Lecture 4

Instrumental Variables

Policy Evaluation

Nova SBE - Universidade Nova de Lisboa / IGC
Pedro C. Vicente
http://www.pedrovicente.org/

Introduction

• We now turn to a method that relaxes the exogeneity assumption of OLS and PSM, and that is also robust to time-varying selection bias, unlike DD

• Recall the simple model:

$$Y_i = aX_i + bT_i + e_i$$

• Treatment assignment *T* may not be random because of two broad factors: endogeneity (programs are placed deliberately in areas that have specific unobservable characteristics correlated with *Y*); unobserved individual heterogeneity (stemming from individual beneficiaries' self-selection into the program)

• In general we then may have:

$$\text{cov}(T, e) \neq 0$$

• IV aims to clean up the correlation between *T* and *ε*, i.e., the variation in *T* that is uncorrelated with *ε* needs to be isolated

• To do so, one needs to find an IV, denoted *Z* that satisfies the following two conditions:

$$1. \quad \text{cov}(T, Z) \neq 0$$

$$2. \quad \text{cov}(\varepsilon, Z) = 0$$

• Thus, instrument *Z* affects selection into the program but is not correlated with factors affecting the outcomes (exclusion restriction)

2SLS: Two-stage Least Squares

• To isolate the part of the treatment variable that is independent of unobserved characteristics affecting the outcome, one first regresses the treatment on the instrument and the other covariates in the original regression; this is the 1st stage regression:

$$T_i = gZ_i + fX_i + u_i$$

• The predicted treatment from this regression $\hat{T}$ reflects the part of the treatment affected only by Z and thus embodies only exogenous variation in the treatment

• The predicted treatment is then substituted in the original equation to produce the 2nd stage regression:

$$Y_i = aX_i + b\hat{T}_i + e_i$$

- Assuming no covariates in the model, 2SLS leads to

$$b_{IV} = \frac{\text{cov}(Y_i, Z_i)}{\text{cov}(T_i, Z_i)}$$

- We can use the covariance formula to deduce the bias in 2SLS:

$$\text{cov}(Y_i, Z_i) = \text{cov}(bT_i + e_i, Z_i) = b\,\text{cov}(T_i, Z_i) + \text{cov}(e_i, Z_i) \Leftrightarrow$$

$$\Leftrightarrow b = \frac{\text{cov}(Y_i, Z_i) - \text{cov}(e_i, Z_i)}{\text{cov}(T_i, Z_i)} \Leftrightarrow$$

$$\Leftrightarrow b_{IV} = b + \frac{\text{cov}(e_i, Z_i)}{\text{cov}(T_i, Z_i)}$$

- Although detailed information on program implementation and participation can directly reveal the presence of selection bias, endogeneity of treatment can also be assessed using the Wu-Hausman test:
    - First regress T on Z and the covariates X, and obtain the corresponding residuals; these residuals reflect all unobserved heterogeneity affecting the treatment not captured by observables
    - Then, regress Y on T, X, and the estimated residuals; if the coefficient on the residuals is statistically different from zero, the null that T is exogenous conditional on observables is rejected

- The IV model has some variations:
  - One can write the instrument equation as a nonlinear binary response model (such as probit or logit); however, one should be cautious regarding the application of 2SLS, which assumes linearity (use maximum likelihood estimation)
  - If panel data exist, IV can be combined with a panel fixed-effects approach in the following model:

$$Y_{it} = \phi T_{it} + \delta X_{it} + \eta_i + \varepsilon_{it}$$

In this specification, time-invariant unobservables are accounted for through fixed effects or time-differencing; IVs can help with time-variant unobservables

Concerns

• The basic drawback of the IV approach is the potential difficulty in finding an appropriate instrumental variable

• Consistency of the IV estimate can be assessed through the expression:

$$b_{IV} = b + \frac{\text{cov}(e_i, Z_i)}{\text{cov}(T_i, Z_i)}$$

• The exclusion restriction assumption guarantees that the bias is zero; however, in real applications this will never happen

- If there is a weak relation between the treatment *T* and the instrument *Z* (WEAK INSTRUMENT PROBLEM), the denominator of the bias will be small; this means:
  - Even if the correlation between *Z* and ε is small, the bias will be large
  - The standard error of the IV estimate is likely to increase (predicted impact on the outcome measured less precisely)
  - One can test for weak instruments by looking at the F-statistic of the 1st stage regression ($F>10$ is the rule of thumb as of Stock et al., JBES, 2002)

