

ECON742 Fall 2010
Final exam December 14th 2010

The exam lasts for 3 hours. Manage your time according to the percentages of each question. No document is allowed. Good luck!

1. (16%) Consider the following linear model:

$$y_i = \beta x_i + \epsilon_i$$

$i \in \{1, \dots, N\}$, $\epsilon_i \text{ iid}(0, \sigma^2)$, $x_i \text{ iid}(0, \sigma_x^2)$, $p \lim(\frac{1}{N} \sum x_i \epsilon_i \neq 0)$.

- (a) Describe the instrumental variables estimator (with one instrument) in this case, and show it is consistent.
- (b) Suppose we find an appropriate instrumental variable w_i that is a dichotomous variable 1/0 independent from ϵ_i and correlated with x_i . Suppose we divide the sample in a first sub-sample $i \in \{1, \dots, N_1\}$ in which $w_i = 1$, and in a second one $i \in \{N_1 + 1, \dots, N\}$ in which $w_i = 0$. Show that the instrumental variable estimator may be rewritten as:

$$\widehat{\beta} = \frac{\frac{1}{N_1} \sum_{j=1}^{N_1} y_j - \frac{1}{N-N_1} \sum_{j=N_1+1}^N y_j}{\frac{1}{N_1} \sum_{j=1}^{N_1} x_j - \frac{1}{N-N_1} \sum_{j=N_1+1}^N x_j}$$

- (c) Show that:

$$\widehat{\beta} = \beta + \frac{\frac{1}{N_1} \sum_{j=1}^{N_1} \epsilon_j - \frac{1}{N-N_1} \sum_{j=N_1+1}^N \epsilon_j}{\frac{1}{N_1} \sum_{j=1}^{N_1} x_j - \frac{1}{N-N_1} \sum_{j=N_1+1}^N x_j}$$

Use this formula to demonstrate that $\widehat{\beta}$ is a consistent estimator of β .

- (d) Make a link between this method and the regression discontinuity design estimate in a subsample restricted around the cut-off.
2. (26%) Consider the following linear model:

$$y = X_1 \beta_1 + X_2 \beta_2 + \epsilon$$

with N observations and $k_1 + k_2$ explanatory variables.

- (a) An econometrician does not realize that X_2 plays a role in this regression, and omits it. Show that the OLS estimator of β_1 is biased if one omits X_2 .
- (b) The “partitioning model” gives us the following formula: $\widehat{\beta}_1 = (X_1' M_{X_2} X_1)^{-1} X_1' M_{X_2} y$, with $M_{X_2} = I_N - X_2 (X_2' X_2)^{-1} X_2'$. Show that this formula gives us a consistent estimator of β_1 (justify well each step of the calculation, with the names of the appropriate theorems, and the assumptions you must make to get to the final result).
- (c) Calculate the variance-covariance matrix of $\widehat{\beta}_1$ (call σ_ϵ^2 the variance of the disturbance term in assumption A4).

- (d) An econometrician suggests to regress $y^+ = M_{X_2}y$ on $X_1^+ = M_{X_2}X_1$, to get $\widehat{\beta}_1^+$, in the model $y^+ = X_1^+\widehat{\beta}_1^+ + \widehat{u}$.
1. What is the relationship between $\widehat{\beta}_1$ and $\widehat{\beta}_1^+$?
 2. What is the relationship between $\widehat{\epsilon}$ and \widehat{u} ?
- (e) An econometrician suggests to regress y on $X_1^+ = M_{X_2}X_1$, to get $\widehat{\beta}_1^{++}$, in the model $y = X_1^+\widehat{\beta}_1^{++} + \widehat{v}$.
1. What is the relationship between $\widehat{\beta}_1$ and $\widehat{\beta}_1^{++}$?
 2. What is the relationship between $\widehat{\epsilon}$ and \widehat{v} ?
- (f) Practical application: consider the model:

$$w_s = \beta_0 + \beta_1x_s + \beta_2y_s + \beta_3z_s + \epsilon_s$$

With $E(\epsilon_s|x_s, y_s, z_s) = 0$. To estimate β_3 , an econometrician suggests to use a 2-step strategy:

1. regress z_s on identity, x_s , et y_s . Get the residuals q_s from this regression.
2. regress w_s on q_s , and get the coefficient.

Is this coefficient a consistent estimator of β_3 ? May we look at the R^2 , and the t-test associated with this coefficient?

