Write your name, Swedish personal number and the number of the question on every cover sheet. Do not write answers for more than one question in the same cover sheet. Explain notions/concepts and symbols. Only legible exams will be marked. No aids are allowed with the exception of calculators provided by exam administrators.

The exam consists of two parts. Part 1 consists of 20 multiple choice questions worth 40 points in total (2 points each). Part 2 consists of three discussion questions worth 60 points in total (20 points each). 100 points in total. For the grade E 40 points are required, for D 50 points, C 60 points, B 75 points and A 90 points.

If you think that a question is vaguely formulated: specify the conditions used for solving it.

Results will be posted on the notice board, House A, floor 3, February 5, 2009 at the latest

Good luck!
Part 1: Multiple Choice Questions (40 points)

Circle the right answer. Only one answer per question. No credit will be given for multiple answers or additional explanations. Two points per question for correct answers.

1. In a multiple regression framework, the slope coefficient on the regressor $X_{2i}$
   a. takes into account the scale of the error term.
   b. is measured in the units of $Y_i$ divided by units of $X_{2i}$.
   c. is usually positive.
   d. is larger than the coefficient on $X_{1i}$.

2. Under imperfect multicollinearity
   a. the OLS estimator cannot be computed.
   b. two or more of the regressors are highly correlated.
   c. the OLS estimator is biased even in samples of $n > 100$.
   d. the error terms are highly, but not perfectly, correlated.

3. When there are omitted variables in the regression, which are determinants of the dependent variable, then
   a. you cannot measure the effect of the omitted variable, but the estimator of your included variable(s) is (are) unaffected.
   b. this has no effect on the estimator of your included variable because the other variable is not included.
   c. this will always bias the OLS estimator of the included variable.
   d. the OLS estimator is biased if the omitted variable is correlated with the included variable.

4. Imagine you regressed earnings of individuals on a constant, a binary variable (“Male”) which takes on the value 1 for males and is 0 otherwise, and another binary variable (“Female”) which takes on the value 1 for females and is 0 otherwise. Because females typically earn less than males, you would expect
   a. the coefficient for Male to have a positive sign, and for Female a negative sign.
   b. both coefficients to be the same distance from the constant, one above and the other below.
   c. none of the OLS estimators to exist because there is perfect multicollinearity.
   d. this to yield a difference in means statistic.
5. For a single restriction \((q = 1)\), the \(F\)-statistic

   a. is the square root of the \(t\)-statistic.
   b. has a critical value of 1.96.
   c. will be negative.
   d. is the square of the \(t\)-statistic.

6. If you reject a joint null hypothesis using the \(F\)-test in a multiple hypothesis setting, then

   a. a series of \(t\)-tests may or may not give you the same conclusion.
   b. the regression is always significant.
   c. all of the hypotheses are always simultaneously rejected.
   d. the \(F\)-statistic must be negative.

7. All of the following are true, with the exception of one condition:

   a. a high \(R^2\) or \(\bar{R}^2\) does not mean that the regressors are a true cause of the dependent variable.
   b. a high \(R^2\) or \(\bar{R}^2\) does not mean that there is no omitted variable bias.
   c. a high \(R^2\) or \(\bar{R}^2\) always means that an added variable is statistically significant.
   d. a high \(R^2\) or \(\bar{R}^2\) does not necessarily mean that you have the most appropriate set of regressors.

8. The general answer to the question of choosing the scale of the variables is

   a. dependent on you whim.
   b. to make the regression results easy to read and to interpret.
   c. to ensure that the regression coefficients always lie between -1 and 1.
   d. irrelevant because regardless of the scale of the variable, the regression coefficient is unaffected.

