1) Prove the following statement: Suppose (i) $\hat{\beta}_n \rightarrow_p \beta_0 \in \mathbb{R}^s$, (ii) $\sup_{\gamma \in \Gamma} \sup_{\beta \in B(\beta_0,\varepsilon)} |L_n(\gamma, \beta) - L(\gamma, \beta)| \rightarrow_\mu 0$ for some $\varepsilon > 0$, and (iii) $L(\gamma, \beta)$ is continuous in $\beta$ at $\beta_0$ uniformly over $\gamma \in \Gamma$ (i.e., $\lim_{\beta \rightarrow \beta_0} \sup_{\gamma \in \Gamma} |L(\gamma, \beta) - L(\gamma, \beta_0)| = 0$). Then,

$$\sup_{\gamma \in \Gamma} |L_n(\gamma, \hat{\beta}_n) - L(\gamma, \beta_0)| \rightarrow_\mathbb{P} 0.$$ 

Note that condition (iii) holds if $\Gamma$ is compact and $L(\gamma, \beta)$ is continuous in $(\gamma, \beta)$ on $\Gamma \times B(\beta_0, \varepsilon)$.

2) (a) Covariance estimation: Suppose $Y_t = X'_t \beta + U_t$ for $t = 1, \ldots, T$, $\{ (X_t, U_t) : t \geq 1 \}$ are iid with $E(U_t | X_t) = 0$ a.s. and $E(U_t^2 | X_t) < \infty$ a.s.. The least squares estimator $\hat{\beta}_{LS}$ satisfies

$$\sqrt{T}(\hat{\beta}_{LS} - \beta) \rightarrow_d Z \sim N(0, \Omega)$$

as $T \rightarrow \infty$, where $\Omega := (E X_t X_t')^{-1} E U_t^2 X_t X_t' (E X_t X_t')^{-1}$. Show that the so-called Eicker-White estimator

$$\hat{\Omega} := \left( \frac{1}{T} \sum_{t=1}^T X_t X_t' \right)^{-1} \left( \frac{1}{T} \sum_{t=1}^T U_t^2 X_t X_t' \left( \frac{1}{T} \sum_{t=1}^T X_t X_t' \right)^{-1} \right),$$

where $\hat{U}_t := Y_t - X'_t \hat{\beta}_{LS}$, satisfies $\hat{\Omega} \rightarrow_p \Omega$ as $T \rightarrow \infty$. State clearly in each step of your argument which rule you use and which assumptions are needed.

(b) In the linear regression model of part a) with $\sigma^2 = E(U_t^2 | X_t) < \infty$ a.s. (conditional homoskedasticity), design a Monte Carlo study in which you compare the precision of the Eicker-White estimator with the one of the covariance matrix estimator that assumes conditional homoskedasticity. For simplicity take $EX_t X_t' = I_k$. Report finite sample bias, variance, and mean-squared error of the variance estimators. Experiment with various error distributions. Note that whenever you pin down distributions for $U$ and $X$, you know what the true $\Omega$ is.

3) (a) For the MD estimator, discussed on problem set 1, establish asymptotic normality of $\hat{\beta}_n$. [Assume here that $\sqrt{n}(\hat{\beta}_n - \tau_0) \rightarrow_d N(0, \Omega_0)]^1$.

(b) Another example of an extremum estimator is the Two-step (TS) Estimator: Suppose the data $\{ W_i : i \leq n \}$ are iid, $\hat{\tau}_n$ is a preliminary consistent estimator of a parameter $\tau_0$, $G_n(\theta, \tau) = \sum_i g(W_i, \theta, \tau)$ is a random $k$-vector that should be close to 0 when $\theta = \theta_0$, $\tau = \tau_0$, and $n$ is large (e.g., in the GMM case, $G_n(\theta, \tau) = \frac{1}{n} \sum_i g(W_i, \theta, \tau)$ and in the MD case $G_n(\theta, \tau) = \hat{\tau}(\tau) - g(\theta, \tau)$), and $A_n$ is a $k \times k$ random weight matrix. Then, the TS estimator $\hat{\theta}_n$ minimizes