- (g) How would your answers be affected if X_1 and X_2 were not correlated? What is the implication for randomized experiments? Should we, or should we not, include the X_2 ?
3. (30%) This exercise refers to the paper, not being studied in class, of Rafael Di Tella, Sebastian Galiani and Ernesto Schargrotsky, (2007), "The Formation of Beliefs: Evidence from the Allocation of Land Titles to Squatters," The Quarterly Journal of Economics February 2007, Vol. 122, No. 1. The purpose of this paper is to study the impact of property rights on the formation of market beliefs (individualism, materialism, meritocracy, trust). The study is located in Argentina and studied squatters who obtained or not, ownership of the land where they lived. Twenty years later, data on these four market beliefs were collected.
- (a) Describe quickly how access to property rights could affect these four market beliefs?
 - (b) Could we do a simple OLS regression of these four beliefs on a dummy variable equal to 1 if the individual has a property right, and 0 otherwise? Describe any problems that may arise.
 - (c) To address this identification problem, the authors use a "natural experiment" that occurred in Argentina in 1981. Read this description of the natural experiment: "We address this endogeneity problem by exploiting a natural experiment. In 1981, about 1,800 families occupied a wasteland area in San Francisco Solano, in the metropolitan area of Buenos Aires, Argentina. The occupants were groups of

landless citizens (organized by a Catholic priest) who wanted to avoid creating a shantytown and immediately partitioned the occupied land into small urban-shaped parcels. At the time of the occupation, the squatters thought the land belonged to the State. However, the occupied area turned out to be made up of thirteen tracts of land belonging to different private owners. The squatters resisted several eviction attempts during the military government. After Argentina's return to democracy, the Congress of the Province of Buenos Aires in 1984 passed a law expropriating these lands from the former owners in exchange for a monetary compensation, in order to transfer the parcels with legal titles to the squatters.

The resulting titling process, however, was incomplete and asynchronous. In 1986 the government made a compensation offer to each owner (or group of owners as some tracts had more than one owner) calculated in proportion to the official tax valuation of each tract of land. These official valuations had been set before the land occupation by the fiscal authority with the purpose of calculating property taxes and were indexed by inflation. After the government made the expropriation offers, the owner(s) of each tract had to decide whether to accept the expropriation compensation and surrender the land or to start a legal dispute in order to obtain a higher compensation. The owners of eight tracts of land accepted the compensation offered by the government. In 1989, the squatters living on those tracts were offered formal titles that secured the property of the parcels. Five former owners, instead, did not accept the compensation offered by the government and sued with the aim of obtaining a higher compensation (the law was approved by Congress and, thus, the expropriation itself could not be challenged, whereas the monetary offers made by the government could be disputed). One of these five lawsuits ultimately ended with a final verdict, and squatters who had settled on this tract of land received titles in 1998. At the time of writing, the other four legal processes are still outstanding in the slow Argentine courts."

Summarize the natural experiment: what is the treatment group? What is the control group? How many tracts of land are in each group?

The number of interviewed households is described in Table 1. How do you make sense of the number 80 (property right=0, but property right availability=1).

TABLE I
ALLOCATION OF LAND TITLES

Year	Property Right Availability = 1			Property Right Availability = 0	
	Property Right = 1	Property Right = 0	Total	Property Right = 0	Total
1989	419	23	442		
1998	173	57	230		
Total	592	80	672	410	1082

Notes: *Property Right Availability* equals 1 if land titles were available for the parcel, i.e., if the former owner surrendered the land to the State. *Property Right* equals 1 if the household has formal titles to the parcel.

- (d) Criticize this natural experiment: is it really "natural"? Is getting a property right, that is to say being in the treatment group, truly exogenous? Give two stories that could cause the treatment is not exogenous.
- (e) To convince us that this natural experiment is good, the authors show us the following Tables II and III. Why are these two tables useful? How do they reassure us that the experiment is natural?

TABLE II
PRETREATMENT CHARACTERISTICS OF THE ORIGINAL HOUSEHOLD HEAD

Characteristics of the original household head	Property Right Availability = 0	Property Right Availability = 1	Difference
Age	48.875 (0.938)	50.406 (0.761)	-1.532 (1.208)
Female	0.407 (0.046)	0.353 (0.035)	0.054 (0.058)
Argentine	0.903 (0.028)	0.904 (0.022)	-0.001 (0.035)
Years of education	6.071 (0.188)	5.995 (0.141)	0.076 (0.235)
Argentine father	0.795 (0.038)	0.866 (0.025)	-0.072 (0.046)
Years of education of the father	4.655 (0.147)	4.417 (0.076)	0.237 (0.165)
Argentine mother	0.804 (0.038)	0.856 (0.026)	-0.052 (0.046)
Years of education of the mother	4.509 (0.122)	4.548 (0.085)	-0.039 (0.149)

Notes: We define the original household head as the family member who was the household head at the time the family arrived to the parcel they are currently occupying. *Property Right Availability* equals 1 if land titles were available for the parcel. The Appendix presents the definition of all the variables. Standard errors are in parentheses.

