ECMA31100 Introduction to Empirical Analysis II

Winter 2022, Week 3: Discussion Session

Conroy Lau

The University of Chicago

January 26, 2022

Welcome to week 3 of ECMA31100!

Problem set

- Please remember to submit your problem set on time!
- Check the latest announcement post on Canvas

TA sessions: new time and location

- Time: Wednesdays, 3:30pm — 4:20pm
- Location: HM 130

My office hours

- Time: Wednesdays, 5:30pm — 7:30pm
- Location: Zoom, see link on Canvas
- Feel free to send me an e-mail if you plan to come, so I can prepare ahead

Review and plans

Last TA session

- Some general advice and resources for further readings
- Imputation methods
- Matching estimators (matching on covariates and propensity scores)

Plans for today

- Finish our discussion on blocking (see my slides from week 2)
- Discuss Abadie (2003)

Please let me know if you have any suggestions or if you find any typos in the slides

Any questions?

Contents

Model and terminology

Notations

- Outcome: Y
- Binary treatment: $D \in \{0, 1\}$
- Binary instrument: $Z \in \{0, 1\}$

Potential outcomes

- $Y = DY_1 + (1 - D)Y_0$, where Y_d is the potential outcome if $D = d$
- $D = ZD_1 + (1 - Z)D_0$, where D_z is the potential outcome if $Z = z$

Four groups

Always-takers (a)	Compliers (c)	Never-takers (n)	Defiers (d)
$D_1 = D_0 = 1$	$D_1 = 1$ and $D_0 = 0$	$D_1 = D_0 = 0$	$D_1 = 0$ and $D_0 = 1$

Abadie (2003)'s κ

Assumptions

- Independence: $(Y_1, Y_0, D_1, D_0) \perp\!\!\!\perp Z | X$
- Exclusion: $\mathbb{P}[Y_{D_1} = Y_{D_0} | X] = 1$
- First stage: $0 < \mathbb{P}[Z = 1|X] < 1$ and $\mathbb{P}[D_1 = 1|X] > \mathbb{P}[D_0 = 1|X]$
- Monotonicity: $\mathbb{P}[D_1 \geq D_0 | X] = 1$

Theorem 3.1 of Abadie (2003)

- For any measurable $f(Y, X, D)$ such that $\mathbb{E}|f(Y, D, X)| < \infty$, we have

$$\mathbb{E}[f(Y, X, D)|D_1 > D_0] = \frac{1}{\mathbb{P}[D_1 > D_0]}\mathbb{E}[\kappa f(Y, X, D)],$$

where

$$\kappa \equiv 1 - \frac{D(1 - Z)}{\mathbb{P}[Z = 0|X]} - \frac{(1 - D)Z}{\mathbb{P}[Z = 1|X]}$$

Abadie (2003)'s κ

Some intuition

- We cannot observe both D_0 and $D_1 \Longrightarrow$ Compliers are not identified individually
- Distribution of observables can be described for the population of the compliers
- κ: Compliers $= 1 -$ Always takers $-$ Never takers

Proof strategy

Proof

- Notations: Let $G \in \{a, c, n\}$ (defiers are ruled out by monotonicity)
  - $G = a$ corresponds to $D_0 = D_1 = 1$
  - $G = c$ corresponds to $D_1 = 1 > 0 = D_0$
  - $G = n$ corresponds to $D_0 = D_1 = 0$
- Step 1: Show that $\mathbb{E}[\kappa | G, X, D, Y] = \mathbb{1}[G = c]$
- Step 2: Use the law of iterated expectations on $\mathbb{E}[\kappa f(Y, X, D)]$

Any questions?

Abadie (2003)'s κ

Estimation: least squares

General set-up

- Let θ_0 be the vector of parameters such that $\mathbb{E}[Y|X, D, D_1 > D_0] = h(D, X, \theta_0)$
- Then, $\theta_0 = \underset{\theta \in \Theta}{\arg\min}\, \mathbb{E}[(Y - h(D, X, \theta))^2 | D_1 > D_0]$ but we cannot identify the compliers
- By the last slide, the above is equivalent to $\theta_0 = \underset{\theta \in \Theta}{\arg\min}\, \mathbb{E}[\kappa(Y - h(D, X, \theta))^2]$
- The sample analog is $\widehat{\theta} = \underset{\theta \in \Theta}{\arg\min}\, \frac{1}{n} \sum_{i=1}^{n} \kappa_i [y_i - h(d_i, x_i, \theta)]^2$

