Part I: Do all 5 questions.

5 pts. 1) Show that the OLS estimate of $\beta$ is unbiased in the model:

$$ y = X\beta + u $$

5 pts. 2) Consider the model

$$ y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t. $$

Find an estimate for the covariance between $\hat{\beta}_1$ and $\hat{\beta}_2$.

5 pts. 3) Consider the models

$$ y_t = \alpha + \beta x_t + u_t $$

$$ x_t = a + by_t + e_t. $$

Find the relationship between the estimates of $\beta$ and $b$.

5 pts. 4) Consider the model

$$ y_t = \alpha + \beta x_t + u_t \quad t = 1,2,\ldots,100. $$

It is known that $\Sigma y_t = 100$, $\Sigma x_t = 80$, $\Sigma y_t^2 = 20000$, $\Sigma x_t^2 = 20000$, and $\Sigma x_t y_t = 15000$. Find $\hat{\alpha}$ and $\hat{\beta}$.

5 pts. 5) Consider the estimated model

$$ C_t = 100 + .7GNP_t + .01WEA_t - 10.2R_t + u_t $$

where $C_t$ = consumption

$GPN_t$ = GNP

$WEA_t$ = wealth

$R_t$ = interest rate

and numbers in parentheses are t-statistics.
a) If GNP increases by 100, how much does consumption change by?

b) Test

\[ H_0: \beta_{\text{WEA}} = 0 \]
\[ H_A: \beta_{\text{WEA}} \neq 0 \]

c) Test

\[ H_0: \beta_{\text{GNP}} = 0.8 \]
\[ H_A: \beta_{\text{GNP}} \neq 0.8 \]

Part II: Do 2 out of 3 questions.

15 pts. 1) Consider the model

\[ y = X\beta + u \]
\[ u \sim N(0, \Omega) \]

Consider the estimator \( \hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y \).

a) Show that \( E\hat{\beta} = \beta \).

b) Find the covariance matrix of \( \hat{\beta} \).

15 pts. 2) Consider the model

\[ y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i \]

Show that the residuals gotten from OLS regression are uncorrelated with any linear combination of \( x_2 \) and \( x_3 \).

15 pts. 3) Consider the relationship

\[ y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t \]

Show that the OLS estimator of \( \beta_1 \) in the equation

\[ y_t = b_0 + b_1 x_{1t} + e_t \]

is a biased estimator of \( \beta_1 \) and find an expression for the bias.
Part III: Do 2 out of 3 questions.

10 pts. 1) A report recently came out which showed that children of working mothers performed better in school than children of nonworking mothers. The equation regressed was of the form:

\[ \text{GPA}_i = \alpha + \beta W_i + u_i \]

where \( \text{GPA}_i \) = school performance

\[ W_i = \begin{cases} 
1 & \text{if child's mother works} \\
0 & \text{if not} 
\end{cases} \]

a) Was \( \hat{\beta} \) positive or negative? Explain.
b) Discuss problems with interpreting the results of this regression.
c) How would you correct the problems?

10 pts. 2) You are hired to measure the elasticity of labor supply with respect to wages using cross section data on hours worked per year and hourly wage rates. [Note: \( E = \frac{w}{h} \frac{\partial h}{\partial w} = \frac{\partial \log h}{\partial \log w} \) where \( E \) is the required elasticity, \( w \) is hourly wage rate and \( h \) is annual hours.]

a) Write down the equation to be estimated and show how to find an estimate of the elasticity.
b) Your employer criticizes your model because you are using hourly wage rates rather than annual earnings. Discuss the advantages and disadvantages of using annual earnings as a substitute for the hourly wage rate.
c) Discuss how to test if the elasticity is zero. Be precise.
d) Give some economic intuition for why it might not be unreasonable to accept the null hypothesis in part (c).

10 pts. 3) Consider a model in which there is an unlimited number of exogenous variables that can be used to explain the endogenous variable. Discuss in statistical terms the merits of trying as many different specifications of the model as you need to get a good fit. Be precise.
Part I. Do 4 out of 5 questions (40 points)

1. Let \( y = X\beta + u \) where \( u \sim N(0, \sigma^2 I) \). Let \( \hat{u} = y - X\hat{\beta} \). Find \( E(u - \hat{u})(u - \hat{u})' \).

