HCMG 901 Problem Set 2

Andrés Rovira

March 4, 2023

1 Theory Portion

Part a

Comment on 2SLS bias in finite sample and its determinants. How does the first stage predictive strength of the instrument affect bias and consistency of the IV estimator?

The 2SLS estimator is biased in finite samples. As discussed in the slides, Hahn and Hausman (2005) show that:

$$E[\hat{\beta}_{2SLS} - \beta] \approx \frac{\sigma_{\eta\xi}}{\sigma_s^2} \frac{K}{N R_{sZ}^2}$$

So, this bias increases with the covariance between the error terms of the two stages $\sigma_{\eta\xi}$, decreases with the variance of the endogenous regressor σ_s^2, increases with the number of instruments K, decreases with the sample size N, and decreases with the R-squared from the first stage model R_{sZ}^2 (the predictive strength of the instrument).

Assuming the exclusion restriction holds, the 2SLS estimator will be consistent in finite samples regardless of the predictive strength of the instrument.

Part b

What assumptions are required to estimate the LATE? Prove the LATE theorem under these maintained assumptions.

The assumptions required to estimate the LATE are:

1. SUTVA: potential outcomes are unrelated to the status of other individuals, so $D_i(Z) = D_i(Z_i)$ and $Y_i(D, Z) = Y_i(D_i, Z_i)$
2. Random assignment: $\{Y_i(d, z)\ \forall d, z\}, D_{1i}, D_{0i} \perp Z_i$
3. Exclusion: Z only affects Y through its effect on D, so $Y_i(1, 0) = Y_i(1, 1) = Y_{1i}$ and $Y_i(0, 0) = Y_i(0, 1) = Y_{0i}$
4. First stage/relevance: Z affects the probability of treatment for some units, so $E[D_{1i} - D_{0i}] \neq 0$
5. Monotonicity: Z makes all individuals weakly more or less likely to be treated, so either $D_{1i} \geq D_{0i}\ \forall i$ or vice-versa

The LATE estimator then estimates:

$$\beta_{LATE} = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]}$$

For the denominator, we have:

$$\begin{aligned} E[D_i|Z_i = 1] - E[D_i|Z_i = 0] &= E[D_{1i}|Z_i = 1] - E[D_{0i}|Z_i = 0] \\ &= E[D_{1i} - D_{0i}] \text{ by random assignment} \\ &= P(D_{1i} > D_{0i}) \text{ by monotonicity} \\ &\neq 0 \text{ by the first stage condition} \end{aligned}$$

For the numerator, we have:

$$\begin{aligned} E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] &= E[Y_{0i} + (Y_{1i} - Y_{0i})D_i|Z_i = 1] - E[Y_{0i} + (Y_{1i} - Y_{0i})D_i|Z_i = 0] \\ &= E[Y_{0i} + (Y_{1i} - Y_{0i})D_{1i}] - E[Y_{0i} + (Y_{1i} - Y_{0i})D_{0i}] \text{ by exclusion} \\ &= E[(Y_{1i} - Y_{0i})(D_{1i} - D_{0i})] \\ &= E[(Y_{1i} - Y_{0i})|D_{1i} > D_{0i}]P(D_{1i} > D_{0i}) \text{ by monotonicity} \end{aligned}$$

Therefore:

$$\begin{aligned} \beta_{LATE} &= \frac{E[(Y_{1i} - Y_{0i})|D_{1i} > D_{0i}]P(D_{1i} > D_{0i})}{P(D_{1i} > D_{0i})} \\ &= E[Y_{1i} - Y_{0i}|D_{1i} > D_{0i}] \\ &= E[Y_{1i} - Y_{0i}|\text{compliers}] \end{aligned}$$

and this proves the LATE theorem.

Part c

How does the monotonicity assumption help us meaningfully interpret the IV estimator? Discuss the benefit for both the first stage and reduced form estimates.

