)$:

$$\int_a^\infty z^2\phi(z)\,dz = \Big[-z\phi(z)\Big]_a^\infty + \int_a^\infty \phi(z)\,dz = a\phi(a) + [1-\Phi(a)].$$

Hence $\mathbb{E}[Z^2 \mid Z > a] = a\lambda_a + 1$.

5: Combine.

Substituting Steps 3 and 4, and recalling $a = -\alpha/\sigma_Y$ so $\sigma_Y^2 a = -\alpha\sigma_Y$:

$$\sigma_Y^2(a\lambda_a + 1) + \alpha\sigma_Y\lambda_a = \underbrace{\sigma_Y^2 a}_{-\alpha\sigma_Y}\,\lambda_a + \sigma_Y^2 + \alpha\sigma_Y\lambda_a = -\alpha\sigma_Y\lambda_a + \sigma_Y^2 + \alpha\sigma_Y\lambda_a = \sigma_Y^2. \qquad \checkmark$$

Why it matters. This identity shows that $\mathbb{E}[(Y^* - \alpha)Y^* \mid Y^* > 0] = \sigma_Y^2$ regardless of the degree of censoring, which is the key step in Goldberger's (1981) argument that $\mathbb{E}[XY^* \mid Y^* > 0] = \Sigma\beta$ under joint normality – the building block of Greene's formula $\beta^{BLP} = \beta(1-\pi)$.

Problem 2 (Tobit Model). Consider the Tobit model $Y^* = \beta_0 + \beta_1 X + e$, $e \mid X \sim N(0, \sigma^2)$, $Y = \max(Y^*, 0)$. You estimate the model on $n = 1{,}000$ households and obtain $\hat\beta_0 = -5$, $\hat\beta_1 = 2$, $\hat\sigma = 4$, where X denotes years of education.

(a) Compute the predicted censoring probability at $X = 3$.

(b) Compute the predicted censored mean $\hat m(3)$ and verify that $\hat m(3) \geq \hat\beta_0 + \hat\beta_1 \cdot 3$.

(c) An OLS regression on the full sample gives $\hat\beta_1^{\text{OLS}} = 0.7$ with estimated censoring proportion $\hat\pi = 0.65$. Is this consistent with Greene's formula? What does it imply about the reliability of OLS here?

Solution.

(a) At $X = 3$: $\hat{\beta}_0 + \hat{\beta}_1 \cdot 3 = -5 + 6 = 1$. The censoring probability is:

$$\Pr[Y^* < 0 \mid X = 3] = \Phi\left(\frac{-1}{4}\right) = \Phi(-0.25) \approx 0.401.$$

About 40% of individuals with 3 years of education are predicted to have $Y = 0$.

(b) Let $z = (X'\hat{\beta})/\hat{\sigma} = 1/4 = 0.25$. The censored mean formula gives:

$$\hat{m}(3) = (X'\hat{\beta}) \cdot \Phi(z) + \hat{\sigma} \cdot \phi(z) = 1 \times \Phi(0.25) + 4 \times \phi(0.25).$$

Using $\Phi(0.25) \approx 0.599$ and $\phi(0.25) \approx 0.387$:

$$\hat{m}(3) \approx 0.599 + 4 \times 0.387 = 0.599 + 1.548 = 2.15.$$

Since $X'\hat{\beta} = 1$ and $\hat{m}(3) = 2.15 > 1$, the inequality $\hat{m}(x) \geq m^*(x)$ holds as expected – the censored mean lies above the latent mean because positive realizations are kept while negative ones are collapsed to zero.

(c) Greene's formula predicts:

$$\beta_1^{BLP} \approx \beta_1(1 - \pi) = 2 \times (1 - 0.65) = 0.70.$$

The OLS estimate $\hat{\beta}_1^{\text{OLS}} = 0.7$ matches exactly. With 65% censoring, OLS recovers only 35% of the true slope – attenuated by a factor of roughly three. OLS should not be trusted here; Tobit MLE or CLAD should be used instead.

Problem 3 (Selection Bias and Heckman). A researcher wants to estimate the returns to education using observed wages. Wages are only observed for employed workers; employment itself is a choice.

(a) Write down Heckman's two-equation model (outcome equation and selection equation) and state the joint normality assumption. Define σ_{21} and explain its role.

(b) Derive the selected-sample CEF $\mathbb{E}[Y \mid X, Z, S = 1]$ and explain why OLS on the selected sample is biased when $\sigma_{21} \neq 0$.

