Introduction to Econometrics Professor Samantha Leorato

SURNAME	NAME	STUDENT'S N.
______________________	______________________	______________________

Exam rules
You have **1 hour and 45 minutes** *to complete the exam*.

Students are not allowed to use notes or books.

Simulated exam (12.12.2018)

1. Which of the following hypotheses can't be tested using a classical F test?
   (a) $\beta_1 = \beta_2 = 0$
   (b) $\beta_0 = \beta_1^2 + \beta_2^2$.
   (c) $\beta_0 = 1$.
   (d) $\beta_0 = -\beta_1$.

---

2. If the regressors of a multivariate linear model suffer the perfect collinearity:
   (a) One or more parameters are not identified.
   (b) The OLS estimator is inconsistent.
   (c) The OLS estimator is inefficient.
   (d) The TSLS estimator should be used.

---

3. The TSLS estimator can be obtained by the following two steps: (please thick two answers)
   (a) In the first stage y is regressed on the instruments $\mathbf{z} = (z_1, \ldots, z_m)$
   (b) In the first stage each of the endogenous variables x_j is regressed on the instruments $\mathbf{z} = (z_1, \ldots, z_m)$ and on the exogenous variables included
   (c) In the second stage y is regressed on the residuals from the first stage.
   (d) In the second stage y is regressed on the predicted values from the first stage.

---

4. Given a sample of $n = 120$ units we observe:
   Sia $\bar{Y} = 0.41667$, $\sum_i X_i = -20$, $n^{-1} \sum_i (Y_i - \bar{Y})^2 = 0.868$; , $n^{-1} \sum_i (X_i - \bar{X})^2 = 2.639$, $\sum_i X_i Y_i = 150$
   (i) consider the following simple linear model, assuming $EX_iU_i = 0$ and $U_i \sim N(0, \sigma^2)$

   $$Y_i = \beta_0 + \beta_1 X_i + U_i, \quad i = 1, \ldots, n$$

   Estimate $\hat{\beta}_0$ and $\hat{\beta}_1$.
   (ii) Compute ESS, TSS e RSS.
   (iii) Compute the unbiased estimate of the variance of the errors $\hat{\sigma}^2$ (SER^2) and the regression l'R^2.
   (iv) Compute an estimate of the variance of $\hat{\beta}_1$ (under the assumption of homoskedasticity of the errors) and test the hypothesis $\beta_1 = 0.7$ at the significance level $\alpha = 0.05$.
   (v) Compute the predicted value $\mathrm{E}(Y \mid X = -0.5)$.

5. Suppose we want to regress the number of cigarettes smoked by a group of men (`Cig`) on the number of cigarettes smoked by their fathers (`Cig_F`) (a non-smoking person has values `Cig` $= 0$, a non smoking father has value `Cig_F` $= 0$)
   (i) It is possible that `Cig_F` is endogenous. If this claim is true, what would be the consequences on the OLS estimates of the model?
   (ii) After introducing some control variables, (age of the son, education, father_job, wage, ...), we decide to solve the problem of an eventual residual endogeneity by using TSLS estimation, using as instrument information on the grandfathers (their education and occupation before retirement). Define briefly the conditions for a variable to be a valid instrument. Do you think these two variables fulfil the necessary conditions?
   (iii) We compute the J test (Sargan test) for the two instruments and we get $J = 5.253$. What does it mean?

6. We observe a balanced panel dataset with $n = 150$ individuals observed in two consecutive years (1999,2000). The variables observed include: education, wage, experience, marital status, number of children,....
   (i) Assume that none of the workers in the sample has a child in the period (1999–2000). What problem we would have if we wanted to estimate the effect of the number of kids on wage using the first difference estimator?
   (ii) By adding to the model the interaction between `educ` (years of schooling) and the dummy for `year=2000` we obtain the following estimate for the parameter associated to the interaction [`dummy2000` $*$ `educ`]: .030 (the interaction [`dummy1999` $*$ `educ`] is the benchmark). How would you interpret this value knowing that the dependent variable is `log_wage`?