$$Q_n(\theta) = ||A_n G_n(\theta, \hat{\tau}_n)||^2 / 2$$

where $\hat{\tau}_n$ is to be treated as a vector of constants. The function $g(\theta)$ does not depend on the data. When working out what $B_0$ and $\Omega_0$ are, replace estimators by their probability limits, e.g. $\hat{\tau}_n$ by $\tau_0$. To show asymptotic normality, let $B_0 \equiv \frac{\partial^2}{\partial \theta \partial \theta'} Q_n(\theta_0) > 0$. Provide primitive assumptions for the assumption $\sqrt{n}(\hat{\theta}_n - \theta_0) \rightarrow_d N(0, \Omega_0)$. What is $\Omega_0$? Then use a Taylor expansion

$$0 = \frac{\partial}{\partial \theta} Q_n(\hat{\theta}_n) - \frac{\partial}{\partial \theta} Q_n(\theta_0) + \frac{\partial^2}{\partial \theta \partial \theta'} Q_n(\theta_0^*) (\hat{\theta}_n - \theta_0),$$

where $\theta_0^*$ lies between $\hat{\theta}_n$ and $\theta_0$. Solve for $(\hat{\theta}_n - \theta_0)$...

---

1 Additional hints: When you differentiate $Q_n(\theta)$ with respect to $\theta$, then $\hat{\tau}_n$ is to be treated as a vector of constants. The function $g(\theta)$ does not depend on the data. When working out what $B_0$ and $\Omega_0$ are, replace estimators by their probability limits, e.g. $\hat{\tau}_n$ by $\tau_0$. To show asymptotic normality, let $B_0 \equiv \frac{\partial^2}{\partial \theta \partial \theta'} Q_n(\theta_0) > 0$. Provide primitive assumptions for the assumption $\sqrt{n}(\hat{\theta}_n - \theta_0) \rightarrow_d N(0, \Omega_0)$. What is $\Omega_0$? Then use a Taylor expansion

$$0 = \frac{\partial}{\partial \theta} Q_n(\hat{\theta}_n) - \frac{\partial}{\partial \theta} Q_n(\theta_0) + \frac{\partial^2}{\partial \theta \partial \theta'} Q_n(\theta_0^*) (\hat{\theta}_n - \theta_0),$$

where $\theta_0^*$ lies between $\hat{\theta}_n$ and $\theta_0$. Solve for $(\hat{\theta}_n - \theta_0)$...
over $\theta \in \Theta$. Under appropriate assumptions (including $G_n(\theta, \tau) \xrightarrow{p} G(\theta, \tau)$) what is $Q(\theta)$? Give primitive conditions that imply consistency (in particular, when does $\theta_0$ uniquely minimize $Q(\theta)$ over $\Theta$?) Asymptotic normality will be discussed in the TA section.

4) Investigate how well the asymptotic normal distribution of the two-stage least squares (TSLS) estimator approximates its finite sample distribution. Consider the linear model

$$y_i = x_i\beta + \varepsilon_i \in \mathbb{R}$$

for $i = 1, 2, \ldots, n$, where the scalar $\beta$ equals 0. Suppose $z_i \in \mathbb{R}^K$ are iid $N(0, I_K)$. Assume that

$$x_i = z_i^T\pi + u_i,$$

where $(\varepsilon_i, u_i)$ are iid $N(0, \Omega)$, where $\Omega$ is a $2 \times 2$-matrix with diagonal and off diagonal elements 1 and $\rho$, respectively. Let $\pi = (\eta, \eta, \ldots, \eta)'$.

(i) Derive the asymptotic distribution of the TSLS estimator in this model. Does it depend on $K$, $\eta$ or $\rho$?

Let $n = 100$ and simulate $R = 5000$ data samples for all 8 parameter combinations of $K = (1; 10)$, $\eta = (.05; 1)$ and $\rho = (0; .95)$.

(ii) Plot the finite sample distribution of the TSLS estimator. Use separate graphs for each of the eight parameter combinations. Does the distribution change with $K$, $\eta$ or $\rho$?

(iii) Simulate the asymptotic normal distribution you derived in (i) for the TSLS estimator and plot it in the corresponding graphs of (ii). Comment on the goodness of the approximation.