(c) Describe the Heckit two-step estimator. Why must standard errors be corrected after step two?

(d) Propose an exclusion restriction for this wage application and explain why it must satisfy two conditions.

Solution.

(a) Heckman's model:

$$Y^* = X'\beta + e \quad \text{(outcome equation)},$$

$$S^* = Z'\gamma + u, \quad S = \mathbf{1}\{S^* > 0\}, \quad Y = Y^* \text{ if } S = 1 \text{ and missing otherwise.}$$

Joint normality assumption:

$$\begin{pmatrix} e \\ u \end{pmatrix} \sim N\left(0, \begin{pmatrix} \sigma^2 & \sigma_{21} \\ \sigma_{21} & 1 \end{pmatrix}\right).$$

The parameter $\sigma_{21} = \text{Cov}(e, u)$ measures how strongly the unobservables driving the wage offer co-move with those driving the employment decision. If $\sigma_{21} \neq 0$, the sample of observed wages is not a random draw from the population in terms of unobserved wage determinants.

(b) Project e onto u: write $e = \sigma_{21} u + \varepsilon$ where $\varepsilon \perp u$. Then:

$$\mathbb{E}[e \mid Z, S = 1] = \sigma_{21}\, \mathbb{E}[u \mid u > -Z'\gamma] + \mathbb{E}[\varepsilon \mid u > -Z'\gamma] = \sigma_{21}\, \lambda(Z'\gamma),$$

using $u \sim N(0, 1)$ so $\mathbb{E}[u \mid u > -Z'\gamma] = \lambda(Z'\gamma)$ and $\mathbb{E}[\varepsilon \mid u > -Z'\gamma] = 0$ by independence. Therefore:

$$\mathbb{E}[Y \mid X, Z, S = 1] = X'\beta + \sigma_{21}\, \lambda(Z'\gamma).$$

OLS on the selected sample omits $\sigma_{21}\lambda(Z'\gamma)$. Since $\lambda(Z'\gamma)$ is generally correlated with X, this is omitted variable bias. The direction depends on $\text{sgn}(\sigma_{21})$: if $\sigma_{21} > 0$ (more able workers are more likely to be employed), OLS overstates the returns to education.

(c) *Step 1:* Run probit of S_i on Z_i to obtain $\hat{\gamma}$. *Step 2:* Construct $\hat{\lambda}_i = \phi(Z_i'\hat{\gamma})/\Phi(Z_i'\hat{\gamma})$ for each selected observation and run OLS of Y_i on $(X_i, \hat{\lambda}_i)$ in the selected subsample. Standard errors from step two are inconsistent because $\hat{\lambda}_i$ is a generated regressor – it carries sampling variability from step one that the step-two OLS formula ignores. The Heckman (1979) corrected covariance matrix or full bootstrap of both steps is required.

(d) A natural exclusion restriction is *non-labor income* (e.g., spousal earnings or dividend income). It satisfies the two required conditions: (i) **relevance** – higher non-labor income reduces the financial need to work, so it affects the employment decision S^*; and (ii) **exclusion** – conditional on individual characteristics, non-labor income does not directly determine the wage offer Y^*, since employers set wages based on productivity, not on the worker's private income sources.

Problem 4 (CQR – Identification and Estimation). Consider the censored quantile regression (CQR) model of Powell (1986):

$$Y^* = X'\beta + e, \qquad Q_\tau[e \mid X] = 0, \qquad Y = \max(Y^*, 0).$$

(a) Using the equivariance property of quantiles, show that $Q_\tau[Y \mid X] = \max(X'\beta, 0)$.

(b) For what values of X is $Q_\tau[Y \mid X]$ identified from the censored data? What goes wrong for observations where $X'\beta \leq 0$?

(c) Write down the CQR criterion function $M_n(\beta; \tau)$ using the check function $\rho_\tau(u) = u(\tau - \mathbf{1}\{u < 0\})$. What special case does $\tau = 0.5$ reduce to?

(d) State one key computational difficulty with CQR that does not arise with Tobit, and explain how practitioners typically deal with it.

Solution.

(a) The equivariance property of quantiles states: for any monotone increasing function h, $Q_\tau[h(W) \mid X] = h(Q_\tau[W \mid X])$. Since $\max(y, 0)$ is monotone increasing in y:

$$Q_\tau[Y \mid X] = Q_\tau[\max(Y^*, 0) \mid X] = \max(Q_\tau[Y^* \mid X], 0) = \max(X'\beta, 0),$$

where we used $Q_\tau[Y^* \mid X] = X'\beta + Q_\tau[e \mid X] = X'\beta$.

