Problem 1.

Solution:

a) We consider a continuous rv Y^* with density $f_{Y^*}(\cdot)$ alongside a censored version of Y^* given by $Y = \min\{Y^*, C\}$ for a known C. We make the assumption that Y^* has a density function that is everywhere positive.

We seek to find the c which solves $\min_{c \in \mathbb{R}} E[\rho_\tau(Y - c)]$, where $\rho_\tau(u) = (\tau - \mathbb{1}(u < 0))u$ for a given $\tau \in (0, 1)$. We begin by computing the expectation for a given value of c. Suppose first that $P(Y < C) > \tau$.

We note that when $P(Y < C) > \tau$, we must have that $q_\tau(Y^*) < C$. That is, the $100\tau\%$ quantile of Y^* is less than C. We also remark that $\forall y < C$, $f_{Y^*}(y) = f_Y(y)$. We consequently have that $q_\tau(Y) = q_\tau(Y^*)$, since $\forall c \leq C$, it follows that

$$\int_{-\infty}^c f_{Y^*}(y)\, dy = \int_{-\infty}^c f_Y(y)\, dy \iff \tau = \int_{-\infty}^{q_\tau(Y^*)} f_{Y^*}(y)\, dy = \int_{-\infty}^{q_\tau(Y^*)} f_{Y^*}(y)\, dy$$

Thus, $q_\tau(Y^*) = q_\tau(Y)$ provided $P(Y < C) > \tau$.

Our approach will be as follows: we seek to show that $E[\rho_\tau(Y - c)]$ is increasing on the interval $(-\infty, c^*]$ and decreasing on the interval (c^*, ∞), where $q_\tau(Y^*)$ is the $100\tau\%$ quantile of Y^*. To start, we consider the case where $c \leq c^* = q_\tau(Y^*)$. We then have that

$$
\begin{aligned}
E[\rho_\tau(Y - c)] =& E[\rho_\tau(Y - c) \mid Y < c]P(Y < c) + E[\rho_\tau(Y - c) \mid Y \in [c, c^*)]P(Y \in [c, c^*)) \\
&+ E[\rho_\tau(Y - c) \mid Y \geq c^*] \\
=& E[(\tau - \mathbb{1}(Y < c))(Y - c) \mid Y < c]P(Y < c) \\
&+ E[(\tau - \mathbb{1}(Y < c))(Y - c) \mid Y \in [c, c^*)]P(Y \in [c, c^*)) \\
&+ E[(\tau - \mathbb{1}(Y < c))(Y - c) \mid Y \geq c^*] \\
=& E[(\tau - 1)(Y - c) \mid Y < c]P(Y < c) + E[\tau(Y - c) \mid Y \in [c, c^*)]P(Y \in [c, c^*)) \\
&+ E[(\tau(Y - c) \mid Y \geq c^*] \\
=& E[(\tau - 1)(Y - c + c^* - c^*) \mid Y < c]P(Y < c) \\
&+ E[\tau(Y - c + c^* - c^*) \mid Y \in [c, c^*)]P(Y \in [c, c^*)) \\
&+ E[(\tau(Y - c + c^* - c^*) \mid Y \geq c^*]P(Y \geq c^*) \\
=& E[(\tau - 1)(Y - c^*) \mid Y < c]P(Y < c) + (\tau - 1)(c^* - c)P(Y < c) \\
&+ E[\tau(Y - c^*) \mid Y \in [c, c^*)]P(Y \in [c, c^*)) + \tau(c^* - c)P(Y \in [c, c^*)) \\
&+ E[\tau(Y - c^*) \mid Y \geq c^*]P(Y \geq c^*) + \tau(c^* - c)P(Y \geq c^*) \\
=& E[|Y - c^*|] + 2E[Y - c^* \mid Y \in [c, c^*)]P(Y \in [c, c^*)) \\
&+ (c^* - c)[(\tau - 1)P(Y < c) + \tau P(Y \geq c)] \\
=& E[|Y - c^*|] + 2E[Y - c^* \mid Y \in [c, c^*)]P(Y \in [c, c^*)) + (c^* - c)[\tau - P(Y < c)]
\end{aligned}
$$