TABLE III
 PRETREATMENT PARCEL CHARACTERISTICS

Parcel characteristics	Property Right Availability = 0	Property Right Availability = 1	Difference
Distance to creek (in blocks)	1.995 (0.061)	1.906 (0.034)	0.088 (0.070)
Distance to nonsquatted area (in blocks)	1.731 (0.058)	1.767 (0.033)	-0.036 (0.067)
Parcel surface (in m ²)	287.219 (4.855)	277.662 (2.799)	9.556* (5.605)
Block corner	0.190 (0.019)	0.156 (0.014)	0.033 (0.023)

Notes: *Property Right Availability* equals 1 if land titles were available for the parcel. The Appendix presents the definition of all the variables. Standard errors are in parentheses.

*Significant at 10 percent.

- (f) Are you reassured by these tables? Are there some problems remaining?
- (g) The main results are presented in Table VI. The four questions are:
- “The first was “Do you believe that it is possible to be successful on your own or a large group that supports each other is necessary?” The two possible answers were “It is possible to be successful on your own” and “A large group is necessary to be successful.” The second was “Do you believe that having money is important to be happy?” The possible answers were “Indispensable to be happy,” “Very important to be happy,” “Important to be happy,” and “Not important to be happy.” The third was “In general, people who put effort working end up much better, better, worst, or much worst than those that do not put an effort?” The possible answers were “Much better than those that do not put an effort,” “Better than those that do not put an effort,” “Worst than those that do not put an effort,” and “Much worst than those that do not put an effort.” The fourth and final question was “In general, in our country, would you say that one can trust other people or that people cannot be trusted?” The possible answers were “You can trust others” and “You cannot trust others.””
- “Market Beliefs is the sum of the dummies for the four questions.”

TABLE IV
BELIEFS AND PROPERTY RIGHTS IN THE SOLANO SETTLEMENT

	(1a)	(2a)	(3a)	(4a)	(5a)
	Success- Alone	Money- Important	Effort- Better	Trust- Others	Market Beliefs
Property right	0.144** (0.064)	0.202*** (0.063)	0.072 (0.056)	0.108* (0.063)	0.527*** (0.131)
Controls	No	No	No	No	No
Observations	312	312	313	313	312
	(1b)	(2b)	(3b)	(4b)	(5b)
Property right	0.169** (0.066)	0.188*** (0.068)	0.022 (0.056)	0.139** (0.065)	0.520*** (0.133)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	312	312	313	313	312

Notes: All columns present 2SLS regressions where *Property Right* is instrumented with *Property Right Availability*. Regressions in the b panel are similar but control for parcel and household characteristics. The former include surface of the parcel, distance to creek, distance to nearest nonsquatted area, and a corner dummy. The latter include age, gender, nationality and years of education of the original household head, and nationality and years of education of her/his parents. The Appendix presents the definition of the variables. Values in parenthesis represent standard errors.

* Significant at 10 percent level;
 ** Significant at 5 percent level;
 *** Significant at 1 percent level.

Comment on these results, forgetting for the moment that they are 2SLS regressions (just pretend they are OLS regressions).

- (h) It is not easy to interpret the magnitude (or economic importance) of the coefficients. Can you use the Table VI below to compare the magnitude of the coefficient "Property Rights" with the magnitude of the coefficient "HH Education" in column 5?

TABLE VI
DIRECT VS. INDIRECT EFFECTS (INCOME, WEALTH, AND EDUCATION)

	(1)	(2)	(3)	(4)	(5)
	Success-Alone	Money-Important	Effort-Better	Trust-Others	Market Beliefs
Property Right	0.214*** (0.075)	0.144* (0.076)	0.021 (0.064)	0.136* (0.074)	0.516*** (0.145)
Income	0.0002 (0.0005)	-0.0003 (0.0005)	0.0004 (0.0004)	-0.0001 (0.0005)	0.0001 (0.001)
Wealth	-0.0001 (0.001)	0.001 (0.001)	-0.001 (0.001)	-0.0001 (0.001)	-0.001 (0.002)
HH Education	0.046*** (0.016)	0.034** (0.016)	-0.003 (0.013)	0.011 (0.016)	0.089*** (0.031)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	254	254	254	254	254

Notes: All columns present 2SLS regressions where *Property Right* is instrumented with *Property Right Availability*. Parcel controls include surface, distance to creek, distance to nearest non-squatted area, and a corner dummy. Household controls include age, gender, and nationality of the original household head, and nationality and years of education of her/his parents. All the variables are described in the Appendix. Values in parenthesis represent standard errors.