(b) $Q_\tau[Y \mid X]$ is identified from the censored data only when $X'\beta > 0$. In this region the censoring constraint is not binding for the τ-th quantile – the censored and uncensored quantile functions coincide. When $X'\beta \leq 0$, we have $Q_\tau[Y \mid X] = 0$ for any such β, so β is not identified from these observations. Identification requires a strictly positive fraction of the population to satisfy $X'\beta > 0$.

(c) The CQR criterion is:

$$M_n(\beta; \tau) = \frac{1}{n} \sum_{i=1}^{n} \rho_\tau(Y_i - \max(X_i'\beta, 0)), \qquad \rho_\tau(u) = u(\tau - \mathbf{1}\{u < 0\}).$$

At $\tau = 0.5$: $\rho_{0.5}(u) = \frac{1}{2}|u|$, so $M_n(\beta; 0.5) \propto \frac{1}{n} \sum |Y_i - \max(X_i'\beta, 0)|$, which is the CLAD criterion. CQR at $\tau = 0.5$ reduces to CLAD.

(d) The criterion $M_n(\beta; \tau)$ is **not globally convex** in β – unlike the Tobit log-likelihood, which is globally concave after Olsen's reparameterization. Standard gradient algorithms may therefore converge to a local rather than global minimum. Practitioners deal with this by using **multiple random starting values** and selecting the solution with the lowest criterion value across all starting points.

Problem 5 (CLAD). Consider the censored regression model

$$Y^* = \beta_0 + \beta_1 X + e, \qquad \text{med}[e \mid X] = 0, \qquad Y = \max(Y^*, 0).$$

(a) State what $\text{med}[Y \mid X]$ equals under this model and justify your answer using the equivariance property of the median.

(b) Write down the CLAD criterion function $M_n(\beta)$ and explain intuitively why minimizing it recovers β consistently even when the Tobit normality assumption fails.

(c) You have five observations: $(X_i, Y_i) = (1,0), (2,3), (3,0), (4,7), (5,4)$. For $\hat{\beta}_0 = 0$, compare M_n at $\hat{\beta}_1 = 1$ vs. $\hat{\beta}_1 = 2$ and identify which gives the lower criterion value.

Solution.

(a) Under the model, $\text{med}[Y^* \mid X] = X'\beta$ since $\text{med}[e \mid X] = 0$. The equivariance property of medians states: for any monotone increasing h, $\text{med}[h(W) \mid X] = h(\text{med}[W \mid X])$. Since $\max(y, 0)$ is monotone increasing:

$$\text{med}[Y \mid X] = \text{med}[\max(Y^*, 0) \mid X] = \max(\text{med}[Y^* \mid X], 0) = \max(X'\beta, 0).$$

Note: equivariance holds for medians (defined by a ranking condition) but not for means, since $\mathbb{E}[\max(Y^*, 0)] \neq \max(\mathbb{E}[Y^*], 0)$ in general.

(b) The CLAD criterion is:

$$M_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} \big| Y_i - \max(X_i'\beta, 0) \big|.$$

Minimizing $M_n(\beta)$ fits the nonlinear median regression $\max(X'\beta, 0)$ to the observed data. Since the population median of Y given X equals $\max(X'\beta, 0)$ by part (a) – regardless of the shape of the error distribution – the sample analogue consistently estimates β without requiring normality or homoskedasticity.

(c) With $\hat{\beta}_0 = 0$, the criterion is $M_n(\beta) = \frac{1}{5} \sum |Y_i - \max(\beta X_i, 0)|$.

At $\hat{\beta}_1 = 1$*:*

i	X_i	Y_i	$\max(\beta X_i, 0)$	$\lvert Y_i - \max(\beta X_i, 0) \rvert$
1	1	0	1	1
2	2	3	2	1
3	3	0	3	3
4	4	7	4	3
5	5	4	5	1
			Total	9

$M_n(1) = 9/5 = 1.8$.

At $\hat{\beta}_1 = 2$*:*

i	X_i	Y_i	$\max(\beta X_i, 0)$	$\lvert Y_i - \max(\beta X_i, 0)\rvert$
1	1	0	2	2
2	2	3	4	1
3	3	0	6	6
4	4	7	8	1
5	5	4	10	6
			Total	16

$M_n(2) = 16/5 = 3.2$. Therefore $\hat{\beta}_1 = 1$ gives the lower criterion value ($1.8 < 3.2$).

