You are hired as a consultant to measure various relationships affecting the work environment of mature men. You are given a large data set with answers to questions asked in the enclosed survey. You must write a coherent (e.g., good English, good logic, etc) proposal to answer all the following questions. Your proposal (and therefore your grade) will be evaluated according to the points indicated to the right of each question.

(1) I want to measure the effect of various characteristics on wages.

a) How can I use the available data to construct a measure of hourly wage and variables that might affect it? (Be precise about what variables I can use and how they can be constructed). (10 points)

b) Suggest some reasonable specifications for wage equations. (5 points)

c) Suggest what econometric method to use and why it is appropriate. (5 points)

d) I am particularly interested in estimating the return to education and how that varies across people. How do I do this? What test statistics do I construct? How do I use the test statistics? (15 points)

e) I am afraid that one of the variables affecting wage is experience on the job but that it might be endogenous. Why so? How can I deal with its potential endogeneity? (15 points)

(2) I am interested in how people decide whether to work.

a) How can I use the available data to construct a measure of whether or not you work and variables that might affect it? (5 points)

b) Suggest a reasonable model specification. (5 points)

c) How can I test whether marital status and number of children have different effects on the work decision for men and women? Be precise and give details (e.g., the test statistic and its asymptotic distribution). (20 points)

(3) I am interested in the joint determination of wage and hours. In particular, I think wage depends on hours in that part-time people get paid less per hour than full-time people. Also hours (supply) depends upon wage. Also, I am afraid that marital status affects hours and may be affected by wage.

a) Suggest a reasonable specification of the relationships I am interested in. Which structural parameters are identified? (20 points)
b) How can I estimate each equation? Be precise and give details. Note that marital status is binary. (20 points)

(4) I can probably get you data similar to this data for each year between 1970 and 1990. This period, income distribution has become more disbursed. Suggest how to measure how much of this is due to

a) increases in the return to education; (10 points)

b) more variation in work status; and (10 points)

c) differences between blacks and whites. (10 points)

Be specific. Provide details of equations to estimate, estimation methods, and test statistics.
Part I. Do 5 out of 6 questions (75 points)

1. Let \( y = X\beta + u \) where \( E(u'u) = \sigma^2 I \). Let \( \hat{\beta} = (X'X)^{-1}X'y \). Let \( \hat{u} = y - X\hat{\beta} \). Find \( EX'u'uX \) and \( EX'u'uX \). Explain why they are different.

2. Assume \( y_i = AX_{1i}x_{2i} + \varepsilon_i \). Explain how to estimate \( \beta_1, \beta_2 \), and the variance of \( \varepsilon_1 \) or the variance of some transformation of \( \varepsilon_1 \).

3. Let \( y_t = X_t\beta + u_t \). Assume \( u_t = \rho u_{t-1} + \alpha_t + \beta_\varepsilon_t \) where \( \varepsilon_t \) is iid \( N(0, \sigma^2) \). Find the covariance matrix of \( u \) and write the GLS estimate of \( \beta \).

4. Let \( y = X\beta + u \) where some of the variables in \( X \) are endogenous. Let \( Z \) be a set of instruments for \( X \). Assume \( \text{rank}(Z) > \text{rank}(X) \). Show how to use all of the instruments efficiently. Be sure to use the notation defined in this problem.

5. Let \( y = X\beta + u \) where some of the variables in \( X \) are endogenous. Let \( Z \) be a set of instruments for \( X \). Use the instrumental variables estimator of \( \beta \) to test \( \mathcal{H}_0: \beta_1 = \beta_2, \beta_3 = \beta_4 \) against \( \mathcal{H}_0: \beta_1 \neq \beta_2, \beta_3 \neq \beta_4 \).

6. Let \( y_t = \sum_{i=0}^{n} \beta_i x_{ti} + u_t \). Consider \( \mathcal{H}_0: \beta_1 = \beta_2, \beta_3 = \beta_4 \) against \( \mathcal{H}_A: \beta_1 \neq \beta_2, \beta_3 \neq \beta_4 \). Suggest how to construct a likelihood ratio test.

Part II. Do 3 out of 4 questions (75 points).

1. Let \( q_t^d = \alpha_0 + \alpha_1 p_t + \alpha_2 y_t + u_t^d \)
   \[ q_t^s = \beta_0 + \beta_1 p_t + \beta_2 y_t + u_t^s \]
   \[ q_t = q_t^d = q_t^s \]

   a) Determine which equations are identified.
   b) Suggest how to estimate the supply equation.
   c) Discuss problems with testing \( \mathcal{H}_0: \alpha_2 = \beta_1 = 0 \) against \( \mathcal{H}_A: \alpha_2 \neq \beta_1 \neq 0 \).
2. Let \( y_t = \alpha_0 + \alpha_1 x_t + u_t \). Assume \( x_t \) (a scalar) is not observed; instead \( z_t = x_t + \epsilon_t \) is observed where \( \epsilon_t \sim iid N(0, \sigma^2) \), \( E \epsilon_t u_t = E x_t \epsilon_t = 0 \). Let \( \nu_t = x_t + e_t \) where \( E e_t u_t = E e_t \epsilon_t = 0 \).

a) Show that the OLS estimator of \( \alpha_1 \) is biased.
b) Find a consistent IV estimator of \( \alpha_1 \).