• With multiple instruments, one can test the exclusion restriction (TEST OF OVERIDENTIFYING RESTRICTIONS):

- First, estimate the equation of interest by 2SLS and obtain the residuals
- Then regress the residuals on X and Z; obtain the R^2
- Use the null hypothesis that all the instrumental variables are uncorrelated with the residuals; this hypothesis is tested against the following statistic

$$nR^2 \sim Chi_q^2$$

where q is the number of instrumental variables minus the number of endogenous variables

LATE: Local Average Treatment Effects

- Which treatment effect does IV estimate? Under some conditions, LATE
- Imbens and Angrist (Econometrica, 1994) and Angrist et al. (JASA, 1996) introduce the LATE; Heckman (JHR, 1997) discusses the assumptions behind this interpretation of IV
- Let the instrument Z be binary; then we have two possibilities for *D* for an individual, D_{1i} and D_{0i} depending on the value of *Z*
- We have the following potential outcomes

$$Y_i(d,z) = \begin{cases} Y_i(1,1) \text{ if } D_i = 1, Z_i = 1 \\ Y_i(1,0) \text{ if } D_i = 1, Z_i = 0 \\ Y_i(0,1) \text{ if } D_i = 0, Z_i = 1 \\ Y_i(0,0) \text{ if } D_i = 0, Z_i = 0 \end{cases}$$

- Assumptions:
  - Independence of the instrument

$$(Y_i(D_{1i},1), Y_i(D_{0i},0), D_{1i}, D_{0i}) \perp Z_i$$

  - Exclusion restriction

$$Y_i(d,0) = Y_i(d,1) \text{ for } d = 0,1$$

  - First stage

$$E(D_{1i} - D_{0i}) \neq 0$$

  - Monotonicity

$$D_{1i} - D_{0i} \geq 0, \forall i$$

- LATE Theorem: Under the above assumptions (IV=Wald) and

$$\beta_{IV} = \frac{E(Y_i \mid Z_i = 1) - E(Y_i \mid Z_i = 0)}{E(D_i \mid Z_i = 1) - E(D_i \mid Z_i = 0)} = E(Y_{1i} - Y_{0i} \mid D_{1i} > D_{0i})$$

• So, under the assumptions, IV estimates

$$E(Y_{1i} - Y_{0i} \mid D_{1i} > D_{0i})$$

• This is the average treatment effect for a group defined by the condition

$$D_{1i} > D_{0i}$$

• Since D_i is zero or one,

$$D_{1i} > D_{0i} \Leftrightarrow (D_{1i} = 1, D_{0i} = 0)$$

• Then this is the group of individuals for whom the instrument changes the treatment

(Remember: Heckman's general selection framework)

- We follow here Heckman and Vytlacil (2005, Econometrica)
- Selection model with two potential outcomes

$$Y_0 = m_0(X, U_0) = m_0(X) + U_0$$

$$Y_1 = m_1(X, U_1) = m_1(X) + U_1$$

$$D^* = m_D(Z) - U_D$$

$$D = \begin{cases} 1 \; if \; D^* \geq 0 \\ 0 \; otherwise \end{cases}$$

$$Y = Y_1 D + Y_0(1 - D)$$

- Main additional assumptions

$$(i)\ \mu_D(Z)\ \textit{nondegenerate conditional on } X$$

$$(ii)\ (U_1, U_0, U_D) \perp Z \mid X$$

$$(iii)\ U_D\ \textit{distributed continuously}$$

- The first two are standard IV assumptions
- The third is for simpler exposition
- There are other technical assumptions that I omit here for simplicity (refer to the paper)

- Treatment effect definitions

$$(i)\ \mathrm{D}^{ATE}(x) \equiv E(\mathrm{D} \mid X = x),\ where\ \mathrm{D} = Y_1 - Y_0$$

$$(ii)\ \mathrm{D}^{TT}(x) \equiv E(\mathrm{D} \mid X = x, D = 1)$$

$$(iii)\ \mathrm{D}^{TU}(x) \equiv E(\mathrm{D} \mid X = x, D = 0)$$

- The paper argues that all treatment effects in the literature are functions of the following treatment effect

$$(iv)\ \mathrm{D}^{MTE}(x, u_D) \equiv E(\mathrm{D} \mid X = x, U_D = u_D)$$

- This is the Marginal Treatment Effect
    - A small value of u_D is likely to produce participation, a large value is likely not to produce participation