As the slides show, without monotonicity, the first stage estimates the following:

$$E[D_i|Z_i = 1] - E[D_i|Z_i = 0] = P(D_{0i} = 0, D_{1i} = 1) - P(D_{0i} = 1, D_{1i} = 0) = P(\text{compliers}) - P(\text{defiers})$$

And the reduced form estimates the following:

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] = E[Y_{1i} - Y_{0i}|\text{compliers}]P(\text{compliers}) - E[Y_{1i} - Y_{0i}|\text{defiers}]P(\text{defiers})$$

Assuming monotonicity means that $P(\text{defiers}) = 0$, which does away with the second term in each case. This allows us to interpret the first stage as the probability of compliers, and the reduced form is no longer a weighted average of the ATEs for compliers and defiers. Then as the LATE theorem shows, we can also interpret the IV estimate as the ATE for compliers.

Part d

Show that for instruments that generate no always takers, LATE=ATT.

The ATT can be split apart as follows:

$$E[Y_{1i}-Y_{0i}|D_i=1] = E[Y_{1i}-Y_{0i}|D_{0i}=1]P(D_{0i}=1|D_i=1)+E[Y_{1i}-Y_{0i}|D_{1i}>D_{0i}]P(D_{1i}>D_{0i}, Z_i=1|D_i=1)$$

But $P(D_{0i}=1|D_i=1)=0$ since it is the share of always takers among treated individuals, and there are no always takers. And $P(D_{1i}>D_{0i}, Z_i=1|D_i=1)=1$ since it is the share of compliers among treated individuals exposed to the instrument, which is everyone since there are no always takers.

Thus, $ATT = E[Y_{1i}-Y_{0i}|D_{1i}>D_{0i}] = LATE$

Part e

What are the benefits of estimating marginal treatment effects? What additional functional form restrictions do researchers usually make to estimate the MTE? How do they help? Discuss the steps to obtain the MTE.

Estimating the MTE is helpful in the context of "essential heterogeneity", where treatment probability for compliers can be correlated with treatment effects. This arises in cases where, for example, individuals who anticipate better treatment effects may select into treatment at higher rates. But estimating the MTE can allow us to still recover the ATE and ATT under some additional assumptions.

First we assume the following about potential outcomes:

$$Y_{ji} = \mu_j(X_i) + U_{ji} \text{ for } j=0,1$$

$$E[Y_{ji}|X=x] = \mu_j(x)$$

$$E[U_{ji}|X_i] = 0 \ \forall j,i$$

Then the additional function form restrictions researchers make to estimate the MTE are as follows. Define a propensity score as:

$$P(D=1|Z=z, X=x) = P(z,x)$$

Assume latent variable D_i^* such that $D_i = 1(D_i^*>0)$ and $D_i^* = \mu_D(Z_i, X_i) - V_i$ where the "unobserved resistance to treatment" V has mean 0.

We can then express $P(z,x)$ as a function of the distribution of V: $P(z,x) = F_V(\mu_D(z,x))$

Then $D_i = 1 \iff U_D < P(z,x)$ where $U_D \sim U[0,1]$ is the CDF of V, and represents the quantile of unobserved resistance. So marginal individuals have $U_D = P(z,x) \equiv p$ and are indifferent to treatment. Then the MTE averages treatment effects across these marginal individuals:

$$MTE(x,z) = E[Y_{1i}-Y_{0i}|X_i=x, Z_i=z, U_D=p]$$

These assumptions help by allowing us to estimate MTE as LATE computed over a small change in propensity score. We take a derivative to find the limit of LATE at p:

$$MTE(X=x, U_D=p) = \frac{\partial E[Y|X=x, P(z,x)=p]}{\partial p}$$

In practice, researchers typically need to make some additional functional form restrictions, assuming they don't have overlap in every $X=x$ cell. We assume Y depends on X, and D depends on X and