9. If the estimates of the coefficients of interest change substantially across specifications,

   a. then this can be expected from sample variation.
   b. then you should change the scale of the variables to make the changes appear to be smaller.
   c. then this often provides evidence that the original specification had omitted variable bias.
   d. then choose the specification for which your coefficient of interest is most significant.
10. In the case of regression with interactions, the coefficient of a binary variable should be interpreted as follows:

a. there are really problems in interpreting these, since the \( \ln(0) \) is not defined.
b. for the case of interacted regressors, the binary variable coefficient represents the various intercepts for the case when the binary variable equals one.
c. first set all explanatory variables to one, with the exception of the binary variables. Then allow for each of the binary variables to take on the value of one sequentially. The resulting predicted value indicates the effect of the binary variable.
d. first compute the expected values of \( Y \) for each possible case described by the set of binary variables. Next compare these expected values. Each coefficient can then be expressed either as an expected value or as the difference between two or more expected values.

11. In the regression model \( Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 (X_i \times D_i) + u_i \), where \( X \) is a continuous variable and \( D \) is a binary variable, \( \beta_3 \)

a. indicates the slope of the regression when \( D=1 \).
b. has a standard error that is not normally distributed even in large samples since \( D \) is not a normally distributed variable.
c. indicates the difference in the slopes of the two regressions.
d. has no meaning since \( (X_i \times D_i) = 0 \) when \( D_i = 0 \).

12. Errors-in-variables bias

a. is present when the probability limit of the OLS estimator is given by
   \[
   \hat{\beta}_1 \to \beta_1 + \frac{\sigma^2}{\sigma^2 + \sigma^2_y}.
   \]
b. arises when an independent variable is measured imprecisely.
c. arises when the dependent variable is measured imprecisely.
d. always occurs in economics since economic data is never precisely measured.

13. The reliability of a study using multiple regression analysis depends on all of the following with the exception of

a. omitted variable bias.
b. errors-in-variables.
c. presence of homoskedasticity in the error term.
d. external validity.
14. Panel data estimation can sometimes be used
   a. to avoid the problems associated with misspecified functional forms.
   b. in case the sum of residuals is not zero.
   c. in the case of omitted variable bias when data on the omitted variable is not available.
   d. to counter sample selection bias.

15. The “before and after” specification, binary variable specification, and “entity-demeaned” specification produce identical OLS estimates
   a. as long as there are observations for more than two time periods.
   b. if you use the heteroskedasticity-robust option in your regression program.
   c. for the case of more than 100 observations.
   d. as long as $T = 2$ and the intercept is excluded from the “before and after” specification.

16. Consider the regression example from your textbook, which estimates the effect of beer taxes on fatality rates across the 48 contiguous U.S. states. If beer taxes were set nationally by the federal government rather than by the states, then
   a. it would not make sense to use state fixed effect.
   b. you can test state fixed effects using homoskedastic-only standard errors.
   c. the OLS estimator will be biased.
   d. you should not use time fixed effects since beer taxes are the same at a point in time across states.
17. In the linear probability model, the interpretation of the slope coefficient is
   a. the change in odds associated with a unit change in $X$, holding other
      regressors constant.
   b. not all that meaningful since the dependent variable is either 0 or 1.
   c. the change in probability that $Y=1$ associated with a unit change in $X$,
      holding others regressors constant.
   d. the response in the dependent variable to a percentage change in the
      regressor.

18. Estimation of the IV regression model
   a. requires exact identification.
   b. allows only one endogenous regressor, which is typically correlated with
      the error term.
   c. requires exact identification or overidentification.
   d. is only possible if the number of instruments is the same as the number of
      regressors.

19. Weak instruments are a problem because
   a. the TSLS estimator may not be normally distributed, even in large
      samples.
   b. they result in the instruments not being exogenous.
   c. the TSLS estimator cannot be computed.
   d. you cannot predict the endogenous variables any longer in the first stage.