• The MTE is the limit form of the LATE, for an infinitesimal change in the instruments Z

• Another way to see this is that the LATE integrates MTE from an u_D equivalent to z' to one equivalent to z

• In doing so, it assumes the referred monotonicity assumption, by which when z increases no one drops participation but some people start participating in the program

• The MTE can then be estimated as a Local IV (LIV):

$$\Delta^{LIV}(x,p) \equiv \frac{\partial E(Y \mid X = x, P(Z) = p)}{\partial p}$$

Example: Childbearing and Labor Supply

- Angrist and Evans (AER, 1998)
  - Research on the labor supply consequences of childbearing is complicated by the endogeneity of fertility
  - Parental preferences for a mixed sibling-sex composition (i.e., parents of same-sex siblings are more likely to go on to have an additional child) are used to construct IV estimates of the effect of childbearing on labor supply:
    - A dummy for whether the sex of the second child matches the sex of the first child provides a plausible instrument for further childbearing among women with at least two children
  - They also use results generated using twins at second birth to construct instruments
  - They are also able to assess the time it takes for the labor-supply consequences of childbearing to disappear
  - Data come from the Census Public Use Micro Samples (PUMS), 1980 and 1990

TABLE 3—FRACTION OF FAMILIES THAT HAD ANOTHER CHILD BY PARITY AND SEX OF CHILDREN

Sex of first child in families with one or more children	All women: 1980 PUMS (649,887 observations) Fraction of sample	All women: 1980 PUMS Fraction that had another child	All women: 1990 PUMS (627,362 observations) Fraction of sample	All women: 1990 PUMS Fraction that had another child	Married women: 1980 PUMS (410,333 observations) Fraction of sample	Married women: 1980 PUMS Fraction that had another child	Married women: 1990 PUMS (477,798 observations) Fraction of sample	Married women: 1990 PUMS Fraction that had another child
(1) one girl	0.488	0.694 (0.001)	0.489	0.665 (0.001)	0.485	0.720 (0.001)	0.487	0.698 (0.001)
(2) one boy	0.512	0.694 (0.001)	0.511	0.667 (0.001)	0.515	0.720 (0.001)	0.513	0.699 (0.001)
difference (2) − (1)	—	0.000 (0.001)	—	0.002 (0.001)	—	0.000 (0.001)	—	0.001 (0.001)

Sex of first two children in families with two or more children	All women: 1980 PUMS (394,835 observations) Fraction of sample	All women: 1980 PUMS Fraction that had another child	All women: 1990 PUMS (380,007 observations) Fraction of sample	All women: 1990 PUMS Fraction that had another child	Married women: 1980 PUMS (254,654 observations) Fraction of sample	Married women: 1980 PUMS Fraction that had another child	Married women: 1990 PUMS (301,588 observations) Fraction of sample	Married women: 1990 PUMS Fraction that had another child
one boy, one girl	0.494	0.372 (0.001)	0.495	0.344 (0.001)	0.494	0.346 (0.001)	0.497	0.331 (0.001)
two girls	0.242	0.441 (0.002)	0.241	0.412 (0.002)	0.239	0.425 (0.002)	0.239	0.408 (0.002)
two boys	0.264	0.423 (0.002)	0.264	0.401 (0.002)	0.266	0.404 (0.002)	0.264	0.396 (0.002)
(1) one boy, one girl	0.494	0.372 (0.001)	0.495	0.344 (0.001)	0.494	0.346 (0.001)	0.497	0.331 (0.001)
(2) both same sex	0.506	0.432 (0.001)	0.505	0.407 (0.001)	0.506	0.414 (0.001)	0.503	0.401 (0.001)
difference (2) − (1)	—	0.060 (0.002)	—	0.063 (0.002)	—	0.068 (0.002)	—	0.070 (0.002)

Notes: The samples are the same as in Table 2. Standard errors are reported in parentheses.