Z in linear additive forms ($Y_{ji} = X_i'\beta_j + U_{ji}$). Then we can write the MTE as the sum of observed and unobserved components which are additively separable:

$$MTE(X = x, U_D = p) = x'(\beta_1 - \beta_0) + E[U_{1i} - U_{0i}|p]$$

We approximate the unobserved portion with a polynomial $K(p)$:

$$E[Y_i|X = x, P = p] = x'\beta_0 + x'(\beta_1 - \beta_0)p + K(p)$$

Taking the derivative wrt p we estimate the MTE as:

$$MTE(X = x, U_D = p) = x'(\beta_1 - \beta_0) + K'(p)$$

The steps to obtain the MTE are as follows:

1. Estimate the propensity score $P(x, z) = P(D = 1|X = x, Z = z)$ by a binary dependent variable model like logit, probit, or the linear probability model
2. Regress Y on X, p, interaction of X and p, and $K(p)$ as a polynomial in p.
3. Estimate the unobserved component of the MTE at each value of p by getting the derivative of $K(p)$ on a grid of values in p
4. Compute $MTE(x, U_D = p) = x'(\beta_1 - \beta_0) + \frac{\partial K(p)}{\partial p}$
5. Obtain standard errors by bootstrapping this procedure

2 Paper Replication Portion

Part a

We have explicit randomization in this setting through the lottery. Present a balance table of individual characteristics between lottery winners and losers and comment on whether you think the randomization appears to be well done. Estimate the intent-to-treat effect of winning the experiment lottery on the likelihood of using different type of health care (extensive margin only). Ensure that you use sampling weights if there are any (follow the paper). Interpret the estimates. Are these estimates policy relevant?

Table 1 presents a balance table between lottery winners and losers. It is modeled after Table A13 in Finkelstein et al. (2012), so for the "Difference" column, the regression also includes household size and survey wave dummmies and their interactions. The balance is pretty convincing, with no significant differences between the groups, so I would say that the randomization appears to be well done.

Table 1: Balance Table of Lottery Winners and Losers

	Control Mean (SD)	Treatment Mean (SD)	Difference
Birth year	1,966.24	1,966.22	–0.0655
	(12.15)	(12.15)	(0.1911)
Female	0.59	0.58	–0.0042
	(0.49)	(0.49)	(0.0068)
English materials requested	0.92	0.91	–0.0003
	(0.28)	(0.29)	(0.0048)
Signed self up	0.88	0.84	–0.0016
	(0.32)	(0.37)	(0.0027)
Signed up first day	0.10	0.11	0.0061
	(0.30)	(0.31)	(0.0049)
Gave phone number	0.91	0.92	0.0059
	(0.28)	(0.27)	(0.0037)
Gave PO box as address	0.13	0.13	–0.0023
	(0.33)	(0.33)	(0.0053)
ZIP is an MSA	0.75	0.75	0.0011
	(0.43)	(0.43)	(0.0070)

Table 2 presents estimates of the intent-to-treat (ITT) effects. Winning the lottery results in a 2.42 percentage point (pp) increase in people taking any prescription medications, and 6.17, 0.65, and 0.22 pp increases in people having any primary care, ER or hospital visits, respectively. These estimates are policy relevant since many policies offer but do not mandate certain services, and these estimates represent the effects of offering Medicaid access broadly to the population, where some would accept and some would not.

Table 2: ITT effects of winning lottery on types of health care usage (extensive margin)

	ITT
Currently taking any prescription medications	0.0252
	(0.0083)
Any primary care visits	0.0617
	(0.0074)
Any ER visits	0.0065
	(0.0067)
Any hospital visits	0.0022
	(0.0040)

Part b

Discuss potential identification concerns with using OLS to estimate the effects of Medicaid coverage on healthcare utilization. Present some evidence on the balance between those who received Medicaid (treated) and the remaining people (controls). Discuss your concerns building on this evidence.