Problem 6 (Quantile Regression by Hand). You observe $n = 5$ observations:

i	X_i	Y_i
1	1	2
2	2	1
3	3	5
4	4	4
5	5	7

Consider the linear quantile regression model with no intercept: $Q_\tau[Y \mid X] = \beta X$.

(a) Write down the check function $\rho_\tau(u)$ and explain asymmetry: why does $\tau = 0.9$ push the fitted line upward relative to $\tau = 0.1$?

(b) For $\tau = 0.5$ (median regression), compute the LAD criterion $M_n(\beta) = \frac{1}{5}\sum_{i=1}^{5} |Y_i - \beta X_i|$ at $\beta = 1$ and $\beta = 1.5$. Which gives a lower value?

(c) For $\tau = 0.9$, compute the check-function criterion $M_n(\beta; 0.9)$ at $\beta = 1$ and $\beta = 1.5$. Which gives a lower value, and does the ranking flip relative to part (b)? Explain why or why not.

(d) At the optimal $\hat{\beta}$, what fraction of observations should lie below the fitted line $\hat{\beta} X_i$? Verify this approximately for the better candidate in part (b).

Solution.

(a) The check function is:

$$\rho_\tau(u) = u(\tau - \mathbf{1}\{u < 0\}) = \begin{cases} \tau \cdot u & u \geq 0, \\ (1-\tau)\cdot |u| & u < 0. \end{cases}$$

For $\tau = 0.9$: residuals above the line ($u > 0$) are penalized at the costly rate 0.9, while residuals below ($u < 0$) attract the cheap rate 0.1. The optimizer therefore pushes the line upward until only 10% of observations remain above it. For $\tau = 0.1$ the penalty structure reverses, pushing the line down.

(b) LAD criterion ($\tau = 0.5$): $M_n(\beta) = \frac{1}{5}\sum |Y_i - \beta X_i|$.

At $\beta = 1$*:* residuals are $2-1 = 1$, $1-2 = -1$, $5-3 = 2$, $4-4 = 0$, $7-5 = 2$. Absolute values: $1+1+2+0+2 = 6$. $M_n(1) = 6/5 = 1.2$.

At $\beta = 1.5$*:* fitted values are $1.5, 3, 4.5, 6, 7.5$. Residuals: $0.5, -2, 0.5, -2, -0.5$. Absolute values: $0.5 + 2 + 0.5 + 2 + 0.5 = 5.5$. $M_n(1.5) = 5.5/5 = 1.1$.

$\beta = 1.5$ gives a lower value ($1.1 < 1.2$).

(c) Check-function criterion at $\tau = 0.9$:

At $\beta = 1$*:* residuals $1, -1, 2, 0, 2$. Penalties: $0.9 \times 1 + 0.1 \times 1 + 0.9 \times 2 + 0 + 0.9 \times 2 = 0.9 + 0.1 + 1.8 + 0 + 1.8 = 4.6$. $M_n(1; 0.9) = 4.6/5 = 0.92$.

At $\beta = 1.5$*:* residuals $0.5, -2, 0.5, -2, -0.5$. Penalties: $0.9 \times 0.5 + 0.1 \times 2 + 0.9 \times 0.5 + 0.1 \times 2 + 0.1 \times 0.5 = 0.45 + 0.20 + 0.45 + 0.20 + 0.05 = 1.35$. $M_n(1.5; 0.9) = 1.35/5 = 0.27$.

$\beta = 1.5$ still wins ($0.27 < 0.92$) – the ranking does not flip. The intuition: at $\tau = 0.9$, underpredictions (positive residuals) attract the costly rate 0.9. At $\beta = 1$, observations 3 and 5 generate large positive residuals (+2 each), each incurring an expensive penalty. Raising β to 1.5 converts these into small negative residuals at the cheap rate 0.1, sharply reducing the criterion.

(d) At the optimal $\hat{\beta}_\tau$, a fraction τ of observations should lie *below* the fitted line. For $\tau = 0.5$, exactly 50% should lie below. Check for $\beta = 1.5$: fitted values $1.5, 3, 4.5, 6, 7.5$. Observations with $Y_i < \hat{\beta} X_i$: obs 2 ($1 < 3$), obs 4 ($4 < 6$), obs 5 ($7 < 7.5$) – 3 out of 5, or 60%. For $\beta = 1$: only obs 2 ($1 < 2$) – 1 out of 5, or 20%. Neither achieves exactly 50% due to data discreteness, but $\beta = 1.5$ is much closer to the target.