TABLE 5—WALD ESTIMATES OF LABOR-SUPPLY MODELS

	1980 PUMS			1990 PUMS			1980 PUMS		
		Wald estimate using as covariate:			Wald estimate using as covariate:			Wald estimate using as covariate:	
Variable	Mean difference by *Same sex*	*More than 2 children*	*Number of children*	Mean difference by *Same sex*	*More than 2 children*	*Number of children*	Mean difference by *Twins-2*	*More than 2 children*	*Number of children*
More than 2 children	0.0600 (0.0016)	—	—	0.0628 (0.0016)	—	—	0.6031 (0.0084)	—	—
Number of children	0.0765 (0.0026)	—	—	0.0836 (0.0025)	—	—	0.8094 (0.0139)	—	—
Worked for pay	−0.0080 (0.0016)	−0.133 (0.026)	−0.104 (0.021)	−0.0053 (0.0015)	−0.084 (0.024)	−0.063 (0.018)	−0.0459 (0.0086)	−0.076 (0.014)	−0.057 (0.011)
Weeks worked	−0.3826 (0.0709)	−6.38 (1.17)	−5.00 (0.92)	−0.3233 (0.0743)	−5.15 (1.17)	−3.87 (0.88)	−1.982 (0.386)	−3.28 (0.63)	−2.45 (0.47)
Hours/week	−0.3110 (0.0602)	−5.18 (1.00)	−4.07 (0.78)	−0.2363 (0.0620)	−3.76 (0.98)	−2.83 (0.73)	−1.979 (0.327)	−3.28 (0.54)	−2.44 (0.40)
Labor income	−132.5 (34.4)	−2208.8 (569.2)	−1732.4 (446.3)	−119.4 (42.4)	−1901.4 (670.3)	−1428.0 (502.6)	−570.8 (186.9)	−946.4 (308.6)	−705.2 (229.8)
ln(*Family income*)	−0.0018 (0.0041)	−0.029 (0.068)	−0.023 (0.054)	−0.0085 (0.0047)	−0.136 (0.074)	−0.102 (0.056)	−0.0341 (0.0223)	−0.057 (0.037)	−0.042 (0.027)

Notes: The samples are the same as in Table 2. Standard errors are reported in parentheses.

The following regression models are used to link labor-supply variables for husbands and wives to the endogenous *More than 2* variable, x_i, and the list of exogenous covariates, including additive effects for the sex of each child:

$$y_i = \alpha_0' \mathbf{w}_i + \alpha_1 s_{1i} + \alpha_2 s_{2i} + \beta x_i + \varepsilon_i, \tag{4}$$

where $\mathbf{w}_i$ is a vector of demographic variables, and s_{1i} and s_{2i} are indicators for the sex of the first two children of mother i. Initially, $\mathbf{w}_i$ is limited to variables that are clearly exogenous to fertility: mother's age and age at first birth, plus race and Hispanic indicators. In the just-identified model where *Same sex* is the only instrument, the first-stage equation relating *More than 2 children* to sex mix is

$$x_i = \pi_0' \mathbf{w}_i + \pi_1 s_{1i} + \pi_2 s_{2i} + \gamma(\mathit{Same\ sex}_i) + \eta_i, \tag{5}$$

where γ is the first-stage effect of the instrument.

TABLE 6—OLS ESTIMATES OF *MORE THAN 2 CHILDREN* EQUATIONS

	All women			Married women		
Independent variable	(1)	(2)	(3)	(4)	(5)	(6)
1980 PUMS						
Boy 1st	—	−0.0080 (0.0015)	0.0001 (0.0021)	—	−0.0111 (0.0018)	−0.0016 (0.0026)
Boy 2nd	—	−0.0081 (0.0015)	—	—	−0.0095 (0.0018)	—
Same sex	0.0600 (0.0016)	0.0617 (0.0015)	—	0.0675 (0.0019)	0.0694 (0.0018)	—
Two boys	—	—	0.0536 (0.0021)	—	—	0.0598 (0.0026)
Two girls	—	—	0.0698 (0.0021)	—	—	0.0789 (0.0026)
With other covariates	no	yes	yes	no	yes	yes
R^2	0.004	0.084	0.084	0.005	0.078	0.078
1990 PUMS						
Boy 1st	—	−0.0081 (0.0015)	−0.0083 (0.0022)	—	−0.0097 (0.0017)	−0.0086 (0.0024)
Boy 2nd	—	0.0002 (0.0015)	—	—	−0.0011 (0.0017)	—
Same sex	0.0628 (0.0016)	(0.0623) (0.0015)	—	0.0702 (0.0018)	0.0703 (0.0017)	—
Two boys	—	—	0.0624 (0.0021)	—	—	0.0692 (0.0023)
Two girls	—	—	0.0621 (0.0022)	—	—	0.0714 (0.0024)
With other covariates	no	yes	yes	no	yes	yes
R^2	0.004	0.082	0.082	0.005	0.082	0.082