Estimation of κ_i (not just for least squares)

- Each κ_i is estimated by $\kappa_i = 1 - \frac{d_i(1-z_i)}{\widehat{\mathbb{P}}[Z=0|x_i]} - \frac{(1-d_i)z_i}{\widehat{\mathbb{P}}[Z=1|x_i]}$
- We can use probit to estimate $\widehat{\mathbb{P}}[Z = 0|X]$ and $\widehat{\mathbb{P}}[Z = 1|X]$

Abadie (2003)'s κ

Estimation: parameterization for least squares

Least squares: linear parametrization

- Suppose that $h(D, X, \theta) = \alpha D + X'\beta$ where $\theta = (\alpha, \beta)$
- Then, we need to solve: $(\widehat{\alpha}, \widehat{\beta}) = \underset{(\alpha,\beta)\in\Theta}{\arg\min} \dfrac{1}{n}\sum_{i=1}^{n} \kappa_i (y_i - \alpha d_i - x_i'\beta)^2$

Least squares: probit

- Why probit? Make sure the $\Phi(\cdot) \in [0, 1]$
- Suppose that $h(D, X, \theta) = \Phi(\alpha D + X'\beta)$ where $\theta = (\alpha, \beta)$ and $\Phi(\cdot)$ is the normal cdf
- Then, we need to solve: $(\widehat{\alpha}, \widehat{\beta}) = \underset{(\alpha,\beta)\in\Theta}{\arg\min} \dfrac{1}{n}\sum_{i=1}^{n} \kappa_i [y_i - \Phi(\alpha d_i + x_i'\beta)]^2$

Abadie (2003)'s κ

Estimation: maximum likelihood

General set-up

- Let $f(Y, D, X, \theta)$ be the density, then we solve $\theta_0 = \arg\max_{\theta \in \Theta} \mathbb{E}[\ln f(Y, D, X, \theta) | D_1 > D_0]$
- Again, the above is equivalent to $\theta_0 = \arg\max_{\theta \in \Theta} \mathbb{E}[\kappa \ln f(Y, D, X, \theta)]$
- The sample analog is $\widehat{\theta} = \arg\max_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^{n} \kappa_i \ln f(y_i, d_i, x_i, \theta)^2$

One possible parameterization

- We may specify f as $f(Y, D, X, \beta) = \Phi(\alpha D + X'\beta)^Y [1 - \Phi(\alpha D + X'\beta)]^{1-Y}$ as Y is binary
- Then, solve $(\widehat{\alpha}, \widehat{\beta}) = \arg\max_{(\alpha,\beta) \in \Theta} \frac{1}{n} \sum_{i=1}^{n} \kappa_i \{y_i \ln \Phi(\alpha d_i + x_i'\beta) + (1 - y_i) \ln[1 - \Phi(\alpha d_i + x_i'\beta)]\}$

Abadie (2003)'s κ

Empirical application

Question: Impact of 401(k) retirement programs on savings

- Only workers in firms that offer 401(k) are eligible
- Do these tax-deferred retirement plans imply:
  - More savings?
  - Crowd out other types of savings?

Some discussion on the assumptions

- 401(k) eligibility is ignorable given some observables $\Longrightarrow$ Can be used as IV
  - Poterba et al. (1994, 1995): eligibility could be unrelated to saving preferences
- Only eligible individuals can open a 401(k) account $\Longrightarrow$ Monotonicity holds

Abadie (2003)'s κ

Empirical application

Table 1: Summary statistics (taken from table 1 of Abadie (2000))