When simply using OLS to estimate the effects of Medicaid coverage, I am very concerned about selection bias. Perhaps those who receive Medicaid are those who are sicker and need it more, so they were more motivated to go the work of applying for Medicaid, but are also more likely to use health care services. Table 3 is a balance table of individual characteristics where "Treatment" is now whether or not the person receives Medicaid. This table is not horribly unbalanced, but is notably less balanced than Table 1. People receiving Medicaid are significantly more likely to be female, request English materials, sign themselves up, sign up on the first day, and give their phone number. Many of these characteristics indeed seem to indicate a slightly more motivated group of people, perhaps because they are sicker, and thus more likely to use health care services.

Table 3: Balance Table by Receipt of Medicaid

	Control Mean (SD)	Treatment Mean (SD)	Difference
Birth year	1,966.21	1,966.30	0.0891
	(12.21)	(11.89)	(0.2224)
Female	0.57	0.63	0.0514
	(0.49)	(0.48)	(0.0084)
English materials requested	0.91	0.93	0.0236
	(0.29)	(0.25)	(0.0051)
Signed self up	0.85	0.89	0.0381
	(0.35)	(0.31)	(0.0055)
Signed up first day	0.10	0.12	0.0225
	(0.30)	(0.33)	(0.0060)
Gave phone number	0.91	0.92	0.0086
	(0.28)	(0.27)	(0.0042)
Gave PO box as address	0.13	0.12	–0.0081
	(0.34)	(0.33)	(0.0062)
ZIP is an MSA	0.75	0.75	0.0073
	(0.43)	(0.43)	(0.0081)

Part c

Regress utilization outcomes from Table 5 on the indicators for Medicaid coverage and obtain OLS estimates. There are several different measures of coverage (see Table 3), use any two that you like. What treatment effect do these coefficients estimate in the context of the potential outcome framework?

See Table 4 for the OLS estimates. As indicators for Medicaid coverage, I chose "Currently on Medicaid" and "Ever on Medicaid". These coefficients estimate the ATE, but due to the problems highlighted in part b, they are biased.

Table 4: OLS Estimates of Utilization Outcomes on Medicaid Coverage Indicators

	Currently on Medicaid	Ever on Medicaid
Any ER visits	0.1112	0.0913
	(0.0086)	(0.0075)
Any hospital visits	0.0650	0.0515
	(0.0057)	(0.0048)
Any primary care visits	0.2722	0.1985
	(0.0076)	(0.0075)
Currently taking any prescription medications	0.1864	0.1316
	(0.0088)	(0.0085)

Part d

What would be the arguments for monotonicity and exclusion in this setting? Comment on whether you find them reasonable.

Monotonicity here is the assumption that exposure to the instrument must (weakly) increase the probability of treatment for all individuals. The argument for this is that winning the lottery presents an appealing opportunity for many people to get Medicaid, and may cause many to take it up. But there is no reason to suspect that someone would seek Medicaid if they lost the lottery, but not if they won - that would certainly be against their interests. This seems like a reasonable argument to me.

The exclusion restriction here means that the winning the lottery must only affect health care usage outcomes through its affect on Medicaid takeup. The argument for this is that this setting offers a truly randomized instrument, which is mechanically only linked to Medicaid takeup and nothing else. Again, this seems rather reasonable.

Part e

Replicate the first stage (Table 3 Columns 5 & 6) for the two measures of Medicaid coverage you used in part (c) above.

See Table 5.

Table 5: First-Stage Estimates

	Control mean	Estimated FS
Ever on Medicaid	0.1350	0.2902
		(0.0066)
Currently on Medicaid	0.1047	0.1906
		(0.0060)

Part f

How are compliers defined in this setting? If we wanted to characterize the size and composition of the complier group, how would we do it? Use some of the background information available about the people (e.g., gender, age, whether they signed up for the lottery themselves) to characterize compliers. What is the size of the complier group?

