Compliance Prob and Response Heterogeneity

1. **The link and the difference between Tbl_Prob_Compliance and Tbl_Response_Hetero**

In Abadie (2003), we know that

Lemma 2.1. *Under Assumption* 2.1,

$$\mathrm{P}(D_1 > D_0|X) = \mathrm{E}[D|Z = 1, X] - \mathrm{E}[D|Z = 0, X] > 0.$$

In the setting of our paper, we get

$$P^{Early}(D_1 > D_0) = E[D|Z^{Early} = 1, Z^{Normal}, X] - E[D|Z^{Early} = 0, Z^{Normal}, X]$$

$$P^{Normal}(D_1 > D_0) = E[D|Z^{Normal} = 1, Z^{Early}, X] - E[D|Z^{Normal} = 0, Z^{Early}, X]$$

And since eligibility for early and normal retirement benefits are mutually exclusive binary instruments, so I turn them into a continuous variable.

According to Kowalski (2021), using the range of the unobserved net cost of retirement we can calculate complier probability and explore selection heterogeneity by comparing the average covariate vectors of always takers, early/normal compliers, and never takers.

In the setting of our paper, we get

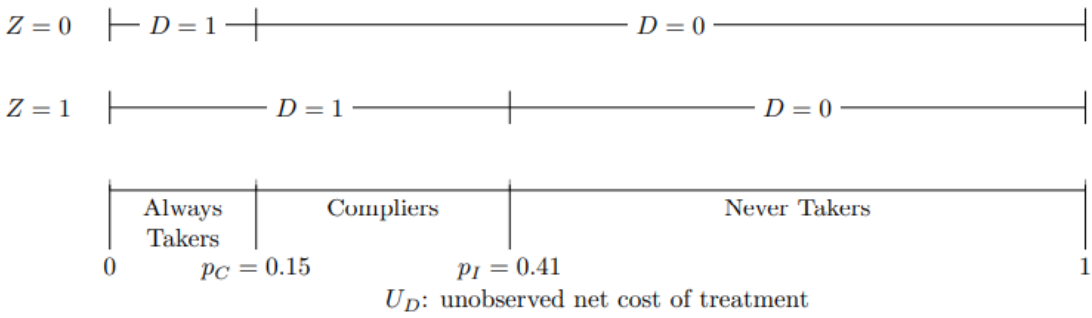
$$P^{Early}(D_1 > D_0) = E[D|Z^{Early} = 1, Z^{Normal} = 0, X] - E[D|Z^{Early} = 0, Z^{Normal} = 0, X]$$

$$P^{Normal}(D_1 > D_0) =$$
$$E[D|Z^{Normal} = 1, Z^{Early} = 0, X] - E[D|Z^{Normal} = 0, Z^{Early} = 1, X]$$

Therefore, while P^{Early} is similar between the two methods, there is a noticeable discrepancy between the two methods when it comes to P^{Normal} .

($E[D|Z^{Normal} = 1, Z^{Early} = 0, X] - E[D|Z^{Normal} = 0, Z^{Early} = 1, X]$ v.s $E[D|Z^{Normal} = 1, Z^{Early}, X] - E[D|Z^{Normal} = 0, Z^{Early}, X]$)

Figure 1: Ranges of U_D Show Ordering from Always Takers to Compliers to Never Takers



Note. Treatment represents enrollment in Medicaid. p_C is the probability of treatment in the control group, and p_I is the probability of treatment in the intervention group.

Similar to the graph above, we can get the probability of treatment (retirement) in the control group ($Z^{normal} = 0, Z^{early} = 0$) , the probability of early treatment in the early intervention group ($Z^{normal} = 0, Z^{early} = 1$), and the probability of normal treatment in the normal intervention group ($Z^{normal} = 1, Z^{early} = 0$)

These probabilities are a set of realizations of the unobserved net cost of treatment and we can use it as weights to calculate the weighted average covariate vectors of always takers, compliers, and never takers.

That is

$$\mathrm{E}[X \mid p_C < U_D \leq p_I] = \mathrm{P}(Z=1)\left[\frac{p_I}{p_I - p_C}\mathrm{E}[X \mid D=1, Z=1] - \frac{p_C}{p_I - p_C}\mathrm{E}[X \mid D=1, Z=0]\right]$$
$$+\mathrm{P}(Z=0)\left[\frac{1-p_C}{p_I - p_C}\mathrm{E}[X \mid D=0, Z=0] - \frac{1-p_I}{p_I - p_C}\mathrm{E}[X \mid D=0, Z=1]\right].$$

I bootstrap for 1000 times to get 1000 realization and calculate se.