We now observe that $E[Y - c^* \mid Y \in [c, c^*)]$ is negative and thus minimized when $Y = c$

(since $c \leq c^*$). As such, we have the following bound:

$$E[\rho_\tau(Y - c)] \geq E[|Y - c^*|] + (c^* - c)P(Y \in [c, c^*)) + (c^* - c)(\tau - P(Y < c))$$
$$= E[|Y - c^*|] + (c^* - c)[P(Y \leq c^*) - P(Y < c) + \tau - P(Y < c)]$$
$$= E[|Y - c^*|] + (c^* - c)[P(Y \leq c^*) + \tau - 2P(Y < c)]$$

We note that $E[|Y - c^*|]$ is constant with respect to c. Furthermore, $\forall c \leq c^*$, we have that $P(Y < c) \leq P(Y \leq c^*) = \tau$, implying that $P(Y \leq c^*) + \tau - 2P(Y < c) > 0$ (the strict inequality holds when $P(Y < c) < P(Y \leq c^*)$ - i.e. when $f_Y(\cdot)$ has positive mass on $[c, c^*)$. We have that $f_{Y^*}(\cdot)$ is everywhere positive, and we also have that $f_Y(\cdot) = f_{Y^*}(\cdot)$ on $[c, c^*)$ and is thus also positive. Hence, this condition is satisfied). Thus, given that $c \leq c^*$, the expression on the right is minimized by setting $c = c^*$!

We now must also check $c > c^*$. Consider c_1, c_2 such that $c^* < c_2 < c_1$. We will claim that $E[\rho_\tau(Y - c_2)] \leq E[\rho_\tau(Y - c_1)]$. To see this, we can compute (in a very similar fashion to the above example):

$$E[\rho_\tau(Y - c_1)] = E[|Y - c_2|] + (c_2 - c_1)(P(Y \leq c_2) + \tau - 2P(Y < c_1))$$
$$> E[|Y - c_2|]$$

since $P(Y \leq c_2) \leq P(Y < c_1)$ and also $\tau \leq P(Y < c_1)$, which both imply that $P(Y \leq c_2) + \tau - 2P(Y < c_1) < 0$. Similarly, as $c_2 < c_1$, we have that $(c_2 - c_1) < 0$, which makes the right expression positive. This implies that $E[\rho_\tau(Y - c)]$ is decreasing in c $\forall c > c^*$. In summary, we have shown that $c^* = q_\tau(Y^*) = \arg\min_{c \in \mathbb{R}} E[\rho_\tau(Y - c)]$. So, in the case where $P(Y < C) > \tau$, the minimizer is unique.

We now assume that $P(Y < C) \leq \tau$. This implies that $q_\tau(Y^*) > C$. We then compute that

$$E[\rho_\tau(Y - c)] = E[(\tau - \mathbb{1}(Y < c))(Y - c)]$$
$$= \tau E[Y - c] - E[(Y - c)\mathbb{1}(Y < c)]$$
$$= \tau E[Y - c] - E[(Y - c) \mid Y < c]P(Y < c)$$
$$= \tau E[Y - c] - E[Y - c]P(Y < c)$$
$$= E[Y - c](\tau - P(Y < c))$$

Note that $E[Y - c]$ is decreasing in c and $P(Y < c)$ is increasing in c (thus $-P(Y < c)$ is decreasing in c), implying that this will be minimized when $c = C$.

b) We now let $Y^* = X'\beta_0 + \varepsilon$ for finite dimensional vector of independent random variables X and $\beta_0 \in B$, where $B \subseteq \mathbb{R}^{d_x}$ is a parameter space which is assumed to be compact. We assume that the $100\tau\%$ quantile of ε given X is zero: that is, $q_\tau(\varepsilon \mid X) = 0$. We additionally make the assumption that the conditional density $f_{\varepsilon \mid X}(\cdot \mid X)$ is everywhere positive $\forall x \in supp(X)$. We consider the criterion function

$$\hat{Q}_n(\beta) = \frac{1}{n}\sum_{i=1}^n \rho_\tau(Y_i - \min\{X_i'\beta, C\})$$

where $\{(Y_i, X_i)\}_{i=1}^n$ is an i.i.d. random sample of random variables $(Y, X')'$. We also impose that $E[|\varepsilon|] < \infty$ and $E[\|X\|] < \infty$. We wish to show the uniform convergence of the criterion function as $n \to \infty$ and to find the limiting criterion function.

