Supplementary Materials

for "Profiling Compliers and Non-compliers for Instrumental-Variable Analysis"

Moritz Marbach
Center for Comparative and International Studies, ETH Zurich, 8092 Zurich, Switzerland

Dominik Hangartner
Center for Comparative and International Studies, ETH Zurich, 8092 Zurich, Switzerland
Department of Government, London School of Economics, London WC2A 2AE, United Kingdom
e-mail: d.hangartner@lse.ac.uk (corresponding author)

Contents

Let N be the sample size, K_{nt} the number of observed never-takers, K_{at} the number of observed always-takers, and $N_{Z=z}$ the number of units with realized assignment status z. Let x_i be the covariate value for the i^{th} unit, Z_i the instrumental variable and D_i^{obs} the corresponding observed treatment status. We then define the following estimators for the mean of the entire sample, $\hat{\mu}$, always-takers $\hat{\mu}_{at}$, and never-takers $\hat{\mu}_{nt}$:

$$\begin{aligned}\hat{\mu} &= \frac{1}{N}\sum_{i=1}^{N} x_i \\ \hat{\mu}_{at} &= \frac{1}{K_{at}}\sum_{i=1}^{K_{at}} x_i \\ \hat{\mu}_{nt} &= \frac{1}{K_{nt}}\sum_{i=1}^{K_{nt}} x_i.\end{aligned} \tag{8}$$

With the sample shares of always-takers, $\hat{\pi}_{at}$, never-takers, $\hat{\pi}_{nt}$, and compliers, $\hat{\pi}_{nt}$, estimated as:

$$\begin{aligned}\hat{\pi}_{at} =& \frac{1}{N_{Z=0}}\sum_{i=1}^{N_{Z=0}} D_i^{obs} \\ \hat{\pi}_{nt} =& \frac{1}{N_{Z=1}}\sum_{i=1}^{N_{Z=1}} (1 - D_i^{obs}) \\ \hat{\pi}_{co} =& 1 - \hat{\pi}_{at} - \hat{\pi}_{nt},\end{aligned} \tag{9}$$

we can back out the complier mean of X:

$$\hat{\mu}_{co} = \frac{1}{\pi_{co}}\hat{\mu} - \frac{\pi_{nt}}{\pi_{co}}\hat{\mu}_{nt} - \frac{\pi_{at}}{\pi_{co}}\hat{\mu}_{at}. \tag{10}$$

Because the sample shares $\hat{\pi}_{co}$, $\hat{\pi}_{at}$, and $\hat{\pi}_{nt}$, have to be estimated, the derivation of the standard error for the complier mean, $\hat{\mu}_{co}$, is somewhat tedious. Given the extremely low computational costs of this estimator, we use the bootstrap method to obtain standard errors that reflect the estimation uncertainty in the covariate means and the sample proportions.

Monte Carlo Experiments

We conducted a Monte Carlo experiment to verify that the empirical coverage rate matches the nominal coverage rate, and that the mean absolute error decreases with larger samples. We used 13 different sample sizes from N=250 to N=24,000, and used 1,000 replications for each run.

For each run we simulated a dataset with N units that are randomly assigned to one of three strata: complier, always-taker and never-taker. For each unit we simulated a realized value of the instrument, the corresponding observed treatment status, and a value for the continuous covariate. We varied the probability that the unit would belong to each strata by drawing from a uniform Dirichlet density under the constraint that no strata's share is less than 0.1. The binary instrument is drawn from a Bernoulli density with a probability parameter that is drawn from a uniform density in a range from 0.1 to 0.9. To simulate the covariate distribution, we draw from three normal densities, one for each strata. The mean parameters for these normal densities are drawn from a uniform distribution with support -2 to 2, and the standard deviations from a uniform density with support 0.25 to 2.

