

pset2_steiner

erika steiner

2025-04-16

In this problem set, we continue analyzing the FTP program, but we reinterpret treatment status. In FTP, as in other programs, there was confusion among respondents about whether they were subject to a time limit. For the purposes of this problem set, we will consider respondents who said that they were subject to a time limit as treated, and those who said they were not as controls. Our goal will be to estimate the effect of a perceived time limit on welfare receipt, employment, and the like. The idea is that you cannot be influenced by a time limit if you don't believe it applies to you, whether or not it applies to you in fact. All data and documentation can be found on Canvas under Problem Set 1.

- (1) *The question about the time limit comes from the four-year survey, the results of which are contained in the file ftp_srv.dta. Start by merging this file to ftp_ar.dta. How many of the original sample members were interviewed for the survey?*

```
# import admin data
ftp_ar <- read_dta("ftp_ar.dta")
# import survey data
ftp_srv <- read_dta("ftp_srv.dta")

# this ftp_ar.dta file already has all survey questions imported
# you can confirm this by running the following:
# setdiff(colnames(ftp_srv), colnames(ftp_ar))

ftp_merged <- ftp_ar %>%
# filter out ftp_ar participants who did not participate in the survey
# using sampleid per data documentation page 13
  filter(.$sampleid %in% (ftp_srv$sampleid))

# count number of sample members interviewed for survey
nrow(ftp_merged)
```

```
## [1] 1729
```

```
# page 8 documentation: Only the 2,160 sample members who entered the study between August 1994 and Feb
```

- (2) *Tabulate the variable fmi2.*

```
# fmi2: "Is/was there a time limit?"

# create a table of fmi2 responses
ftp_merged %>%
  # tabulate
  count(fmi2) %>%
  # per survey page 48, question I2
  # "1" = "Yes". "2" = "No", "8" = "Don't Know"
  mutate(
    fmi2 = case_when(
```

```

fmi2 == 1 ~ "Yes",
fmi2 == 2 ~ "No",
fmi2 == 8 ~ "I don't know",
# Edit from chatgpt - did not know how to re-format NA counts
# Prompt: how to handle "NA" format in mutate case when argument r
# Response: Check for NA explicitly: In case_when(), you need to handle NA values explicitly by u
is.na(fmi2) ~ "Missing"
)) %>%
# make table
kable(col.names = c("Response", "Frequency"), caption = "(Is/Was) there a time limit on how long you

```

Table 1: (Is/Was) there a time limit on how long you (are/were) allowed to receive cash assistance?

Response	Frequency
Yes	666
No	365
I don't know	118
Missing	580

How many of the survey respondents have valid (i.e., non-missing) responses?

1149

How many thought they were subject to a time limit?

666

How many thought otherwise?

365

How many were unsure?

118

Finally, what explains why the number of valid responses is less than the number of people interviewed for the survey? (Hint: examine variables fmi1 and fmi1a.)

```

# fmi1: currently receiving cash assistance
# fmi1a: have received cash assistance

# fmi1 table
ftp_merged %>%
# tabulate
count(fmi1) %>%
# per survey page 48, question I1
# "1" = "Yes". "2" = "No", "8" = "Don't Know"
mutate(
  fmi1 = case_when(
    fmi1 == 1 ~ "Yes",
    fmi1 == 2 ~ "No",
    fmi1 == 8 ~ "I don't know",
    is.na(fmi1) ~ "Missing"
  )) %>%
# make table
kable(col.names = c("Response", "Frequency"), caption = "Are you currently receiving cash assistance

```

Table 2: Are you currently receiving cash assistance through AFDC, the WAGES program, or TANF?

Response	Frequency
Yes	230
No	1499

```
# fmila table
ftp_merged %>%
  # tabulate
  count(fmila) %>%
  # per survey page 48, question I1a
  # "1" = "Yes". "2" = "No", "8" = "Don't Know"
  mutate(
    fmila = case_when(
      fmila == 1 ~ "Yes",
      fmila == 2 ~ "No",
      fmila == 8 ~ "I don't know",
      is.na(fmila) ~ "Missing"
    ) %>%
  # make table
  kable(col.names = c("Response", "Frequency"), caption = "Have you ever received cash assistance since
```

Table 3: Have you ever received cash assistance since (RAD)?

Response	Frequency
Yes	919
No	565
I don't know	15
Missing	230

Only 230 people reported actively receiving cash assistance. Of the ~1500 people who did not report actively receiving cash assistance, 565 people said they hadn't received any cash assistance since randomization (which could have occurred up to two years prior TK). Thus, many respondents may not have interacted with cash assistance programs recently and would not how to respond to this question.