To demonstrate uniform convergence, it suffices to verify that some Uniform Law of Large Numbers holds and use this to show uniform convergence in probability of the criterion function. In particular, we verify that the assumptions of the Newey and McFadden (94) ULLN (referred to in the notes as ULLN1) hold in our context. We must verify the following four conditions:

(a) The parameter space B is compact

(b) $\forall \beta \in B$, $\hat{Q}_n(\beta)$ is continuous with probability 1

(c) $\forall \beta \in B$, $\hat{Q}_n(\beta)$ is dominated by a measurable function $G(\beta)$

(d) $E[|G(\beta)|] < \infty$

If all of these are satisfied, then we have that $\sup_{\beta \in B} \|\hat{Q}_n(\beta) - E[\hat{Q}_n(\beta)]\| \to_p 0$.

We start with 1 - the easiest one! We assume B is compact, so this is not too arduous to verify. We now must show that $\hat{Q}_n$ is continuous in X_i and Y_i with probability 1. It suffices to show that $\rho_\tau(Y_i - \min\{X_i'\beta, C\})$ is continuous at each point β with probability 1. We note that

$$\rho_\tau(Y_i - \min\{X_i'\beta, C\})) = (\tau - \mathbb{1}(Y_i < \min\{X_i'\beta, C\}))(Y_i - \min\{X_i'\beta, C\})$$

For a fixed (Y_i, X_i), this function is discontinuous only at the point β such that where $Y_i = X_i'\beta \leq C$. Since we assume that $Y_i = X_i'\beta + \varepsilon_i$ and $f_{\varepsilon|X}(\cdot \mid X)$ is everywhere positive $\forall X$ and Y^* is continuous, it follows that $Y_i = X_i'\beta$ with probability zero. Thus, condition 2 is satisfied.

For condition 3, we must find a dominating function for $\hat{Q}_n(\beta)$. Again, it suffices to find a dominating function for $\rho_\tau(Y_i - \min\{X_i'\beta, C\})$, since $\hat{Q}_n(\beta)$ is an average of these evaluated at different (X_i, Y_i) pairs. We observe that

$$
\begin{aligned}
\|\rho_\tau(Y_i - \min\{X_i'\beta, C\})\| &= \|(\tau - \mathbb{1}(Y_i < \min\{X_i'\beta, C\}))(Y_i - \min\{X_i'\beta, C\})\| \\
&\leq \|(\tau - \mathbb{1}(Y_i < \min\{X_i'\beta, C\}))\| \|Y_i - \min\{X_i'\beta, C\}\| \\
&\leq \|\tau + 1\|(\|Y_i - X_i'\beta\| + \|Y_i - C\|) \\
&\leq 2(\|Y_i\| + \|X_i'\beta\| + \|Y_i\| + |C|) \\
&\leq 2(2\|Y_i\| + \|X_i'\|\|\beta\| + |C|) \\
&\leq 2(2y + x \cdot \mathrm{diam}(B) + |C|)
\end{aligned}
$$

In other words, we have found a function $G(x, y) = 2(2y + x\mathrm{diam}(B) + |C|)$ which dominates $\hat{Q}_n(\beta)$ [it may not be the tightest bound ever constructed, but it works :)]. We let $y = \|Y_i\|$, $x = \|X_i\|$, and $\mathrm{diam}(B)$ refer to the diameter of the parameter space B (which we know is finite due to the fact that B is compact).