If we can get the realization of the probability of always taker using Abadie (2003), then we are able to calculate the weighted average covariate vectors of always takers compliers, and never takers.

⇒ The connection between Tbl_Prob_Compliance and Tbl_Response_Hetero can be described as follows: the average covariate vectors of always takers compliers, and never takers in Tbl_Response_Hetero are calculated utilizing weights based on the probabilities specified in Tbl_Prob_Compliance.

2. **what compliance probability can explain that the first-stage coefficients couldn't.**

$$P^{Early}[D_1 > D_0 | D = 1] = P^{Early}[D_1 > D_0] * P[Z^{Early} = 1]/P[D = 1]$$

$$P^{Normal}[D_1 > D_0 | D = 1] = P^{Normal}[D_1 > D_0] * P[Z^{Normal} = 1]/P[D = 1]$$

$$P^{Early}[D_1 > D_0 | D = 0] = P^{Early}[D_1 > D_0] * P[Z^{Early} = 0]/P[D = 0]$$

$$P^{Normal}[D_1 > D_0 | D = 1] = P^{Normal}[D_1 > D_0] * P[Z^{Normal} = 0]/P[D = 0]$$

But we did not get $P[D_1 = D_0 = 0]$ and $P[D_1 = D_0 = 1]$ yet.

$P^{Normal}[D_1 > D_0]$ and $P^{Early}[D_1 > D_0]$ provide information about the proportion of the population that complies with normal retirement eligibility and early retirement eligibility, respectively.

$P^{Normal}[D_1 > D_0|D = 1]$ and $P^{Early}[D_1 > D_0|D = 1]$ refer to the intervention group and reveal the percentage of retirees who retired due to their eligibility for normal retirement or early retirement benefits.

$P^{Normal}[D_1 > D_0|D = 0]$ and $P^{Early}[D_1 > D_0|D = 0]$ correspond to the control group and indicate the proportion of active workers who would have retired had they been eligible for normal retirement or early retirement benefits but have continued to work.

When $P[Z^{Early} = 1]$ is relatively bigger than $P[D = 1]$, it suggests a higher likelihood for a retiree to be eligible for early retirement benefits, resulting in a greater proportion of retirees complying with early retirement eligibility. It is noteworthy that $P^{Early}[D_1 > D_0]$ must lie between $P^{Early}[D_1 > D_0|D = 1]$ and $P^{Early}[D_1 > D_0|D = 0]$.

3. **Please take this text from the current Data Appendix (see below), update it to reflect the new time frame**

Updated Version:

> For this paper, the sample is restricted to active workers who were between ages 52 and 65 as of April 2016 and who were members of TSERS or LGERS.
> In addition, we exclude individuals who were retired or deceased as of April 2016, as well as those who had less than 5 years of service. Additionally, we further restrict the sample to include only those who were hired before March 2014.
> To create our final analysis sample, we merge the administrative records with responses to a survey of active employees fielded in May 2016.

Restriction list

- Not retired as of April 2016 (Active Benefit Account and Claim Benefit)
- Active Membership
- TSERS or LGERS (Original: TSERS or LGERS but not both)
- Drop law enforcement, firefighters
- Workers aged 52 to 65 or (Original: Take integer of age and restrict workers from age 52 to 64)
- Drop indi. claiming disability or other benefit types(non-regular) as of April 2016
- Drop data with unidentified gender
- Drop individuals deceased as of April 2016
- Drop individuals/memberships with missing contribution dates
- First hired before 3/4/2014
- Drop individuals with last contribution year before 2014 in April 2016 data
- Drop if last contribution date is mssing or is earlier than April 2016 in Admin 5 data
- Drop if Marital status missing
- Drop if health status missing
- Drop if Race missing
- Drop if Agency Classification missing

Blue: Not sure

Orange: Updated Restriction; some changes are made to the code; resulting different sample size.

Green: Updated Restriction; some changes are made to the code; resulting same sample size.

Updated Appendix Table A1 is in output folder in RA.

Question about time:

The fielding date for Survey 3 is recorded as 5/10, but some notations in the original table seems to refer to the survey date as April 2016 instead of May 2016.

```
Survey 3 timeline:

  Sun 5/10 First day of fielding survey
  Tue 5/17 Reminder sent to benefit claimants, with date of first drawing
  Wed 5/18 Reminder sent to actives, with date of first drawing
  Mon 5/23 First iPad drawing
  Tue 5/31 Reminders sent to all groups, announcing iPad winners
  Tue 6/8 Reminder sent to all groups, with date of second drawing
  Fri 6/10 Second iPad drawing
```