Notes: Other covariates in the models are indicators for *Age*, *Age at first birth*, *Black*, *Hispanic*, and *Other race*. The variable *Boy 2nd* is excluded from columns (3) and (6). Standard errors are reported in parentheses.

TABLE 7—OLS AND 2SLS ESTIMATES OF LABOR-SUPPLY MODELS USING 1980 CENSUS DATA

	All women			Married women			Husbands of married women		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Estimation method	OLS	2SLS	2SLS	OLS	2SLS	2SLS	OLS	2SLS	2SLS
Instrument for *More than 2 children*	—	*Same sex*	*Two boys, Two girls*	—	*Same sex*	*Two boys, Two girls*	—	*Same sex*	*Two boys, Two girls*
Dependent variable:									
Worked for pay	−0.176	−0.120	−0.113	−0.167	−0.120	−0.113	−0.008	0.004	0.001
	(0.002)	(0.025)	(0.025)	(0.002)	(0.028)	(0.028)	(0.001)	(0.009)	(0.008)
			[0.013]			[0.013]			[0.013]
Weeks worked	−8.97	−5.66	−5.37	−8.05	−5.40	−5.16	−0.82	0.59	0.45
	(0.07)	(1.11)	(1.10)	(0.09)	(1.20)	(1.20)	(0.04)	(0.60)	(0.59)
			[0.017]			[0.071]			[0.030]
Hours/week	−6.66	−4.59	−4.37	−6.02	−4.83	−4.61	0.25	0.56	0.50
	(0.06)	(0.95)	(0.94)	(0.08)	(1.02)	(1.01)	(0.05)	(0.70)	(0.69)
			[0.030]			[0.049]			[0.71]
Labor income	−3768.2	−1960.5	−1870.4	−3165.7	−1344.8	−1321.2	−1505.5	−1248.1	−1382.3
	(35.4)	(541.5)	(538.5)	(42.0)	(569.2)	(565.9)	(103.5)	(1397.8)	(1388.9)
			[0.126]			[0.703]			(0.549)
ln(*Family income*)	−0.126	−0.038	−0.045	−0.132	−0.051	−0.053	—	—	—
	(0.004)	(0.064)	(0.064)	(0.004)	(0.056)	(0.056)			
			[0.319]			[0.743]			
ln(*Non-wife income*)	—	—	—	−0.053	0.023	0.016	—	—	—
				(0.005)	(0.066)	(0.066)			
						[0.297]			

Notes: The table reports estimates of the coefficient on the *More than 2 children* variable in equations (4) and (6) in the text. Other covariates in the models are *Age*, *Age at first birth*, plus indicators for *Boy 1st*, *Boy 2nd*, *Black*, *Hispanic*, and *Other race*. The variable *Boy 2nd* is excluded from equation (6). The *p*-value for the test of overidentifying restrictions associated with equation (6) is shown in brackets. Standard errors are reported in parentheses.