- (3) Define a new treatment dummy that is equal to one for people who believed they were subject to a time limit and equal to zero for those who did not or were not sure. Call it *TLyes*. Cross-tabulate *TLyes* and the experimental dummy (“e”).

```
ftp_merged %<>%
  # add new treatment dummy TLyes
  mutate(TLyes =
    case_when(
      # 1 if believed subject to time limit
      fmi2 == 1 ~ 1,
      # 0 if did not or were not sure if subject to time limit
      fmi2 == 2 ~ 0,
      fmi2 == 8 ~ 0
    )
  ) %>%
  # remove rows where TLyes = NA
```

```

filter(!is.na(TLyes))

# cross-tabulate TLyes and e
ftp_merged %>%
  # tabulate
  count(TLyes, e) %>%
  # enhance readability of TLyes
  mutate(
    TLyes = case_when(
      TLyes == 1 ~ "Answered Yes TL",
      TLyes == 0 ~ "Answered No TL",
    ) %>%
  # enhance readability of e
  mutate(
    e = case_when(
      e == 1 ~ "e Treatment",
      e == 0 ~ "e Control"
    )
  ) %>%
  # pivot into crosstab
  pivot_wider(names_from = e, values_from = n) %>%
  # make table
  kable(caption = "Cross-tabulating TLyes and e")

```

Table 4: Cross-tabulating TLyes and e

TLyes	e Control	e Treatment
Answered No TL	369	114
Answered Yes TL	205	461

Discuss confusion about the time limit.

Approximately 25% of participants who were subject to time limits incorrectly believed they were not subject to time limits. Similarly, 31% of participants who were not subject to time limits incorrectly believed they were subject to time limits.

The inaccuracy of these beliefs is very relevant to this study, which seeks to understand how imposed time limits change the behavior of welfare recipients. If participants have inaccurate information about these time limits, their behaviors may differ from what would be expected if they had a more accurate understanding. Conclusions drawn from this study should thus consider how this confusion may confound our understanding of results.

- (4) *Using TLyes as your treatment indicator, estimate by OLS the effect of the time limit on the number of months of welfare receipt during years 1-2 post_RA. To do this, estimate a regression of your dependent variable on TLyes and the set of control variables used by MDRC in their publications (the ones with the string "cova:" in their labels).*

```

# create variable which equals number of months on welfare receipt during years 1-2 post_RA

ftp_merged %<>%
  # variable = number of months rec welfare y1-2 (nmrecyr12)
  mutate(nmrecyr12 =

```

```

      # number of months y1 + number of months y2
      krecc2t5 + krecc6t9)

# identify covas columns
# pulled from pset1
covas <- ftp_merged %>%
  # select
  select(
    # columns where...
    where(~ {
      # the column's label
      column_label <- attributes(.)$label
      # is not null AND contains 'cova' (returns TRUE)
      !is.null(column_label) && str_detect(column_label, 'cova')
    }) %>%
  colnames()

# write formula for nmrecyr12 regressed on TLyes and control variables (covas)
formula <- as.formula(paste(
  # nmrecyr12 regressed on TLyes and...
  "nmrecyr12 ~ TLyes+",
  # list covas as string, separated by +
  paste(covas, collapse = " + ")))

model <- lm(formula, data = ftp_merged)

# summary(model)

# regress nmrecyr12 on TLyes and set of control variables

stargazer(model, title = "Results")

```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Wed, Apr 16, 2025 - 16:56:21

```
# summary(lm(formula, data = ftp_merged))
```

Do you believe these regressions consistently estimate the effect of the time limit? Explain.

- (5) Provide conditions under which the experimental dummy would be a valid instrument for the time limit.
- (6) Estimate the first-stage regression of TLyes on e. Do you have a weak-instrument problem?
- (7) Use the experimental dummy as an instrument to estimate the effect of the time limit on the number of months of welfare receipt during years 1-2 post-RA. How do the welfare receipt results compare to the OLS estimates from question (4)? Can you summarize the direction of bias associated with the OLS estimates?
- (8) Explain why you would expect the exclusion condition to fail for the above regression.
- (9) Now estimate a model with age-group interactions, along the lines of what you did in PS 1. Define the same four age groups, and estimate the same model, but interact the age-group dummies with TLyes rather than e. Estimate the model by OLS. Omit the control variables.
- (10) Now estimate the same model, but use interactions between e and the age-group dummies as instruments for the interactions between TLyes and the age-group dummies. Comment on how estimating the model with age-group interactions helps to satisfy the exclusion restriction.
- (11) Discuss your estimated effects of time limits and how they compare with the estimates from question

Table 5: Results

	<i>Dependent variable:</i>	
	nmrecyr12	
TLyes	0.389	(0.438)
yr2adc	0.001***	(0.0003)
yradc	-0.0001	(0.001)
yr2rec	2.141	(1.537)
yrrec	1.631	(1.595)
yrkrec	0.190	(0.149)
yrearn	-0.0001	(0.0003)
yr2earn	-0.0001	(0.0001)
yrearnsq	0.000	(0.000)
yrem	-0.785	(1.061)
yr2emp	-0.118	(0.846)
yrkemp	-0.006	(0.611)
yr2kemp	0.040	(0.332)
yr2fs	-0.0002	(0.0003)
yrfs	-0.0002	(0.001)
yr2rfs	-1.071	(1.753)
yrdfs	-0.454	(1.830)
yrkrfs	0.162	(0.137)
Constant	8.190***	

(9). Can you diagnose why the two sets of coefficients differ?