3. Let \( y = X\beta + u \) where \( u \sim N(0, \sigma^2 I) \). Suggest an estimator for \( \sigma^2 \) and determine its plim. Be explicit about the arguments you are using to show consistency of your estimator.

4. Let \( u \sim N(0, \Omega) \). Find the distribution of \( u' \Omega^{-1} u \). Hint: Let \( RR' = \Omega^{-1} \) and find the distribution of \( Ru \) first.

Part III. Do 2 out of 3 questions (30 points).

1. Suggest how to predict the Dow-Jones stock market index. Be precise about what variables to use in predicting it, what estimation method you are using (and why), and any significant problems you expect to encounter. Define the standard deviation of your prediction.

2. Republican "supply side economics" works by tax reductions giving people incentive to work more than enough so that total revenue increases. Suggest how to estimate the labor supply effect of lower taxes (i.e., higher net wages). Be precise about what explanatory variables to use, what estimation method to use (and why), and any significant problems you expect to encounter. Suggest a test of the "supply side economics."

3. Suggest how to measure the effect of class size on the school performance of elementary school children. Be precise about what explanatory variables to use, what econometric methods to use (and why), and any problems you expect to incur. There is a negative correlation between class size and performance, and yet most researchers find no significant effect of class size on performance when using regression analysis. Explain why.
You are hired as a consultant to measure various relationships affecting the work environment of mature men. You are given a large data set with answers to questions asked in the enclosed survey. You must write a coherent (e.g., good English, good logic, etc) proposal to answer all the following questions. Your proposal (and therefore your grade) will be evaluated according to the points indicated to the right of each question.

(1) I want to measure the effect of various characteristics on wages.
   a) How can I use the available data to construct a measure of hourly wage and variables that might affect it? (Be precise about what variables I can use and how they can be constructed). (10 points)

   b) Suggest some reasonable specifications for wage equations. (5 points)

   c) Suggest what econometric method to use and why it is appropriate. (5 points)

   d) I am particularly interested in estimating the return to education and how that varies across people. How do I do this? What test statistics do I construct? How do I use the test statistics? (15 points)

   e) I am afraid that one of the variables affecting wage is experience on the job but that it might be endogenous. Why so? How can I deal with its potential endogeneity? (15 points)

(2) I am interested in how people decide whether to work.
   a) How can I use the available data to construct a measure of whether or not you work and variables that might affect it? (5 points)

   b) Suggest a reasonable model specification. (5 points)

   c) How can I test whether marital status and number of children have different effects on the work decision for men and women? Be precise and give details (e.g., the test statistic and its asymptotic distribution). (20 points)

(3) I am interested in the joint determination of wage and hours. In particular, I think wage depends on hours in that part-time people get paid less per hour than full-time people. Also hours (supply) depends upon wage. Also, I am afraid that marital status affects hours and may be affected by wage.

   a) Suggest a reasonable specification of the relationships I am interested in. Which structural parameters are identified? (20 points)
b) How can I estimate each equation? Be precise and give details. Note that marital status is binary. (20 points)

(4) I can probably get you data similar to this data for each year between 1970 and 1990. This period, income distribution has become more disbursed. Suggest how to measure how much of this is due to

a) increases in the return to education; (10 points)

b) more variation in work status; and (10 points)

c) differences between blacks and whites. (10 points)

Be specific. Provide details of equations to estimate, estimation methods, and test statistics.
Part I. Answer 6 out of 8 questions (90 points)

1. Let \( y = X\beta + u \) where \( u \sim N(0,\Omega) \). What is the covariance matrix of the OLS estimator of \( \beta \)?

2. Consider the following three equations:
   
   i. \( y_1 = \alpha_0 + \alpha_1 \text{Female}_i + \alpha_2 \text{Black}_i + Z\alpha_3 + u_1 \)
   
   ii. \( y_1 = \beta_0 + \beta_1 \text{Male}_i + \beta_2 \text{Black}_i + Z\beta_3 + e_1 \)
   
   iii. \( y_1 = \gamma_{11} \text{Male}_i + \gamma_{12} \text{Female}_i + \gamma_2 \text{White}_i + Z\gamma_3 + \epsilon_1 \).

   a. Write \( \beta_0, \beta_1, \beta_2, \) and \( \beta_3 \) in terms of \( \alpha_0, \alpha_1, \alpha_2, \) and \( \alpha_3 \).
   
   b. Write \( \gamma_{11}, \gamma_{12}, \gamma_2, \) and \( \gamma_3 \) in terms of \( \alpha_0, \alpha_1, \alpha_2, \) and \( \alpha_3 \).
   
   c. Write \( \text{Var} \epsilon_1 \) and \( \text{Var} e_1 \) in terms of \( \text{Var} u_1 \).