Table 8—OLS and 2SLS Estimates of Labor-Supply Models Using 1990 Census Data

	All women			Married women			Husbands of married women		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Estimation method	OLS	2SLS	2SLS	OLS	2SLS	2SLS	OLS	2SLS	2SLS
Instrument for *More than 2 children*	—	*Same sex*	*Two boys, Two girls*	—	*Same sex*	*Two boys, Two girls*	—	*Same sex*	*Two boys, Two girls*
Dependent variable:									
Worked for pay	−0.155	−0.092	−0.092	−0.147	−0.104	−0.104	−0.102	0.017	0.017
	(0.002)	(0.024)	(0.024)	(0.002)	(0.024)	(0.024)	(0.001)	(0.009)	(0.009)
			[0.743]			[0.576]			[0.989]
Weeks worked	−8.71	−5.66	−5.64	−8.25	−5.76	−5.76	−1.03	1.01	1.01
	(0.08)	(1.16)	(1.16)	(0.09)	(1.15)	(1.15)	(0.05)	(0.63)	(0.63)
			[0.391]			[0.670]			[0.708]
Hours/week	−6.80	−4.08	−4.10	−6.39	−3.94	−3.95	−0.06	0.85	0.83
	(0.07)	(0.98)	(0.98)	(0.07)	(0.96)	(0.96)	(0.05)	(0.69)	(0.69)
			[0.489]			[0.665]			[0.180]
Labor income	−3984.4	−2099.6	−2096.2	−3753.9	−2457.5	−2456.3	929.7	1348.7	1354.8
	(44.2)	(664.0)	(663.8)	(50.7)	(669.7)	(669.7)	(114.9)	(1536.0)	(1535.9)
			[0.830]			[0.893]			[0.711]
ln(*Family income*)	−0.119	−0.124	−0.122	−0.103	−0.054	−0.054	—	—	—
	(0.005)	(0.071)	(0.071)	(0.004)	(0.051)	(0.051)			
			[0.270]			[0.878]			
ln(*Non-wife income*)	—	—	—	−0.004	0.020	0.020	—	—	—
				(0.005)	(0.068)	(0.068)			
						[0.452]			

Notes: The table reports the coefficient on the *More than 2 children* variable in equations (4) and (6) in the text estimated with 1990 Census data. Other covariates in the models are *Age, Age at first birth,* plus indicators for *Boy 1st, Boy 2nd, Black, Hispanic,* and *Other race.* The variable *Boy 2nd* is excluded from equation (6). The p-value for the test of overidentifying restrictions associated with equation (6) is shown in brackets. Standard errors are reported in parentheses.

TABLE 11—COMPARISION OF 2SLS ESTIMATES USING *SAME SEX* AND *TWINS-2* INSTRUMENTS IN 1980 CENSUS DATA

	All women		Married women		Husbands	
Model	(1)	(2)	(1)	(2)	(1)	(2)
Instrument for *More than 2 children*	*Same sex*	*Twins-2*	*Same sex*	*Twins-2*	*Same sex*	*Twins-2*
Dependent variable:						
Worked for pay	−0.125	−0.079	−0.123	−0.087	0.004	−0.001
	(0.026)	(0.013)	(0.028)	(0.017)	(0.009)	(0.005)
Weeks worked	−5.82	−3.64	−5.47	−4.21	0.65	−0.35
	(1.15)	(0.60)	(1.23)	(0.72)	(0.61)	(0.36)
Hours/week	−4.76	−3.33	−4.91	−3.49	0.57	−0.49
	(0.98)	(0.51)	(1.03)	(0.61)	(0.71)	(0.42)
Labor income	−1961.7	−1262.2	−1329.8	−1453.1	−1194.8	616.8
	(560.5)	(292.8)	(579.1)	(339.8)	(1421.4)	(836.9)
ln(*Family income*)	−0.021	−0.071	−0.049	−0.025	—	—
	(0.067)	(0.035)	(0.057)	(0.033)		
ln(*Non-wife income*)	—	—	0.026	0.051	—	—
			(0.068)	(0.040)		

Notes: The table reports 2SLS estimates of the coefficient on *More than 2 children* in equation (4) in the text using *Same sex* and *Twins-2* as instruments. Other covariates in the models are *Age*, *Age at first birth*, ages of the first two children, plus indicators for *Boy 1st*, *Boy 2nd*, *Black*, *Hispanic*, and *Other race*. Data are from the 1980 Census. Standard errors are reported in parentheses.

What have we learnt?

- IV enables taking into account time-varying unobservables in estimating treatment effects
- However, it is demanding on the conditions for suitable instruments (problem of weak instruments, exclusion restriction)
- However, in an heterogeneous effect world, it estimates a very specific treatment effect (under restrictive assumptions)
    - Estimating the MTE and building the TE of interest from there may be the way to go (but not very popular to date in the applied literature)

Problem Set 4

Exercise: Follow the instructions in chapter 15 of Khandker et al (2009). You should produce a do file and a log file, which should be commented to show that you understood the results. These should be emailed to the grader (Matilde Grácio): only one email per group, please.

Note: We should be able to run the do file on a computer given the original datafile and produce the raw log file.

Due date: Tuesday, March 15.