2. Consider the following three equations:
   
   a. \( \ln W_i = \alpha_0 + \alpha_1 \text{Black}_i + \alpha_2 \text{Educ}_i + u_i \quad u_i \sim \text{iidN}(0, \sigma_u^2) \)
   
   b. \( \ln W_i = \beta_0 + \beta_1 \text{White}_i + \beta_2 \text{Educ}_i + e_i \quad e_i \sim \text{iidN}(0, \sigma_e^2) \)
   
   c. \( \ln W_i = \gamma_1 \text{Black}_i + \gamma_2 \text{White}_i + \epsilon_i \quad \epsilon_i \sim \text{iidN}(0, \sigma_\epsilon^2) \)

   Equation (a) is estimated using OLS
   
   \[
   \ln W_i = 1.2 - 0.3 \text{Black}_i + 0.07 \text{Educ}_i + u_i \\
   (0.2) \quad (0.05) \quad (0.02)
   \]

   \[ R^2 = 0.23 \quad \sigma_u^2 = 1.6 \]

   where numbers in parentheses are coefficient standard errors.

   Using this information, write down analogous results for equations (b) and (c).

3. Let \( y_t = \sum_{i=0}^{n} \beta_1 x_{t_i} + u_t \). Consider the test:

   \( H_0: \beta_2 = \beta_3 + 4\beta_1, \ 4 = \beta_6 / \beta_1 \)

   \( H_A: \beta_2 \neq \beta_3 + 4\beta_1, \ 4 \neq \beta_6 / \beta_1 \).

   Construct a Wald test statistic, report its distribution, and explain when to reject or not reject \( H_0 \).
4. Let \( y = X/3 + Z + u \) where 0 is nx1 and 7 is mx1. Let \( Q = (X|Z) \) and 
\( Ot = l \) so that \( y = QGt + u \). Consider the following estimators of 0.

a. \( O = \) first n elements of \( a = (Q'Q)^{-1} Q'y \)

b. \( O = (X'X)^{-1} X' \left[ I - Z(Z'Z)^{-1} Z' \right] y \)

c. Let \( H = [I - Z(Z'Z)^{-1} Z']X \) and \( O = (H'X)^{\cdot} \). 

Explain in detail why all three of these estimators are identical.

5. Let \( y = X0 + u \) where \( u \sim N(0,\sigma^2) \). Let \( 0 \) be the OLS estimator of 0. 

Find \( E0 \) and \( D(0) \). Construct a Wald test statistic for 

\[ V K = \frac{\sum_{i=1}^{n} (X_i - \hat{X}_i)^2}{\sigma^2} \]

against \( H_0: \beta_1 = \beta_2 = \beta_3, \beta_4 + \beta_5 = 5 \)

Part II. Do 1 out of 2 questions (25 points).

1w Consider the model

\[ y = X0 + u \]

where \( u \sim iidN(0,\sigma^2) \).

You are interested in the test 

\( H_0: \beta_1 = \beta_2 = \beta_3, \beta_4 + \beta_5 = 5 \)

\( H_A: \beta_1 = \beta_2 = \beta_3, \beta_4 + \beta_5 = 5 \)

a. Construct a Wald test statistic and derive its distribution.

b. Suggest how to estimate \( (0_0, \ldots, 0_n) \) while imposing the restrictions implied by \( H_0 \).

c. Suggest another test statistic for the test.
2. Let \( y = X\beta + Z\gamma + u \). Consider the regression \( y = Xb + e \). (Hint: think of \( e \) as \( e = Z\gamma + u \).) The OLS estimate of \( b \) is

\[
\hat{b} = (X'X)^{-1}X'y.
\]

a. Find \( E\hat{b} \), and explain in words when \( E\hat{b} = \beta \).

b. Let \( y = \log \text{wage} \), \( X \) a constant and education, and \( Z \) equal (unobserved) ability. Explain in words how not observing ability biases your estimate of the coefficient in education. Is the bias positive or negative? Explain.
Part I. Do three out of four questions. (45 points)

1. Let \( y_t = \sum_{L=0}^{n} \beta_i X_{it} + u_t \), \( u_t \sim iidN(0, \sigma^2) \), \( t = 1, 2, \ldots, T \). Let \( \hat{u}_t = y_t - \hat{y}_t \)
where \( \hat{y}_t = \sum_{i=0}^{n} \hat{\beta}_i X_{it} \) and \( \hat{\beta}_i \) is the OLS estimator of \( \beta_i \) for \( i = 0, 1, \ldots, n \).