3. Let \( y = X\beta + Z\gamma + u \). Consider regressing
   
   i. \( y = X\beta + e \)
   
   ii. \( y = Za + \epsilon \).

   Let \( \hat{e} \) and \( \hat{\epsilon} \) be residuals from the two OLS regressions for (i) and (ii). Let \( \hat{u} \) be residuals from regressing \( y \) on \( Z \) and \( X \).

   a. Find \( \hat{e}'\hat{e} \) and \( E\hat{e}'\hat{e} \).
   
   b. Find \( \hat{e}'\hat{u} \) and \( E\hat{e}'\hat{u} \).

4. Let \( y = X\beta + Z\gamma + u \) where \( X \) are \( k \) endogenous regressors and \( Z \) are \( m \) exogenous regressors. Let \( Q \) be a set of instruments for \( X \).

   a. What is the rank of \( Q'[X|Z]? \)
   
   b. How can we get 2SLS estimates of \( \beta \)?

5. Consider \( Y = X\beta + u \). Let \( H_0: \frac{\beta_1}{\beta_2} = 5 \) against \( H_A: \frac{\beta_1}{\beta_2} \neq 5 \).

   a. Construct a Wald test statistic.
   
   b. Construct a t-test statistic.
   
   c. Construct a likelihood ratio test statistic.
6. Let \( y^*_i = X_i \beta + u_i \), \( i = 1, 2, \ldots, n \). Let

\[
    y_1 = \begin{cases} 
    0 & \text{if } y^*_1 \leq 0 \\
    y^*_1 & \text{if } 0 \leq y^*_1 \leq 10 \\
    10 & \text{if } 10 \leq y^*_1 
    \end{cases}
\]

Write down the likelihood function to estimate \( \beta \) given data on \((y_i, X_i)\), \( i = 1, 2, \ldots, n \).

7. Consider \( y^*_i = X_i \beta + u_i \), \( i = 1, 2, \ldots, n \). Let

\[
    y_i = \begin{cases} 
    0 & \text{if } y^*_i \leq 0 \\
    1 & \text{if } y^*_i > 0 
    \end{cases}
\]

Compare i) \( u_i \sim \text{iidN}(0,1) \) with ii) \( u_i \sim \text{iidN}(0, \sigma^2) \) where \( \sigma^2 \) is unknown. Write down the likelihood function for both cases and show that for any pair \((\beta_1, \sigma^2)\) with \( \sigma^2 > 0 \), there exists another pair \((\beta_2, 1)\) with the same value of the likelihood function. Explain intuitively why this implies that \( \sigma^2 \) is not identified.

8. Why does satisfaction of the rank condition for identification imply satisfaction of the order condition? Why doesn't satisfaction of the order condition imply satisfaction of the rank condition?

Part II. Do 1 out of 2 questions (40 points)

1. Consider the model \( y^*_i = X_i \beta + u_i \), \( u_i \sim \text{iidN}(0,1) \), \( i = 1, 2, \ldots, n \). Let

\[
    y_i = \begin{cases} 
    0 & \text{if } y^*_i \leq 0 \\
    1 & \text{if } y^*_i > 0 
    \end{cases}
\]

   a. Show how to write the derivative of the log likelihood function as \( Z'e = 0 \) where \( Z \) is a matrix of functions of explanatory variables and \( e \) is a vector of residuals.

   b. Assume some of the elements of \( X \) are endogenous, i.e., \( EX'u \neq 0 \). Let \( Q \) be a set of instruments. Suggest how to estimate \( \beta \) using \( Q \).
4. Consider the following model:

\[
q^d_t = \alpha_0 + \alpha_1 p_t + \alpha_2 y_t + e_t \\
q^s_t = \beta_0 + \beta_1 p_t + \beta_3 z_t + u_t \\
q^d_t = q^s_t
\]

where \( p_t \) is endogenous and \( y_t \) and \( z_t \) are exogenous. Suggest how to test

\[
H_0: \alpha_2 = \beta_3 = 0 \\
H_A: \alpha_2 \neq 0, \beta_3 \neq 0.
\]
Part III. Do 2 out of 3 questions (50 points)

1. Using the tables and figures from "Lengths of Spells of Psychiatric Hospital Stays and Community Stays," explain in terms of the coefficients in Table 6 the curves in figures labeled "Kaplan-Meier Survival Curves for Hospital Stays" and "Estimated Survival Curves for Hospital Stays." What does econometrics theory imply about the "SLOPE" coefficients in Table 6 relative to the "SLOPE" coefficients in Table 5?

2. Using the tables in "A Disaggregate Discrete Choice Model of Transportation Demand by Elderly and Disabled People in Rural Virginia," explain in detail how to interpret the results in Tables 4 and 5. Explain in detail how to derive the results in Table 5 given the estimates in Table 4.

3. Using Table 4 in "Estimating Family Long-Term Care Decisions in the Presence of Endogenous Child Characterstics," take your family and compute the probability of each care arrangement being used once your parents are 80 years old. Hint: If there are n children in your family, there are n + 2 care alternatives. Provide detail on what you are assuming about your family and on each calculation you do.