Finally, we must show that his dominating function is integrable. We assumed that $E[\|X_i\|] < \infty$, and furthermore that $E[|\varepsilon|] < \infty$. Since $Y_i = \min\{X_i'\beta + \varepsilon, C\}$ and

$|C| < \infty$, we have that

$$E[\|Y_i\|] = E[\|X_i'\beta + \varepsilon\|] \vee E[|C|] \le E[\|X_i\|]\|\beta\| + E[\|\varepsilon\|] \vee E[|C|]$$
$$\le (E[\|X_i\|]\text{diam}(B) + E[\|\varepsilon\|]) \vee E[|C|]$$
$$< \infty$$

Thus, $E[\|Y_i\|] < \infty$ as well. Putting this together, we have that

$$E[\|G(x,y)\|] = E[2(2\|Y_i\| + \|X_i\| \cdot \text{diam}(B) + |C|)] < \infty$$

which demonstrates that condition 4 is satisfied. We are done! We have verified the four conditions of the ULLN and thus we conclude that

$$\hat{Q}_n(\beta) \to_p Q(\beta)$$

where

$$Q(\beta) = E[\rho_\tau(Y - \min\{X'\beta, C\})]$$

c) We would like to show that $\beta_0 = \arg\min_{\beta \in B} E[\rho_\tau(Y - \min\{X'\beta, C\})]$. We note that since $Y = \min\{Y^*, C\}$ and $Y^* = X'\beta_0 + \varepsilon$, it follows that $Y^* - X'\beta_0 = \varepsilon$. We know that the conditional $100\tau\%$ quantile of ε given X is zero. This fits exactly into the framework of the first problem! Here, we have that $c = \min\{X'\beta, C\}$. From (a), we know that when $P(Y < C) \ge \tau$ the expectation is minimized when $c = q_\tau(Y^*)$ (the $100\tau\%$ percentile of Y^* given X). It is therefore minimized at β_0, since the $100\tau\%$ quantile of Y^* will be $X'\beta_0$. When $P(Y < C) \ge \tau$, the expectation will be minimized by $\beta_0 = C$ in a similar argument as part (a).

d) We know that if $E[XX'\mathbb{1}(X\beta_0 < C)]$ has full rank, then X is invertible and β_0 is the unique solution to $X'\beta_0 = q_\tau(Y)$. This implies that β_0 is the unique parameter value which minimizes $Q(\beta)$.

□

Problem 2.

Solution:

□

In the below results, we observe that independent draw offers lower RMSE than same draw in every scenario we considered. Furthermore, it has smaller bias for $\hat{\beta}$ (in absolute value) in every situation except for $S = 25$. However, in that case, we see that the standard deviation of the coefficient for independent draw is lower than that of same draw. The standard deviation of each component of $\hat{\theta}^{S2}$ is less than the corresponding component of $\hat{\theta}^S$ in each scenario tested.

It appears that increasing the sample size S with the non-normalized simulation draws generally reduces the RMSE. However, the pattern is much more clear with same draw vs independent draw. There is a more clear decrease for same draw, which makes sense: if

Table 1: Summary statistics, $S = 50$

Estimator	Bias	Standard Deviation	RMSE
$\hat{\alpha}^S$	0.0018	0.0779	0.0779
$\hat{\alpha}^{S2}$	-0.0019	0.0740	0.0740
$\hat{\beta}^S$	0.0857	0.4129	0.4220
$\hat{\beta}^{S2}$	-0.0557	0.3142	0.3191
$\hat{\sigma^2}^S$	0.7721	2.2022	2.3337
$\hat{\sigma^2}^{S2}$	0.0647	1.5857	1.5870

Table 2: Summary statistics, $S = 100$

Estimator	Bias	Standard Deviation	RMSE
$\hat{\alpha}^S$	0.0013	0.0758	0.0758
$\hat{\alpha}^{S2}$	-0.0030	0.0750	0.0750
$\hat{\beta}^S$	0.0934	0.3866	0.3977
$\hat{\beta}^{S2}$	0.0046	0.3275	0.3275
$\hat{\sigma^2}^S$	0.7293	2.0221	2.1497
$\hat{\sigma^2}^{S2}$	0.2621	1.6341	1.6549

Table 3: Summary statistics, $S = 25$

Estimator	Bias	Standard Deviation	RMSE
$\hat{\alpha}^S$	-0.0008	0.0799	0.0799
$\hat{\alpha}^{S2}$	-0.0013	0.0673	0.0673
$\hat{\beta}^S$	0.0852	0.4504	0.4583
$\hat{\beta}^{S2}$	-0.1609	0.2527	0.2996
$\hat{\sigma^2}^S$	0.8773	2.3495	2.5079
$\hat{\sigma^2}^{S2}$	-0.3691	0.9928	1.0592