Compliers in this setting are those who receive Medicaid by winning the lottery, but would not otherwise have received Medicaid. To characterize the size of the complier group, we can difference the means of the Medicaid variables for lottery winners and losers. This is equivalent to the first-stage estimates, without using the additional controls for household size and survey wave dummmies and their interactions, which I present in the first row of Table 6. Compliers are 19% and 29% of the sample for "Currently on Medicaid" and "Ever on Medicaid", respectively.

To characterize the composition of the complier group by some indicator variables of background information, we can look at the following ratio:

$$\frac{P[X_i = 1|D_{1i} > D_{0i}]}{P[X_i = 1]} = \frac{P[D_i|Z_i = 1, X_i = 1] - P[D_i|Z_i = 0, X_i = 1]}{P[D_{1i} > D_{0i}]}$$

Equivalent to the first stage coefficient for those with $X_i = 1$ divided by the first stage coefficient for the full sample. See the rest of Table 6 for estimates of these ratios. These tell us that, when compared to the full sample, compliers are older, less often female, request English materials more, sign themselves up more, sign up on the first day noticeably more often, give their phone numbers at about the same rate, gave PO boxes as addresses slightly more often, and were in MSAs at about the same rate.

Table 6: Shares of Compliers and Ratios of Covariates Among Compliers vs Full Sample

	Currently on Medicaid	Ever on Medicaid
Share of compliers	0.19	0.29
Born before 1968	1.12	1.03
Female	0.98	0.98
English materials requested	1.04	1.04
Signed self up	1.09	1.06
Signed up first day	1.21	1.18
Gave phone number	1.00	1.01
Gave PO box as address	1.02	1.03
ZIP is an MSA	0.99	0.99

Part g

Follow the approach used by the authors to instrument for Medicaid coverage using the offer of Medicaid through the lottery. Generate 2SLS estimates for extensive margin health care utilization outcomes in Table 5. Interpret the 2SLS estimate in the context of heterogeneous treatment effects. Can we interpret the LATE as the ATT?

See Table 7 for 2SLS estimates. These are estimates of the LATE, which is a ATE estimate "local" to the compliers, and not necessarily representative of other groups. When there are heterogeneous treatment effects, we do not expect compliers to be representative of either the whole sample or even the treated group (all Medicaid recipients). Thus we cannot safely interpret the LATE as the ATE or the ATT.

Table 7: 2SLS Estimates of LATE

	LATE
Currently taking any prescription medications	0.0878
	(0.0288)
Any primary care visits	0.2124
	(0.0252)
Any ER visits	0.0223
	(0.0231)
Any hospital visits	0.0077
	(0.0136)