* Significant at 10 percent level;
 ** Significant at 5 percent level;
 *** Significant at 1 percent level.

- (i) Let us come back to 2SLS. Remember the 80 households that we talked about above. What are the implications for the variable "Property Right" (endogenous, exogenous?)? As you can see in the footnotes, "Property Right is Instrumented With Property Right Availability". Why would it be a good instrument? What are the statistical tests to be done? Are they presented in the paper? What do you think the potential R-squared of the 1st stage?
- (j) Look at Table VIII: the average on the four questions were also collected for the general population of Buenos Aires. By looking at this table, can you tell if it the lucky squatters becoming more pro-market, or if it is the unlucky squatters experiencing frustration and uncertainty due to this situation.

TABLE VIII
BELIEFS AMONGST SOLANO SQUATTERS AND THE GENERAL POPULATION

	(1) Success-Alone	(2) Money-Important	(3) Effort-Better	(4) Trust-Others	(5) Market Beliefs
Average for:					
Buenos Aires General Population	0.440 (0.021)	0.671 (0.019)	0.726 (0.019)	0.476 (0.021)	2.342 (0.046)
Squatters with Property Right = 0	0.330 (0.040)	0.503 (0.042)	0.735 (0.037)	0.335 (0.040)	1.906 (0.086)
Squatters with Property Right = 1	0.433 (0.037)	0.676 (0.035)	0.791 (0.030)	0.393 (0.037)	2.294 (0.074)

Notes: *Property Right* equals 1 if the household has formal titles to the parcel. The Appendix describes the definition of the variables. Values in parenthesis represent standard errors.

- (k) Conclude on the contribution of the paper. Does this paper teach us anything? What? Are there any political implications? Does the empirical methodology seem sound? What would you do with unlimited time and resources? What other hypotheses would you want to test?
4. (20%) Angrist, J. and A. Krueger (1991), "Does Compulsory School Attendance Affect Schooling and Earnings?", *The Quarterly Journal of Economics*, 106(4), 979-1014.
- Explain, using an appropriate specification, how compulsory schooling laws can be used to estimate the impact of education on earnings. What assumptions must be made for this analysis to work? Do you think that quarter of birth may have an impact on income, other than through the number of years of schooling. How would this bias the estimators? Describe all the tests that should be done to test this instrumental variable. Why is each test critical? Are these tests performed in this paper? Criticize this paper.
5. (18%) Duflo, E. (2001), "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment", *American Economic Review*, 91(4), 795-813.

Write down the basic equation of a difference-in-differences estimator (with the notation that you may choose and define appropriately). What assumption(s) are needed to isolate the causal impact of a program, reform, politics? Look at Table 3 below. Describe the identification strategy of the program in panel A. Would it be rigorous

to compare the numbers one by one ("simple difference")? Why should we estimate a "double difference"? Is the assumption underlying the estimation checked (see Panel B)? Criticize this approach. What would you do if you had no budget or time constraints?

TABLE 3—MEANS OF EDUCATION AND LOG(WAGE) BY COHORT AND LEVEL OF PROGRAM CELLS

	Years of education			Log(wages)		
	Level of program in region of birth			Level of program in region of birth		
	High (1)	Low (2)	Difference (3)	High (4)	Low (5)	Difference (6)
<i>Panel A: Experiment of Interest</i>						
Aged 2 to 6 in 1974	8.49 (0.043)	9.76 (0.037)	-1.27 (0.057)	6.61 (0.0078)	6.73 (0.0064)	-0.12 (0.010)
Aged 12 to 17 in 1974	8.02 (0.053)	9.40 (0.042)	-1.39 (0.067)	6.87 (0.0085)	7.02 (0.0069)	-0.15 (0.011)
Difference	0.47 (0.070)	0.36 (0.038)	0.12 (0.089)	-0.26 (0.011)	-0.29 (0.0096)	0.026 (0.015)
<i>Panel B: Control Experiment</i>						
Aged 12 to 17 in 1974	8.02 (0.053)	9.40 (0.042)	-1.39 (0.067)	6.87 (0.0085)	7.02 (0.0069)	-0.15 (0.011)
Aged 18 to 24 in 1974	7.70 (0.059)	9.12 (0.044)	-1.42 (0.072)	6.92 (0.0097)	7.08 (0.0076)	-0.16 (0.012)
Difference	0.32 (0.080)	0.28 (0.061)	0.034 (0.098)	0.056 (0.013)	0.063 (0.010)	0.0070 (0.016)

Notes: The sample is made of the individuals who earn a wage. Standard errors are in parentheses.