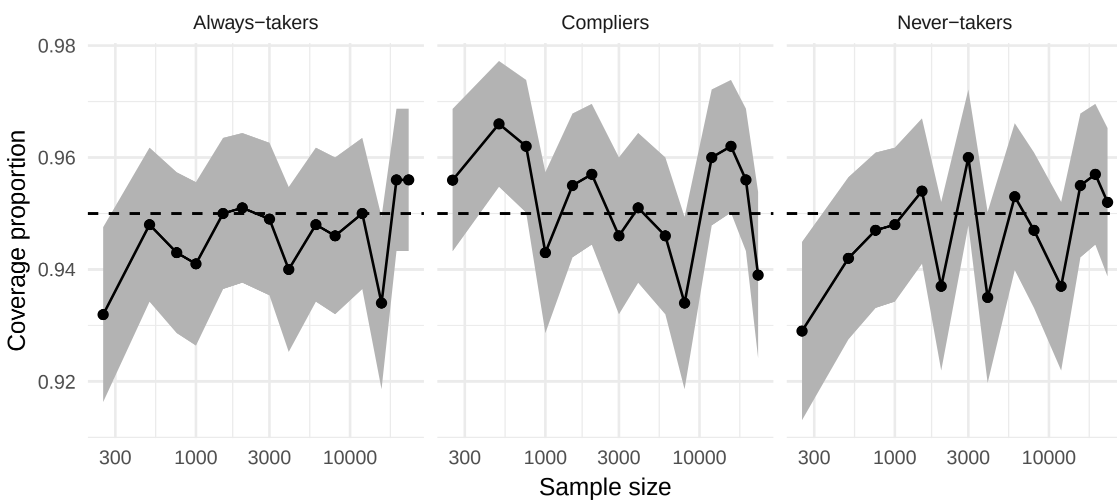

Figure S.1: The average coverage rate for the 95% bootstrap confidence interval across 13 Monte Carlo experiments with varying sample sizes.

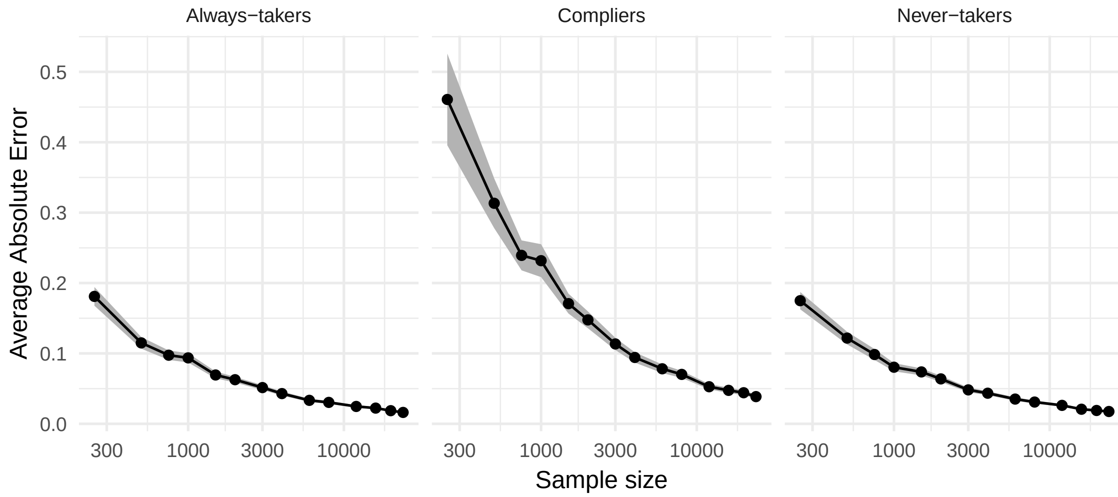

Figure S.2: The average absolute error for the mean across 13 Monte Carlo experiments with varying sample sizes.

Further Results for Main Application

	First-stage	ITT	LATE
Washington Post			
Assigned [0,1]	0.353	0.079	
	(0.042)	(0.045)	
Received [0,1]			0.223
			(0.129)
(Intercept)	0.211	0.411	0.364
	(0.024)	(0.029)	(0.051)
Num. obs.	503	503	503

Table S.1: OLS and 2SLS estimates with heteroscedasticity-robust standard errors in parentheses using the data from Gerber, Karlan and Bergan (2009) as described in the main text. The outcome is vote intention for the Democratic candidate in the 2005 Virgina gubernatorial election.

Variable	Always-taker	Never-taker	Complier	Sample
Proportion [0-1]	0.20 (0.02)	0.45 (0.03)	0.35 (0.03)	-
Age [17-94]	52.26 (1.20)	49.18 (1.11)	47.75 (1.54)	49.30 (0.48)
Female [0,1]	0.32 (0.05)	0.39 (0.04)	0.28 (0.06)	0.34 (0.02)
Voted 2004 [0,1]	0.48 (0.05)	0.47 (0.04)	0.45 (0.06)	0.46 (0.02)
Voted 2002 [0,1]	0.33 (0.05)	0.26 (0.04)	0.28 (0.05)	0.28 (0.02)
Voted 2001 [0,1]	0.04 (0.02)	0.04 (0.02)	0.05 (0.03)	0.04 (0.01)
Preferes Republican candidate [0,1]	0.20 (0.04)	0.24 (0.04)	0.19 (0.05)	0.21 (0.01)
Preferes Democratic candidate [0,1]	0.18 (0.04)	0.20 (0.03)	0.22 (0.05)	0.20 (0.01)
No candidate preference [0,1]	0.18 (0.04)	0.11 (0.03)	0.09 (0.04)	0.12 (0.01)

Table S.2: Descriptive statistics (mean and 95% bootstrap confidence intervals) for the complier and non-complier subpopulations in the study by Gerber, Karlan and Bergan (2009).