Answers

1. (b)
2. (a)
3. (b)-(d)
4. (i) $\hat{\beta}_1 = 0.5$; $\hat{\beta}_0 = 0.5$

   (ii)
$$TSS = n * 0.86806 = 104.1667$$
$$ESS = n * \hat{\beta}_1^2 * n^{-1} \sum_i (X_i - \bar{X})^2 = 120 * 2.639 * (0.5)^2 = 79.1667$$
$$RSS = TSS - ESS = 25$$

   (iii) $\hat{\sigma}^2 = RSS/(n-2) = 25/118 = 0.21186$. The R^2 is $R^2 = ESS/TSS = 0.76$.

   (iv) $\widehat{\text{var}(\hat{\beta}_1)} = \hat{\sigma}^2 / \sum_i (X_i - \bar{X})^2 = 0.21186/(120*2.639) = 0.00067$. Thus, $\hat{se}(\hat{\beta}_1) = \sqrt{0.00067} = 0.02587$ and the $t-$statistic to test $\beta_1 = 0.7$ is equal to $t = \frac{0.5-0.7}{0.02587} = -7.73$. If $H_1 : \beta_1 \neq 0.7$, then we reject the null hypothesis at $\alpha = 0.05$, because $|t| > 1.96$ (critical value from the normal distribution).

   (iv) $\hat{\text{E}}(Y \mid X = -0.5) = 0.5 - 0.5^2 = 0.25$.

5. (i) If the claim is true, the OLS estimator for the parameter of `Cig_F` would be biased and inconsistent

   (ii) A variable z is a valid instrument if it satisfies simoultaneously the two conditions: (1) It is exogenous, namely $Cov(z_i, u_i) = 0$; (2) It is relevant, that, in the case of a single endogenous regressor (x_i) means that it is correlated with it, namely $Cov(x_i, z_i) \neq 0$. In the case of two instruments (z_1, z_2) and 1 endogenous regressor (without other exogenous regressors, for simplicity), both instruments have to be uncorrelated with the error, and the exclusion restriction on the coefficients associated to the instruments in the reduced form model $x_i = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + v$ have to be rejected, (that is, the null $H_0 : \pi_1 = \pi_2 = 0$ has to be rejected). The two variables `Granfather_educ` and `Grandfather_job` can be correlated with the regressor `Cig_F`, because the household social and cultural background has an impact on the smoking habits in adulthood. This can suggest that the instruments are likely to be relevant. Each of the instruments is also exogenous if we assume that grandparents didn't have a role in early years nurture of their grandchildren, and had an impact on their smoking behavior only through their influence on their fathers; whether this assumption is plausible or not depends on the characteristic of the sample (age of the individuals, year of the survey, social characteristics of their household, and nationality of individuals)

   (iii) Sargan's J test is a test to verify the overidentification restriction (a test for exogeneity of the instruments, in case of overidentification). In this case, the endogenous variable is CF and we have 2 instruments. Under the null hypothesis that both instruments are exogenous, the test follows, asymptotically a χ^2_1 distribution. The Sargan J is larger than the critical value of a χ^2_1 at level $\alpha = 0.05$ (that is 3.84), then we reject the null hypothesis and conclude that there is a potential problem of endogeneity in (at least) one of the two variables used as instruments (although we wouldn't reject at 1% level).

6. (i) After the first difference, all constants and variables that are constant in time for each unit would disappear, thus we can't estimate the parameter associated to the number of children because the variable is eliminated together with the individual effects.

   (ii) The estimate 0.03 gives the effect that one more year of schooling in 2000 has on wage, relative to the benchmark year (1999). Since the wage is in logarithmic scale, this means that one more year of schooling in the year 2000 determines, on average, an effect on wage that is 3% higher than one more year of schooling in 1999.