Let \( u \) be the vector of errors and \( \hat{u} \) be the vector of residuals. Find the distribution of \( u - \hat{u} \).

2. Consider the normal equations:
\[ q = X'(y - X\hat{\beta}). \]

Find the covariance matrix of \( q \). Provide intuition for your answer.

3. Consider \( y = X\beta + Z\gamma + u \). Let \( \tilde{y} = (I - P_Z)y \) and \( \tilde{X} = (I - P_Z)X \). Show using matrix algebra that the OLS estimator of \( b_1 \) in \( y = \tilde{X}b_1 + e_1 \) is the same as the OLS estimator of \( b_2 \) in \( \tilde{y} = \tilde{X}b_2 + e_2 \).

4. Let \( y_t = \beta_0 + \beta_1 X_{1t} + u_t \), \( t = 1, 2, \ldots, T \). Let the t-statistic for \( \beta_1 \) be 2.73. Find \( R^2 \) for the equation.

Part II. (30 Points)

1. You have been given data set with 1000 observations. Each column has
   1. wage
   2. education
   3. male
   4. race (0 = white, 1 = black, 2 = asian).

Write a LIMDEP program to test whether the rate of return to education varies with a) race, b) sex, and c) education.
Part I. answer 3 out of 4 questions (30 points)

1. Let \( y = X\beta + u \). Assume \( u \sim N(0, \sigma^2 I) \). Find the distribution of the OLS estimator of \( \beta \). Show all work.

2. Let \( U = F_{m,n} \). Find the distribution of \( m\bar{u} \) as \( n \to \infty \). Show all work.

3. Let \( y = X\beta + u \). Let \( \hat{\beta} \) be the OLS estimate of \( \beta \). Let \( \hat{u} = y - X\hat{\beta} \). Show that \( \hat{u} \) and \( X_j \) are uncorrelated where \( X_j \) is the \( j \)th column of \( X \). Show all work.

4. Let \( s^2 = \frac{\hat{u}'\hat{u}}{[n-(m+1)]} \) where \( \hat{u} \) are residuals from the OLS regression for \( y = X\beta + u \), \( n \) is the sample size, and \( m + 1 \) is the number of columns in \( X \). Show that \( E s^2 = \sigma^2 \) where \( \sigma^2 = E \hat{u}_1^2 \).

Part II. Answer 1 out of 2 questions (20 points)

1. Assume \( y = X\beta + 2\gamma + u \). Consider running the regression \( y = Xb + e \), i.e., do not include \( Z \). Let \( \hat{b} \) be the OLS estimate of \( b \). Find the \( E\hat{b} \). Show that \( E\hat{b} = \beta \) if and only if \( X \) and \( Z \) are uncorrelated.

2. Consider the following two equations:
   a. \( \ln W_i = \beta_0 + \beta_1 \text{Educ}_i + \beta_2 \text{Age}_i + \beta_3 \text{Black}_i + \beta_4 \text{Oriental}_i + u_i \)
   b. \( \ln W_i = \gamma_1 \text{Educ}_i + \gamma_2 \text{Age}_i + \gamma_3 \text{Black}_i + \gamma_4 \text{Oriental}_i + \gamma_5 \text{White}_i + u_i \)

   Show that, given reasonable data, both equations can be estimated. Derive the relationship between the \( \beta \)'s in equation (a) and the \( \gamma \)'s in equation (b).