Stata Code

```
/*******************************************************************************
AUTHORS: Andres Rovira and Fabio Schanaider
CREATED: 2023-02-24
PURPOSE: Solve coding portions of HCMG 901 Problem Set 2
*******************************************************************************/

global home "~"
if "$S_OS" == "Windows" global home "`:env USERPROFILE'"
global code "$home/Dropbox (Penn)/Classes/2_health_applied_metrics/ps2/code"
global ol "$home/Dropbox (Penn)/Apps/Overleaf/hcmg901_ps2"
cd "$code"

// ssc install texsave


*******************************************************************************
**# Replicate analysis data from Finkelstein et al. (2012) using their code
*******************************************************************************

// Downloaded data from "https://data.nber.org/oregon/4.data.html"
// Unzipped into the input folder

global repl_code "../input/OHIE_Public_Use_Files/OHIE_QJE_Replication_Code"
global repl_data "$repl_code/Data"

// Manual step: As instructed in "oregon_hie_qje_replication.do" copy data files
//        from input/OHIE_Public_Use_Files/OHIE_Data to a Data folder under $repl_code

// Run their data prep code
cd "$repl_code"
do "SubPrograms/prepare_data.do"
cd "$code"

// Relabel some variables for tables created later
use "$repl_data/data_for_analysis.dta", clear
label var birthyear_list  "Birth year"
label var female_list     "Female"
label var english_list    "English materials requested"
label var self_list       "Signed self up"
label var first_day_list  "Signed up first day"
label var have_phone_list "Gave phone number"
label var pobox_list      "Gave PO box as address"
label var zip_msa         "ZIP is an MSA"
save "$repl_data/data_for_analysis.dta", replace

// Define variable lists used by paper
global baseline_list birthyear_list female_list english_list self_list ///
                     first_day_list have_phone_list pobox_list zip_msa
global survey_useext_list rx_any_12m doc_any_12m er_any_12m hosp_any_12m
```

```
global mdcd_covg_vars ohp_all_ever_survey ohp_all_at_12m


*******************************************************************************
**# Part a, balance table by lottery outcome
*******************************************************************************

use "$repl_data/data_for_analysis.dta", clear

tempname tabpost
tempfile tabfile
postfile `tabpost' str100 varname sd_ind control treatment diff using `tabfile'
foreach var of varlist $baseline_list {
        regress `var' if treatment == 0 & returned_12m == 1 [pweight = weight_12m]
                local c_mean = _b[_cons]
                local c_sd   = e(rmse)
        regress `var' if treatment == 1 & returned_12m == 1 [pweight = weight_12m]
                local t_mean = _b[_cons]
                local t_sd   = e(rmse)
        regress `var' treatment ddd* if returned_12m == 1 [pweight = weight_12m], vce(cluster househo
                local d_mean = _b[treatment]
                local d_se   = _se[treatment]
        local lbl : var label `var'
        post `tabpost' ("`lbl'") (0) (`c_mean') (`t_mean') (`d_mean')
        post `tabpost' ("`lbl'") (1) (`c_sd')   (`t_sd')   (`d_se')
}
postclose `tabpost'
use `tabfile', clear

format control treatment %13.2fc
format diff %13.4fc
foreach var of varlist control treatment diff {
        tostring `var', replace force usedisplayformat
        replace `var' = "(" + `var' + ")" if sd_ind == 1 & !missing(`var')
}
replace varname = "" if sd_ind == 1
drop sd_ind
label var control   "Control Mean (SD)"
label var treatment "Treatment Mean (SD)"
label var diff      "Difference"
texsave * using "$ol/tab_a_balance.tex", varlabels frag replace location("H") ///
        title("Balance Table of Lottery Winners and Losers")


*******************************************************************************
**# Part a, ITT lottery effect on extensive margin of healthcare types
*******************************************************************************

use "$repl_data/data_for_analysis.dta", clear

tempname tabpost
```

```
tempfile tabfile
postfile ‘tabpost’ str100 varname sd_ind itt using ‘tabfile’
foreach var of varlist $survey_useext_list {