Problem 7 (CLAD by Hand). You observe $n = 5$ observations:

i	X_i	Y_i
1	1	0
2	2	0
3	3	2
4	4	5
5	5	6

Assume the model $Y^* = \beta X + e$, $\text{med}[e \mid X] = 0$, $Y = \max(Y^*, 0)$, with no intercept.

(a) Two observations are censored ($Y_i = 0$). For each, compute the predicted latent value $\hat{Y}_i^* = \beta X_i$ at $\beta = 0.8$ and $\beta = 1.2$. Are these consistent with $Y^* \leq 0$ having occurred?

(b) Write down the CLAD criterion $M_n(\beta) = \frac{1}{5}\sum_{i=1}^{5} |Y_i - \max(\beta X_i, 0)|$ and evaluate it at $\beta = 0.8$ and $\beta = 1.2$.

(c) Which β gives the lower criterion value? Compare to what a naive LAD (ignoring censoring, fitting $Y_i = \beta X_i$ directly) would give at the same two candidate values. Does ignoring censoring bias the estimate upward or downward here?

(d) Explain why CLAD is consistent even if $e \mid X$ is heteroskedastic, whereas Tobit MLE would not be.

Solution.

(a) The censored observations are $i = 1$ ($X_1 = 1$) and $i = 2$ ($X_2 = 2$).

At $\beta = 0.8$: $\hat{Y}_1^* = 0.8$ and $\hat{Y}_2^* = 1.6$. Both are positive, which is inconsistent with $Y^* \leq 0$ having caused $Y = 0$ – the model over-predicts the latent outcome for these observations.

At $\beta = 1.2$: $\hat{Y}_1^* = 1.2$ and $\hat{Y}_2^* = 2.4$. Same problem, but worse. Neither candidate is consistent with censoring at these observations; the true β is likely smaller.

(b) CLAD criterion:

At $\beta = 0.8$:

i	X_i	Y_i	$\max(0.8X_i, 0)$	$\lvert Y_i - \max(0.8X_i, 0)\rvert$
1	1	0	0.8	0.8
2	2	0	1.6	1.6
3	3	2	2.4	0.4
4	4	5	3.2	1.8
5	5	6	4.0	2.0
			Total	6.6

$M_n(0.8) = 6.6/5 = 1.32$.

At $\beta = 1.2$*:*

i	X_i	Y_i	$\max(1.2X_i, 0)$	$\lvert Y_i - \max(1.2X_i, 0)\rvert$
1	1	0	1.2	1.2
2	2	0	2.4	2.4
3	3	2	3.6	1.6
4	4	5	4.8	0.2
5	5	6	6.0	0.0
			Total	5.4

$M_n(1.2) = 5.4/5 = 1.08$.

(c) $\beta = 1.2$ gives the lower CLAD criterion ($1.08 < 1.32$).

Naive LAD criterion $\frac{1}{5}\sum |Y_i - \beta X_i|$:

At $\beta = 0.8$*:* $0.8 + 1.6 + 0.4 + 1.8 + 2.0 = 6.6$, so $M^{LAD}(0.8) = 1.32$.

At $\beta = 1.2$*:* $1.2 + 2.4 + 1.6 + 0.2 + 0 = 5.4$, so $M^{LAD}(1.2) = 1.08$.

At these two candidates the criteria coincide because $\beta X_i > 0$ for all i, so $\max(\beta X_i, 0) = \beta X_i$. The difference emerges for small β such that $\beta X_i < 0$ for the censored observations: naive LAD then computes $|0 - \beta X_i| = |\beta||X_i| > 0$ and is pushed toward larger β to reduce the penalty. CLAD instead computes $|0 - \max(\beta X_i, 0)| = 0$ and correctly treats the censored observation as perfectly explained. Ignoring censoring therefore biases the naive estimator **upward**.

(d) CLAD is consistent under the single assumption $\text{med}[e \mid X] = 0$, which places no restriction on $\text{Var}(e \mid X)$. The argument: the population CLAD criterion is minimized at the true

β because the population median of Y given X equals $\max(X'\beta, 0)$ regardless of how the variance of e varies with X.

Tobit MLE is derived under $e \mid X \sim N(0, \sigma^2)$, which requires both normality and homoskedasticity. If $\text{Var}(e \mid X) = \sigma^2(X)$ varies with X, the censoring probability $\Phi(-X'\beta/\sigma(X))$ is misspecified in the likelihood, the score equations are incorrect, and the MLE converges to a pseudo-true value that generally differs from the true β.