	Entire Sample	By 401(k) participation		By 401(k) eligibility	
		Participants	Non-participants	Eligibles	Non-eligibles
Treatment:					
Participation in 401(k)	0.28 (0.45)			0.70 (0.46)	0.00 (0.00)
Instrument:					
Eligibility for 401(k)	0.39 (0.49)	1.00 (0.00)	0.16 (0.37)		
Outcome variables:					
Family Net Financial Assets	19,071.68 (63,963.84)	38,472.96 (79,271.08)	11,667.22 (55,289.23)	30,535.09 (75,018.98)	11,676.77 (54,420.17)
Participation in IRA	0.25 (0.44)	0.36 (0.48)	0.21 (0.41)	0.32 (0.47)	0.21 (0.41)
Covariates:					
Family Income	39,254.64 (24,090.00)	49,815.14 (26,814.24)	35,224.25 (21,649.17)	47,297.81 (25,620.00)	34,066.10 (21,510.64)
Age	41.08 (10.30)	41.51 (9.65)	40.91 (10.53)	41.48 (9.61)	40.82 (10.72)
Married	0.63 (0.48)	0.70 (0.46)	0.60 (0.49)	0.68 (0.47)	0.60 (0.49)
Family Size	2.89 (1.53)	2.92 (1.47)	2.87 (1.55)	2.91 (1.48)	2.87 (1.56)

Abadie (2003)'s κ

Empirical application

Table 2: Summary statistics (taken from table 2 of Abadie (2000))

Dependent Variable: Family Net Financial Assets (in $)

		Endogenous Treatment		
		Two Stage Least Squares		
	Ordinary Least Squares (1)	First Stage (2)	Second Stage (3)	Least Squares Treated (4)
Participation in 401(k)	13,527.05 (1,810.27)		9,418.83 (2,152.89)	10,800.25 (2,261.55)
Constant	-23,549.00 (2,178.08)	-0.0306 (0.0087)	-23,298.74 (2,167.39)	-27,133.56 (3,212.35)
Family Income (in thousand $)	976.93 (83.37)	0.0013 (0.0001)	997.19 (83.86)	982.37 (106.65)
Age (minus 25)	-376.17 (236.98)	-0.0022 (0.0010)	-345.95 (238.10)	312.30 (371.76)
Age (minus 25) square	38.70 (7.67)	0.0001 (0.0000)	37.85 (7.70)	24.44 (11.40)
Married	-8,369.47 (1,829.93)	-0.0005 (0.0079)	-8,355.87 (1,829.67)	-6,646.69 (2,742.77)
Family Size	-785.65 (410.78)	0.0001 (0.0024)	-818.96 (410.54)	-1,234.25 (647.42)
Eligibility for 401(k)		0.6883 (0.0080)		

Abadie (2003)'s κ

Empirical application

Table 3: Summary statistics (taken from table 3 of Abadie (2000))

Dependent Variable: IRA Account

	Linear Response			Probit Response				
		Endogenous Treatment				Endogenous Treatment		
	Least Sq. (1)	Two Stage Least Sq. (2)	Least Sq. Treated (3)	Probit (4)	Least Sq. (5)	Bivariate Probit (6)	Probit Treated (7)	Least Sq. Treated (8)
Participation in 401(k)	0.0569 (0.0103)	0.0274 (0.0132)	0.0253 (0.0131)	0.0712 (0.0121)	0.0699 (0.0126)	0.0407 (0.0156)	0.0358 (0.0161)	0.0264 (0.0172)
Family Income (in thousand $)	0.0059 (0.0002)	0.0060 (0.0002)	0.0060 (0.0003)	0.0069 (0.0003)	0.0070 (0.0003)	0.0069 (0.0003)	0.0069 (0.0004)	0.0072 (0.0005)
Age (minus 25)	0.0074 (0.0014)	0.0076 (0.0014)	0.0119 (0.0025)	0.0149 (0.0022)	0.0153 (0.0023)	0.0147 (0.0021)	0.0183 (0.0034)	0.0207 (0.0037)
Age (minus 25) square	0.0000 (0.0000)	0.0000 (0.0000)	-0.0001 (0.0001)	-0.0001 (0.0001)	-0.0001 (0.0001)	-0.0001 (0.0001)	-0.0002 (0.0001)	-0.0002 (0.0001)
Married	0.0312 (0.0110)	0.0313 (0.0110)	0.0440 (0.0184)	0.0590 (0.0152)	0.0477 (0.0166)	0.0577 (0.0148)	0.0627 (0.0231)	0.0535 (0.0244)
Family Size	-0.0264 (0.0032)	-0.0266 (0.0032)	-0.0340 (0.0053)	-0.0424 (0.0050)	-0.0403 (0.0056)	-0.0415 (0.0049)	-0.0472 (0.0075)	-0.0480 (0.0082)

Any questions?