        local lbl : var label ‘var’
        regress ‘var’ treatment ddd* if sample_12m_resp == 1 [pweight = weight_12m], vce(cluster hous
        post ‘tabpost’ ("‘lbl’") (0) (_b[treatment])
        post ‘tabpost’ ("‘lbl’") (1) (_se[treatment])
}
postclose ‘tabpost’
use ‘tabfile’, clear

format itt %13.4fc
tostring itt, replace force usedisplayformat
replace itt = "(" + itt + ")" if sd_ind == 1 & !missing(itt)
replace varname = "" if sd_ind == 1
drop sd_ind
label var itt "ITT"
texsave * using "$ol/tab_a_itt.tex", varlabels frag replace location("H") ///
        title("ITT effects of winning lottery on types of health care usage (extensive margin)")


*******************************************************************************
**# Part b, balance table by medicaid receipt
*******************************************************************************

use "$repl_data/data_for_analysis.dta", clear

tempname tabpost
tempfile tabfile
postfile ‘tabpost’ str100 varname sd_ind control treatment diff using ‘tabfile’
foreach var of varlist $baseline_list {
        regress ‘var’ if ohp_all_at_12m == 0 & returned_12m == 1 [pweight = weight_12m]
                local c_mean = _b[_cons]
                local c_sd   = e(rmse)
        regress ‘var’ if ohp_all_at_12m == 1 & returned_12m == 1 [pweight = weight_12m]
                local t_mean = _b[_cons]
                local t_sd   = e(rmse)
        regress ‘var’ ohp_all_at_12m if returned_12m == 1 [pweight = weight_12m], vce(cluster househo
                local d_mean = _b[ohp_all_at_12m]
                local d_se   = _se[ohp_all_at_12m]
        local lbl : var label ‘var’
        post ‘tabpost’ ("‘lbl’") (0) (‘c_mean’) (‘t_mean’) (‘d_mean’)
        post ‘tabpost’ ("‘lbl’") (1) (‘c_sd’)   (‘t_sd’)   (‘d_se’)
}
postclose ‘tabpost’
use ‘tabfile’, clear

format control treatment %13.2fc
format diff %13.4fc
foreach var of varlist control treatment diff {
        tostring ‘var’, replace force usedisplayformat
```

```
        replace `var' = "(" + `var' + ")" if sd_ind == 1 & !missing(`var')
}
replace varname = "" if sd_ind == 1
drop sd_ind
label var control   "Control Mean (SD)"
label var treatment "Treatment Mean (SD)"
label var diff      "Difference"
texsave * using "$ol/tab_b_balance.tex", varlabels frag replace location("H") ///
        title("Balance Table by Receipt of Medicaid")


*******************************************************************************
**# Part c, OLS of utilization on medicaid coverage
*******************************************************************************

use "$repl_data/data_for_analysis.dta", clear

tempname tabpost
tempfile tabfile
postfile `tabpost' str100 util_var sd_ind str100 mdcd_var est using `tabfile'
foreach yvar of varlist $survey_useext_list {
        local lbl : var label `yvar'
        foreach xvar of varlist $mdcd_covg_vars {
                regress `yvar' `xvar' if sample_12m_resp == 1 [pweight = weight_12m], vce(cluster hou
                post `tabpost' ("`lbl'") (0) ("`xvar'") (_b[`xvar'])
                post `tabpost' ("`lbl'") (1) ("`xvar'") (_se[`xvar'])
        }
}
postclose `tabpost'
use `tabfile', clear

reshape wide est, i(util_var sd_ind) j(mdcd_var) string
rename est* *
foreach var of varlist $mdcd_covg_vars {
        format `var' %13.4fc
        tostring `var', replace force usedisplayformat
        replace `var' = "(" + `var' + ")" if sd_ind == 1 & !missing(`var')
}
replace util_var = "" if sd_ind == 1
drop sd_ind
label var ohp_all_ever_survey "Ever on Medicaid"
label var ohp_all_at_12m      "Currently on Medicaid"
texsave * using "$ol/tab_c.tex", varlabels frag replace location("H") ///
        title("OLS Estimates of Utilization Outcomes on Medicaid Coverage Indicators")


*******************************************************************************
**# Part e, first stage estimates
*******************************************************************************

// Table 3 Columns 5 & 6
```

```
use "$repl_data/data_for_analysis.dta", clear

label var ohp_all_ever_survey "Ever on Medicaid"
label var ohp_all_at_12m      "Currently on Medicaid"

tempname tabpost
tempfile tabfile