Table 4: Summary statistics, $S = 50$, normalized draws

Estimator	Bias	Standard Deviation	RMSE
$\hat{\alpha}^S$	-0.00009	0.0732	0.0732
$\hat{\alpha}^{S2}$	-0.00073	0.0748	0.0748
$\hat{\beta}^S$	0.08712	0.3582	0.3686
$\hat{\beta}^{S2}$	0.05569	0.3415	0.3460
$\hat{\sigma^2}^S$	0.6762	2.0372	2.1465
$\hat{\sigma^2}^{S2}$	0.5156	1.8667	1.9366

Table 5: Summary statistics, $S = 100$, normalized draws

Estimator	Bias	Standard Deviation	RMSE
$\hat{\alpha}^S$	0.0014	0.0789	0.0789
$\hat{\alpha}^{S2}$	-0.0018	0.0755	0.0755
$\hat{\beta}^S$	0.0768	0.3454	0.3538
$\hat{\beta}^{S2}$	0.0609	0.3285	0.3341
$\hat{\sigma^2}^S$	0.6077	1.9365	2.0297
$\hat{\sigma^2}^{S2}$	0.5097	1.7965	1.8674

Table 6: Summary statistics, $S = 25$, normalized draws

Estimator	Bias	Standard Deviation	RMSE
$\hat{\alpha}^S$	-0.00047	0.0757	0.0757
$\hat{\alpha}^{S2}$	-0.00002	0.0728	0.0728
$\hat{\beta}^S$	0.0968	0.3593	0.3721
$\hat{\beta}^{S2}$	0.0301	0.3231	0.3245
$\hat{\sigma^2}^S$	0.8462	2.2220	2.3777
$\hat{\sigma^2}^{S2}$	0.3991	1.7629	1.8075

same draw uses the same, small sample for each X_i, a bad simulation draw is more likely and will be more likely to lead to incorrect/inconsistent simulated likelihoods (and thus bad parameter estimates). However, when the sample size S increases, the likelihood of a 'bad' sample goes down, so we expect our simulated likelihoods to be more accurate/consistent.

We contrast this with independent draw: with independent draw, there seems to be a bias-variance tradeoff. For independent draw, as we increase S, the bias of β goes down and its standard deviation increases. Intuitively, since we are using a different simulated draw to compute the likelihood for each X_i, we are going to get a different likelihood each time. On one hand, this is good because we reduce the importance of a particular draw, meaning that on average our estimates are less likely to be biased due to a bad simulation draw. However, using many different simulated draws will lead to higher variance in our estimates. This is the classic bias-variance tradeoff, and it is reflected in the results we observe. Overall, we see an increase in RMSE for the coefficients $\hat{\beta}^{2S}$ and $\hat{\sigma^2}^{2S}$ as we increase S. It is worth noting that they are still lower than the same-draw RMSE's.

We now compare with the results for normalized simulation draws. We see that normalizing the simulation draw leads to lower bias, standard deviation, and RMSE for same-draw simulated likelihood. This makes sense: if we get a bad simulation draw and use it to find the simulated likelihood sample $\{X_i\}$, the bad draw will bias the estimates for this sample. When we repeat this a bunch of times, our estimates will have high variance (since they are more susceptible to simulation error). When we correct for 'bad' draws, however, our estimates will become more consistent and less noisy. This intuitively explains why we see

lower RMSE for same-draw simulated likelihood with normalized simulation draws.

Counterintuitively, normalizing the simulation draws for independent-draw simulated likelihood seems to lead to worse performance (i.e. more bias, higher standard deviation, and higher RMSE compared with the non-normalized case). I don't exactly know why this would be, and it is entirely possible that there is some error in my code. My guess is that since we're already taking independent simulation draws, the impact of a 'bad' draw is already pretty low. If we go further and normalize samples to be mean zero with variance 1, then we may be getting rid of some draws that are 'bad' but may, for example, provide some information about the tails of the distribution (I don't know how plausible this is). It may be the case that normalizing gets rid of some information, which leads to less consistent and more noisy estimates overall. However, I don't totally understand why this would be the case; I am just rationalizing ex post :)