Supplementary Application: Fox News Study

Albertson and Lawrence (2009) report the results of a randomized experiment in which 259 out of $N = 507$ participants in Orange County were assigned to watch a Fox News debate on affirmative action. This debate took place the evening before California residents voted in a referendum on whether to abolish affirmative action (Proposition 209, known as the California Civil Rights Initiative). When the participants were re-interviewed shortly after casting their ballot, they indicated whether they had watched the debate, and whether they had voted to abolish affirmative action. The baseline survey also included various questions about the participants' socio-economic background. Based on these data, and using the standard 2SLS IV estimator, the authors estimated the effect of watching the Fox debate on knowledge of and support for the proposition within the complier strata. Among the compliers, the authors found that viewing the debate moderately increased viewers' self-assessed knowledge about the proposition on a 4-point scale ($\mathrm{coef} = 0.27$, $p = 0.09$) but did not affect their support for it ($\mathrm{coef} = -0.07$, $p = 0.47$). The full replication results based on the replication materials of Aronow and Carnegie (2013) are included in the SM Table S.3.

Next, we used the estimator described above to profile compliers and non-compliers. Figure S.3 shows the covariate means for the entire sample, and the sample shares and covariate means for compliers, always-takers and never-takers for eight socio-economic background characteristics. Numerical estimates are provided in SM Table S.4. Approximately 55% of the experimental participants are never-takers, 41% are compliers, and only 4% are always-takers. The low number of always-takers implies that any estimate for that group will likely have a large variance.

The upper panel of SM Figure S.3 shows that all strata have relatively similar socio-economic backgrounds. Compliers and never-takers tend to be similar in terms of race, education, income, gender and ideology. However, we do find politically and statistically significant differences for political interest and news consumption. Compared to never-takers, compliers tend to be more politically interested, watch more TV news, and read the newspaper more often.

In the Fox debate study, participants in the encouragement group were asked to watch the program, sent a reminder (including a refrigerator magnet with the day and the time) and given a token $2 incentive. Thus, a larger financial reward for watching the Fox debate might strengthen the instrument by incentivizing some marginal never-takers to become compliers. Would we expect the treatment effect for these additional compliers to be the same as the LATE in the original study? Given the differences in the level of political interest between compliers and never-takers, and the frequency with which they to watch TV news and read the newspaper (all factors that are likely related to attitudes on affirmative action (Kuklinski et al. 1997)), we have little reason to assume that the LATE estimated for compliers can be readily generalized to other study participants.