postfile `tabpost' str100 varname sd_ind c_mean first_stage using `tabfile'
foreach var of varlist $mdcd_covg_vars {
        local lbl : var label `var'
        regress `var' if treatment == 0 & sample_12m_resp == 1 [pweight = weight_12m]
        local c_mean = _b[_cons]
        regress `var' treatment ddd* if sample_12m_resp == 1 [pweight = weight_12m], vce(cluster hous
        post `tabpost' ("`lbl'") (0) (`c_mean') (_b[treatment])
        post `tabpost' ("`lbl'") (1) (.)        (_se[treatment])
}
postclose `tabpost'
use `tabfile', clear

foreach var of varlist c_mean first_stage {
        format `var' %13.4fc
        tostring `var', replace force usedisplayformat
        replace `var' = "" if `var' == "."
        replace `var' = "(" + `var' + ")" if sd_ind == 1 & !missing(`var')
}
replace varname = "" if sd_ind == 1
drop sd_ind
label var c_mean      "Control mean"
label var first_stage "Estimated FS"
texsave * using "$ol/tab_e.tex", varlabels frag replace location("H") ///
        title("First-Stage Estimates")


*******************************************************************************
**# Part f, compliers analysis
*******************************************************************************

use "$repl_data/data_for_analysis.dta", clear

gen born_before_1968 = (birthyear_list < 1968)
label var born_before_1968 "Born before 1968"

global background_dummies born_before_1968 female_list english_list self_list ///
                          first_day_list have_phone_list pobox_list zip_msa

tempname tabpost
tempfile tabfile
postfile `tabpost' str100 xvarname str100 yvar coef using `tabfile'
foreach yvar of varlist $mdcd_covg_vars {
```

```
        // Size of complier group? Do a simpler first-stage with weights but no controls
        regress ‘yvar’ treatment if sample_12m_resp == 1 [pweight = weight_12m]
                local complier_share = _b[treatment]
                post ‘tabpost’ ("Share of compliers") ("‘yvar’") (‘complier_share’)

        // Characterize compliers by covariates: coef from first-stage with covar turned on vs overal
        foreach xvar of varlist $background_dummies {
                local lbl : var label ‘xvar’
                regress ‘yvar’ treatment if sample_12m_resp == 1 & ‘xvar’ == 1 [pweight = weight_12m]
                        local fs_xeq1 = _b[treatment]
                local x_comp_ratio = ‘fs_xeq1’ / ‘complier_share’
                post ‘tabpost’ ("‘lbl’") ("‘yvar’") (‘x_comp_ratio’)
        }
}
postclose ‘tabpost’
use ‘tabfile’, clear

gen orig_order = _n
egen yvar_order = min(orig_order), by(yvar)
bysort yvar_order (orig_order): gen order_within = _n
drop orig_order yvar_order
reshape wide coef order_within, i(xvarname) j(yvar) string
sort order*
drop order*
rename coef* *
format $mdcd_covg_vars %13.2fc
tostring $mdcd_covg_vars, replace force usedisplayformat
label var ohp_all_ever_survey "Ever on Medicaid"
label var ohp_all_at_12m      "Currently on Medicaid"
texsave * using "$ol/tab_f.tex", varlabels frag replace hlines(1) location("H") ///
        title("Shares of Compliers and Ratios of Covariates Among Compliers vs Full Sample")


*******************************************************************************
**# Part g, LATE estimation with 2SLS
*******************************************************************************

use "$repl_data/data_for_analysis.dta", clear

tempname tabpost
tempfile tabfile
postfile ‘tabpost’ str100 varname sd_ind late using ‘tabfile’
foreach var of varlist $survey_useext_list {
        local lbl : var label ‘var’
        ivregress 2sls ‘var’ (ohp_all_ever_survey = treatment) ddd* ///
                if sample_12m_resp == 1 [pweight = weight_12m], vce(cluster household_id)
        post ‘tabpost’ ("‘lbl’") (0) (_b[ohp_all_ever_survey])
        post ‘tabpost’ ("‘lbl’") (1) (_se[ohp_all_ever_survey])
}
postclose ‘tabpost’
use ‘tabfile’, clear
```

```
format late %13.4fc
tostring late, replace force usedisplayformat
replace late = "(" + late + ")" if sd_ind == 1 & !missing(late)
replace varname = "" if sd_ind == 1
drop sd_ind
label var late "LATE"
texsave * using "$ol/tab_g.tex", varlabels frag replace location("H") ///
        title("2SLS Estimates of LATE")
```