1. (20 points) Suppose we want to estimate the following linear regression model:

\[ y_i = \beta x_i + \varepsilon_i, \]

where \( y_i \) and \( x_i \) are scalars that always assume positive values, and \( \varepsilon_i \sim i.i.d.(0, \sigma^2_{\varepsilon}) \). Unfortunately, we are unable to observe \( y_i \) directly. Instead, we have a “noisy” measure of \( y_i \),

\[ y_i^* = y_i + u_i, \]

where \( u_i \sim i.i.d.(0, \sigma^2_u) \) is measurement error.

(a) Is the OLS estimator \( \hat{\beta} \) obtained by regressing \( y_i^* \) on \( x_i \) a consistent estimator of \( \beta \)? Show why or why not.

(b) If \( E(u_i) = \gamma \) for all \( i \), is the OLS estimator \( \hat{\beta} \) consistent? If not, find the asymptotic bias associated with the estimator.

2. (20 points) Consider the following regression model:

\[ y_i = x_i\beta + \varepsilon_i, \]

where \( \varepsilon_i = c_i u_i, \ u_i \sim i.i.d.(0, \sigma^2), \) and \( c_i^2 = \exp(z_i^\prime \theta) \). Assume that \( \sigma^2 \) and \( \theta \) are unknown, whereas \( z_i \) are observed.

(a) Is the OLS estimator consistent? Is it efficient? Please explain.
(b) Define a FGLS estimator of $\beta$, if one exists. Please provide all the details of the estimation strategy.

3. (15 points) Consider the following panel data model:

$$y_{it} = x_{it}'\beta + \varepsilon_{it}, \quad t = 1, ..., T; \quad i = 1, ..., N$$

where $x_{it}$ is a $(k \times 1)$ vector, $E(\varepsilon_{it}|x_{it}) = 0$, and the following assumptions hold about the covariance structure:

$$\sigma_{it}^2 \equiv E(\varepsilon_{it}^2) = \sigma_i^2, \quad t = 1, ..., T$$

$$\sigma_{ij,ts} \equiv E(\varepsilon_{it}\varepsilon_{js}) = 0, \quad t \neq s; \quad i, j = 1, ..., N$$

$$\sigma_{ij,t} \equiv E(\varepsilon_{it}\varepsilon_{jt}) = \sigma_{ij}, \quad i \neq j; \quad t = 1, ..., T$$

Further assume that $\sigma_{ij} = \gamma(z_i - z_j)^\phi$, and $\sigma_i^2 = \alpha(w_i)^\delta$, where $\{z_i, w_i\}$ are observed and $(\alpha, \gamma, \delta, \phi)$ are unknown parameters.

Please describe in detail a FGLS strategy to estimate $\beta$.

4. (15 points) Take the linear model

$$Y = X\beta + e.$$  

Let the OLS estimator for $\beta$ be $\hat{\beta}$ and the OLS residual be $\hat{e} = Y - X\hat{\beta}$. Also let the 2SLS estimator for $\beta$ using some instrument $Z$ be $\tilde{\beta}$ and the 2SLS residual be $\tilde{e} = Y - X\tilde{\beta}$.

If $X$ is indeed endogenous, will 2SLS “fit” better than OLS, in the sense that $\tilde{e}'\tilde{e} < \hat{e}'\hat{e}$, at least in large samples?

5. (30 points) Take the linear model

$$y_i = x_i\beta + \varepsilon_i$$

where $x_i$ and $\beta$ are $(1 \times 1)$.

(a) Show that $E(x_i\varepsilon_i) = 0$ and $E(x_i^2\varepsilon_i) = 0$. Is $z_i = (x_i \quad x_i^2)$ a valid instrumental variable for estimation of $\beta$?

(b) Define the 2SLS estimator of $\beta$, using $z_i$ as an instrument for $x_i$. How does this differ from OLS?

(c) Find the efficient GMM estimator of $\beta$, based on the moment condition

$$E[z_i(y_i - x_i\beta)] = 0.$$  

Does this differ from 2SLS and/or OLS?