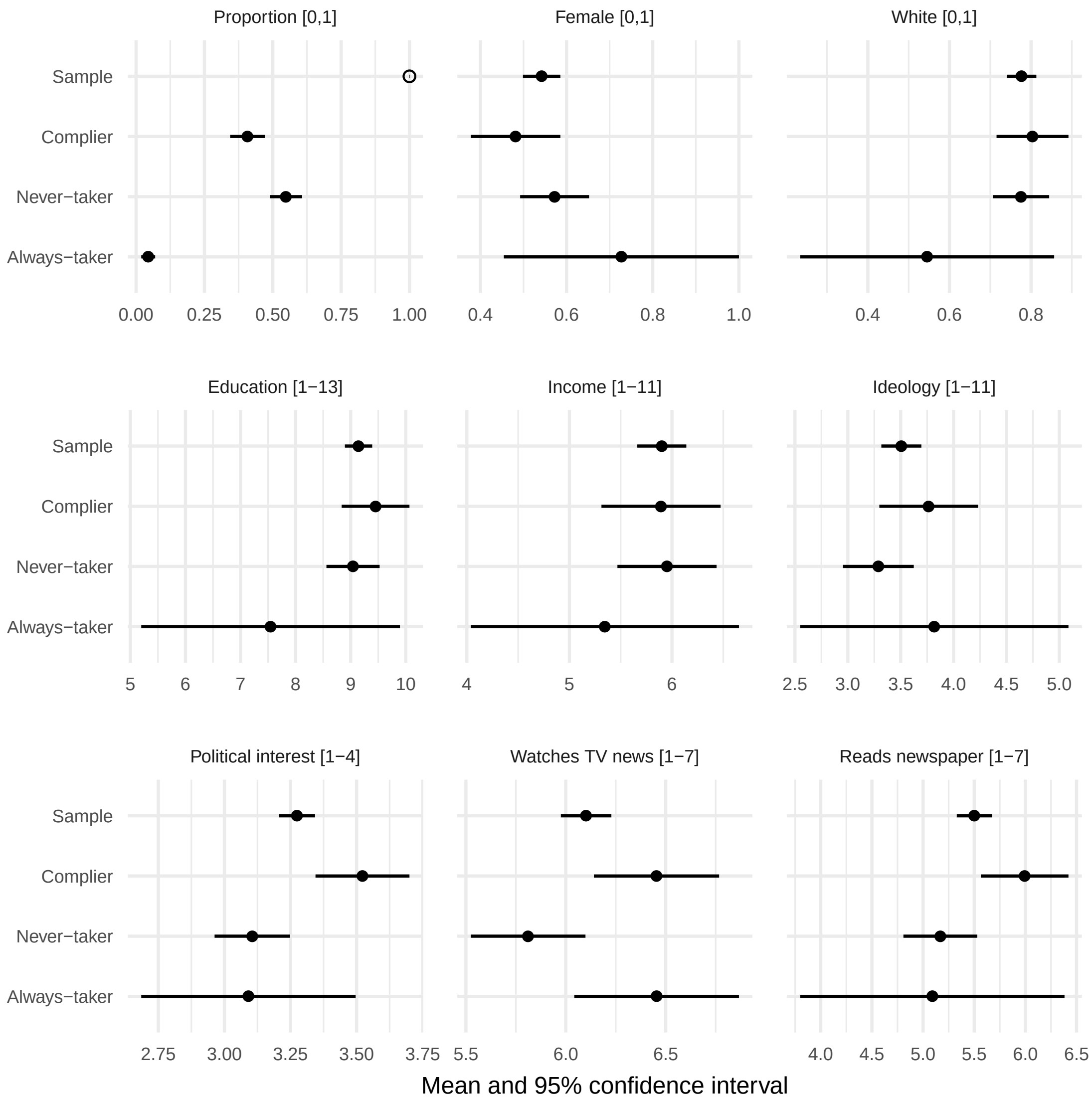

Figure S.3: Descriptive statistics (mean and 95% bootstrap confidence intervals) for the complier and non-complier subpopulations in a field experiment about the effect of watching a Fox News debate on support for affirmative action. While compliers and non-compliers tend to have similar socio-economic backgrounds, compliers are more politically interested and consume more news.

	Support [0,1]	Information [1-4]
Watched Debate [0,1]	−0.066	0.271
	(0.093)	(0.158)
Ideology [1-11]	−0.086	−0.017
	(0.010)	(0.016)
Political interest [1-4]	−0.037	0.252
	(0.030)	(0.047)
Watches TV news [1-7]	0.008	0.000
	(0.014)	(0.024)
Education [1-13]	−0.013	0.003
	(0.008)	(0.013)
Reads newspaper [1-7]	−0.007	0.108
	(0.012)	(0.019)
Female [0,1]	−0.026	−0.053
	(0.041)	(0.069)
Income [1-11]	0.007	−0.009
	(0.008)	(0.013)
White [0,1]	0.183	0.066
	(0.052)	(0.086)
(Intercept)	1.029	1.797
	(0.142)	(0.234)
Num. obs.	441	498

Table S.3: 2SLS estimates with standard errors in parentheses. The estimates are qualitatively similar, but numerically not exactly identical to the published estimates in Table 4, Models 3-4 in Albertson and Lawrence (2009).

Variable	Always-taker	Never-taker	Complier	Sample
Proportion [0,1]	0.04 (0.01)	0.55 (0.03)	0.41 (0.03)	-
Female [0,1]	0.73 (0.14)	0.57 (0.04)	0.48 (0.05)	0.54 (0.02)
White [0,1]	0.55 (0.16)	0.78 (0.03)	0.80 (0.04)	0.78 (0.02)
Education [1-13]	7.55 (1.22)	9.04 (0.25)	9.45 (0.31)	9.14 (0.13)
Income [1-11]	5.35 (0.67)	5.95 (0.25)	5.89 (0.30)	5.90 (0.13)
Ideology [1-11]	3.82 (0.65)	3.29 (0.17)	3.76 (0.24)	3.51 (0.10)
Political interest [1-4]	3.09 (0.21)	3.11 (0.07)	3.52 (0.09)	3.27 (0.04)
Watches TV news [1-7]	6.45 (0.21)	5.81 (0.14)	6.45 (0.16)	6.10 (0.06)
Reads newspaper [1-7]	5.09 (0.67)	5.17 (0.19)	5.99 (0.22)	5.50 (0.09)

Table S.4: Descriptive statistics (mean and 95% bootstrap confidence intervals) for the complier and non-complier subpopulations in the study by Albertson and Lawrence (2009).