Part III. Answer the question (20 points)

1. If one regresses a measure of elementary school performance by a child as a function of time spent at home by the child's mother, one finds a negative relationship. Comment on this, and, if relevant, suggest how you might alternatively estimate the effect in elementary school performance of mother's time at home.
Econ 372
Midterm
March 24, 1994

Part I. Do 3 out of 4 questions. (30 points)

1. Consider the model
\[ y = X\beta + u. \]
Let \( \hat{u} \) be OLS residuals. Define
\[ e = u - \hat{u}. \]

a. Find \( Ee \) and \( Eee' \).

b. Find \( Eeu' \).

2. Let \( y_t = X_t \beta + Z_t \gamma + u_t \)
\[ u_t = \rho u_{t-1} + e_t \]
\[ e_t \sim iidN(0, \sigma^2). \]

Consider regressing \( y \) on \( X \):
\[ y = Xb + \epsilon. \]
Find \( Eb \).

3. Let \( y_t = X_t \beta + Z_t \gamma + e_t \)
\[ e_t = a u_t + b u_{t-2} + c u_{t-3} \]
\[ u_t \sim iidN(0, \sigma^2). \]

Let \( \Omega \) be the covariance matrix of \( e \).

a. What is \( \Omega \)?

b. How could you estimate \( \Omega \)?
Part I. Do three out of four questions (45 points).

1. Let \( y = X\beta + Z\gamma + u \). Show that \( Z'u = 0 \).

2. Consider \( y_1 = \beta_0 + \beta_1 \text{Black}_1 + \beta_2 \text{Asian}_1 + X_1\gamma + u_1 \)
   \[ y_1 = \delta_0 + \delta_1 \text{White}_1 + \delta_2 \text{Black}_1 + X_1\theta + \epsilon_1 \]
   \[ y_1 = \alpha_1 \text{White}_1 + \alpha_2 \text{Black}_1 + \alpha_3 \text{Asian}_1 + X_1\psi + \epsilon_1 \]

   Assuming these three equations are estimating the same relationship with the same data, write \( (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\gamma}) \) and \( (\hat{\delta}_0, \hat{\delta}_1, \hat{\delta}_2, \hat{\theta}) \) in terms of \( (\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\psi}) \).

3. Let \( y = X\beta + Z\gamma + u \).
   a. Derive a Wald statistic for \( H_0: \gamma = 0 \) vs. \( H_A: \gamma \neq 0 \).
   b. Derive a likelihood ratio test for \( H_0: \gamma = 0 \) vs. \( H_A: \gamma \neq 0 \).

4. Let \( y = X\beta + u \). Assume \( X \) is \( T \times (m+1) \). Let \( Q \) be the \( F \) statistic for the equation. Show that \( mQ \rightarrow \chi^2_m \) as \( T \rightarrow \infty \).

Part II. Do one out of two questions (20 points)

1. You are given data on output, inputs, and prices of output and inputs. You observe each of these variables over a large number of firms. Let \( y_i \) = output of firm \( i \), \( X_{ij} \) = input \( j \) of firm \( i \), \( p_i \) = output price of firm \( i \), and \( w_{ij} \) = price of input \( j \) for firm \( i \). Assume
   \[ y_i = A \prod_{j=1}^{K} x_{ij} \alpha_j u_i \]

   Suggest how to estimate \((A, \alpha_1, \alpha_2, \ldots, \alpha_k, \sigma_u^2)\). Discuss problems likely to exist.

2. Your job is to estimate the relationship between pollution and changes in weather (e.g., the greenhouse effect). Suggest what data to collect and how to estimate the relationship. Be specific about data, unit of observation, explanatory variables, etc. Suggest problems likely to exist.
3. Let $y_t = \sum_{i=0}^{n} \beta_i x_{it} + u_t$, $t = 1,2, \ldots, T$; $u_t \sim iidN(0, \sigma)$.

Consider $H_0: \beta_2 = \beta_3$ against $H_A: \beta_2 \neq \beta_3$. Construct a $t$-statistic to test the null hypothesis.

Part III. Do 1 out 1. (15 points)

1. Consider a sample where we observe wages of people every year between 1980 and 1989 (ten years) for 200 people; i.e., there are 2000 observations. You want to find the effect of gender, race, and education on log wages and how it is changing over time. Discuss how to do this using an efficient estimation procedure. Be as precise as possible.