20. The following are reasons for studying randomized controlled experiment in an
    econometrics course, with the exception of
    a. at a conceptual level, the notion of an ideal randomized controlled
       experiment provides a benchmark against which to judge estimates of
       causal effects in practice.
    b. when experiments are actually conducted, their results can be very
       influential, so it is important to understand the limitations and threats to
       validity of actual experiments as well as their strength.
    c. randomized controlled experiments in economics are common.
    d. external circumstances sometimes produce what appears to be
       randomization.
Part 2: Discussion Questions

On separate sheets of paper, answer the following discussion questions. Write your name, personal number (personnummer) and the question number on each sheet. Answer each question clearly and concisely. Only legible answers will be considered, others will be disregarded. If you think that a question is vaguely formulated, specify the conditions used for solving it. Each question is worth 20 points.

1) Two authors published a study in 1992 of the effect of minimum wages on teenage employment using a U.S. state panel. The paper used annual observations for the years 1977-1989 and included all 50 states plus the District of Columbia. The estimated equation is of the following type

\[ E_{it} = \beta_0 + \beta_1 \left( \frac{M_{it}}{W_{it}} \right) + \gamma_2 D_{2i} + \ldots + \gamma_n D_{ni} + \delta_2 B_{2i} + \ldots + \delta_{13} B_{13i} + u_{it}, \]

where \( E \) is the employment to population ratio of teenagers, \( M \) is the nominal minimum wage, and \( W \) is average wage in the state. In addition, other explanatory variables, such as the prime-age male unemployment rate, and the teenage population share were included.

(a) Briefly discuss the advantage of using panel data in this situation rather than pure cross sections or time series.

(b) Estimating the model by OLS but including only time fixed effects results in the following output

\[ \hat{E}_{it} = \hat{\beta}_0 - 0.33 \times \left( \frac{M_{it}}{W_{it}} \right) + 0.35 \times (SHY_{it}) - 1.53 \times uram_{it}; \quad \hat{R}^2 = 0.20 \]

\[ (0.08) \quad (0.28) \quad (0.13) \]

where \( SHY \) is the proportion of teenagers in the population, and \( uram \) is the prime-age male unemployment rate. Coefficients for the time fixed effects are not reported. Numbers in parenthesis are homoskedasticity-only standard errors.

Comment on the above results. Are the coefficients statistically significant? Since these are level regressions, how would you calculate elasticities?

(c) Adding state fixed effects changed the above equation as follows:

\[ \hat{E}_{it} = \hat{\beta}_0 + 0.07 \times \left( \frac{M_{it}}{W_{it}} \right) - 0.19 \times (SHY_{it}) - 0.54 \times uram_{it}; \quad \hat{R}^2 = 0.69 \]

\[ (0.10) \quad (0.22) \quad (0.11) \]

Compare the two results. Why would the inclusion of state fixed effects change the coefficients in this way?

(d) The significance of each coefficient decreased, yet \( \hat{R}^2 \) increased. How is that possible? What does this result tell you about testing the hypothesis that all of the state fixed effects can be restricted to have the same coefficient? How would you test for such a hypothesis?
2) Earnings functions, whereby the log of earnings is regressed on years of education, years of on the job training, and individual characteristics, have been studied for a variety of reasons. Some studies have focused on the returns to education, others on discrimination, union non-union differentials, etc. For all these studies, a major concern has been the fact that ability should enter as a determinant of earnings, but that it is close to impossible to measure and therefore represents an omitted variable.

Assume that the coefficient on years of education is the parameter of interest. Given that education is positively correlated to ability, since, for example, more able students attract scholarships and hence receive more years of education, the OLS estimator for the returns to education could be upward biased. To overcome this problem, various authors have used instrumental variable estimation techniques. For each of the instruments potential instruments listed below briefly discuss instrument validity.

(a) The individual’s postal zip code.
(b) The individual’s IQ or testscore on a work related exam.
(c) Years of education for the individual’s mother or father.
(d) Number of siblings the individual has.

3) Describe the major differences between a randomized controlled experiment and a quasi-experiment.