Contents

Proof of Theorem 3.1 in Abadie (2003)

Back

Proof

- Consider the case with $G = c$:

$$\begin{aligned}
\mathbb{E}[\kappa | G = c, X, D, Y] &= \mathbb{E}\left[1 - \frac{D(1-Z)}{\mathbb{P}[Z=0|X]} - \frac{(1-D)Z}{\mathbb{P}[Z=1|X]} \middle| G = c, X, D, Y\right] \\
&= \mathbb{E}\left[1 - 0 - 0 \middle| G = c, X, D, Y\right] \\
&= 1
\end{aligned}$$

Proof of Theorem 3.1 in Abadie (2003)

Proof (Cont'd)

- Consider the case with $G = a$:

$$
\begin{aligned}
\mathbb{E}[\kappa|G=a,X,D,Y] &= \mathbb{E}\left[1 - \frac{D(1-Z)}{\mathbb{P}[Z=0|X]} - \frac{(1-D)Z}{\mathbb{P}[Z=1|X]}\middle| G=a,X,D,Y\right] \\
&= \mathbb{E}\left[1 - \frac{1-Z}{\mathbb{P}[Z=0|X]} - 0\middle| G=a,X,D,Y\right] \\
&= 1 - \mathbb{E}\left[\frac{1-Z}{\mathbb{P}[Z=0|X]}\middle| X\right] \\
&= 1 - \frac{\mathbb{P}[Z=0|X]}{\mathbb{P}[Z=0|X]} \\
&= 0
\end{aligned}
$$

Proof of Theorem 3.1 in Abadie (2003)

Back

Proof (Cont'd)

- Consider the case with $G = n$:

$$
\begin{aligned}
\mathbb{E}[\kappa | G = n, X, D, Y] &= \mathbb{E}\left[1 - \frac{D(1-Z)}{\mathbb{P}[Z=0|X]} - \frac{(1-D)Z}{\mathbb{P}[Z=1|X]} \middle| G = n, X, D, Y\right] \\
&= \mathbb{E}\left[1 - 0 - \frac{Z}{\mathbb{P}[Z=1|X]} \middle| G = n, X, D, Y\right] \\
&= 1 - \mathbb{E}\left[\frac{Z}{\mathbb{P}[Z=1|X]} \middle| X\right] \\
&= 1 - \frac{\mathbb{P}[Z=1|X]}{\mathbb{P}[Z=1|X]} \\
&= 0
\end{aligned}
$$

Proof of Theorem 3.1 in Abadie (2003)

Back

Proof (Cont'd)

- Combining the last three slides, we have $\mathbb{E}[\kappa|G, X, D, Y] = \mathbb{1}[G = c]$
- By the law of iterated expectations, we obtain

$$\begin{aligned}
\mathbb{E}[\kappa f(Y, X, D)] &= \mathbb{E}[\mathbb{E}[\kappa f(Y, X, D)|G, Y, X, D]] \\
&= \mathbb{E}[f(Y, X, D)\mathbb{E}[\kappa|G, Y, X, D]] \\
&= \mathbb{E}[f(Y, X, D)\mathbb{1}[G = c]] \\
&= \mathbb{P}[G = c]\mathbb{E}[f(Y, X, D)|G = c],
\end{aligned}$$

- Rearranging yields

$$\mathbb{E}[f(Y, X, D)|G = c] = \frac{1}{\mathbb{P}[G = c]}\mathbb{E}[\kappa f(Y, X, D)]$$

References 1

Abadie, A. (2000): "Semiparametric instrumental variable estimation for causal effects," *Working paper*.

——— (2003): "Semiparametric instrumental variable estimation for causal effects," *Journal of Econometrics*, 113, 231–263.

Andrews, I., J. H. Stock, and L. Sun (2019): "Weak Instruments in Instrumental Variables Regression: Theory and Practice," *Annual Review of Economics*, 11, 727–753.

Poterba, J. M., S. F. Venti, and D. A. Wise (1994): "401(k) plans and tax-deferred savings," in *Studies in the Economics of Aging*, ed. by D. A. Wise, University of Chicago Press.

——— (1995): "Do 401(k) contributions crow out other personal savings?" *Journal of Public Economics*, 56, 1–32.