ECON 2320 – Economics of Labor and Population
Problem Set 2
Samuel Brown
Fall 2011
Due: October 26

1. (a) To identify the average treatment effect of maternal smoking by comparing unadjusted group means, smoking must be randomly assigned to mothers. An estimate of the effect of smoking on birthweight using this method can be seen in the first column of Table 2. The mean characteristics of non-smokers and smokers can be seen in Table 1. As is evident from these tables, non-smokers and smokers exhibit significantly different observable characteristics. This is evidence against random assignment of smoking.

Table 1: Descriptive Statistics

	Non-Smokers	Smokers	Total	Difference
Mother Age	27.239	25.368	26.889	1.871***
	(5.466)	(5.299)	(5.483)	(0.037)
Mother Education	13.126	11.871	12.892	1.255***
	(2.228)	(1.642)	(2.186)	(0.015)
Mother White	0.851	0.843	0.849	0.008**
	(0.356)	(0.364)	(0.358)	(0.002)
Mother Unmarried	0.179	0.413	0.223	−0.234***
	(0.384)	(0.492)	(0.416)	(0.003)
Father Age	29.540	28.354	29.319	1.187***
	(6.180)	(6.509)	(6.260)	(0.043)
Father Education	13.256	12.038	13.029	1.218***
	(2.359)	(1.773)	(2.310)	(0.016)
Father White	0.844	0.823	0.840	0.021***
	(0.363)	(0.382)	(0.367)	(0.003)
Mother Drank	0.009	0.051	0.017	−0.043***
	(0.092)	(0.221)	(0.128)	(0.001)
Prenatal Visits	11.167	10.256	10.997	0.911***
	(3.451)	(3.985)	(3.574)	(0.024)
Previous Deaths	0.325	0.436	0.346	−0.111***
	(0.694)	(0.830)	(0.723)	(0.005)
Mother Diabetic	0.017	0.015	0.016	0.002
	(0.129)	(0.122)	(0.127)	(0.001)
Mother Anemic	0.010	0.018	0.011	−0.008***
	(0.099)	(0.133)	(0.106)	(0.001)
N	112955	25934	138889	138889

Mean values; standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(b) An adjusted estimate of the average treatment effect of smoking on birthweight can be seen in the second column of Table 2. This effect is identified if maternal smoking is randomly assigned conditional on the observables, viz., all unobservable determinants of birthweight are *independent* of smoking assignment, conditional on the observables.

(c) The restrictive linear structure of (b) is likely a source of specification bias. An estimate of the birthweight equation in which higher-order terms and interactions have been added can be seen in the third column of Table 2. Though this method may produce more accurate estimates, it can be less precise and is computationally more intensive.

Table 2: OLS Estimates of Birthweight Equations

	(1)	(2)	(3)
Constant	3426.051***	3114.409***	2367.558***
	(1.696)	(42.935)	(701.418)
Mother Smoked	−257.865***	−216.574***	−213.959***
	(3.992)	(3.986)	(4.025)
Mother Age		18.371***	133.672
		(2.462)	(102.127)
Mother Age^2		−0.340***	−5.736
		(0.044)	(5.449)
Mother Education		−28.845***	−49.557***
		(5.092)	(7.350)
Mother Education^2		1.131***	1.873***
		(0.194)	(0.273)
Mother Unmarried		−41.134***	
		(4.631)	
Mother Black		−195.256***	
		(5.595)	
Mother Hispanic		−119.586***	
		(9.984)	
Mother Other Race		−184.963***	
		(11.724)	
Mother Drank		−48.676***	−41.954***
		(11.461)	(11.402)
Prenatal Visits		25.427***	24.285***
		(0.536)	(0.688)
Mo. Since Last		0.934***	3.394***
		(0.052)	(0.118)
Prev. Birth Pre-Term		−489.033***	−488.293***
		(15.326)	(16.918)
Prev. Birth > 4kg		518.148***	492.288***
		(13.581)	(14.371)
Twins or Greater		−952.379***	−962.984***
		(13.759)	(15.964)
Hypertension		−191.248***	−189.317***
		(11.405)	(13.650)
Mother Diabetic		42.840**	41.053**
		(13.167)	(14.806)
Int. and 2nd Order	No	No	Yes
R^2	0.030	0.152	0.159
F	4173.329	1126.417	.
N	138889	138889	138889

Dependent variable is infant birthweight. Robust standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(d) The propensity score approach assigns a probability of treatment, conditional on the observables, to each individual. If the treatment is randomly assigned conditional on the observables, then the propensity score is a sufficient statistic for selection bias. For this reason, it can be used to correct the unadjusted estimates. A benefit of this approach is that it only requires a single index

to adjust estimates, while that of (c) requires many regressors.

(e) A propensity score was estimated for each individual as follows: (i) a logit model containing interactions and higher order terms of the the observables was used to estimate propensity scores; (ii) the sample was stratified into blocks in which p-scores were not significantly different from each other at the 1% level; (iii) the observables in each block were tested for significant differences in their sample means; and (iv) if 90% of the observables in 90% of the blocks were statistically indistinguishable at the 1% level, the algorithm terminated. This produced the 29 blocks that are summarized in Table 3. Note that there are treatments and controls in every block.

Table 3: Propensity Score Blocks

Block No.	Non-smokers	Smokers	Total
1	9334	251	9585
2	17434	748	18182
3	4432	362	4794
4	6205	626	6831
5	6468	746	7214
6	1773	201	1974
7	1824	271	2095
8	4205	654	4859
9	11779	2023	13802
10	11687	2278	13965
11	7533	1731	9264
12	5851	1644	7495
13	2567	760	3327
14	2156	738	2894
15	3587	1200	4787
16	5151	2222	7373
17	3051	1567	4618
18	1147	659	1806
19	882	610	1492
20	2055	1553	3608
21	1387	1281	2668
22	944	1034	1978
23	783	1110	1893
24	435	808	1243
25	134	362	496
26	54	179	233
27	26	102	128
28	7	84	91
29	9	130	139
Total	112900	25934	138834

A box plot of propensity scores for non-smokers and smokers can be seen in Figure 1. Evidently there is overlap in the treatments and controls, though there is selection on the observables.

An estimate of the average treatment effect, adjusted using the p-score, can be seen in the first column of Table 4. If the average effect of treatment on the treated is calculated by taking a weighted average of differences in group means across the 29 blocks, this produces a statistic of -207.792. This statistic is slightly smaller than the p-score-adjusted estimate, but it is not significantly different.

(f) Estimates of the average treatment effect and average effect of treatment on the treated can be seen in the second and third columns of Table 4, respectively. Note that these estimates are not significantly different from those of (e). This suggests that the p-score adjustment is robust to selection bias. A benefit of using the p-scores as weights is that it produces more efficient estimates. However, it puts greater weight on treatments with low propensity scores and controls

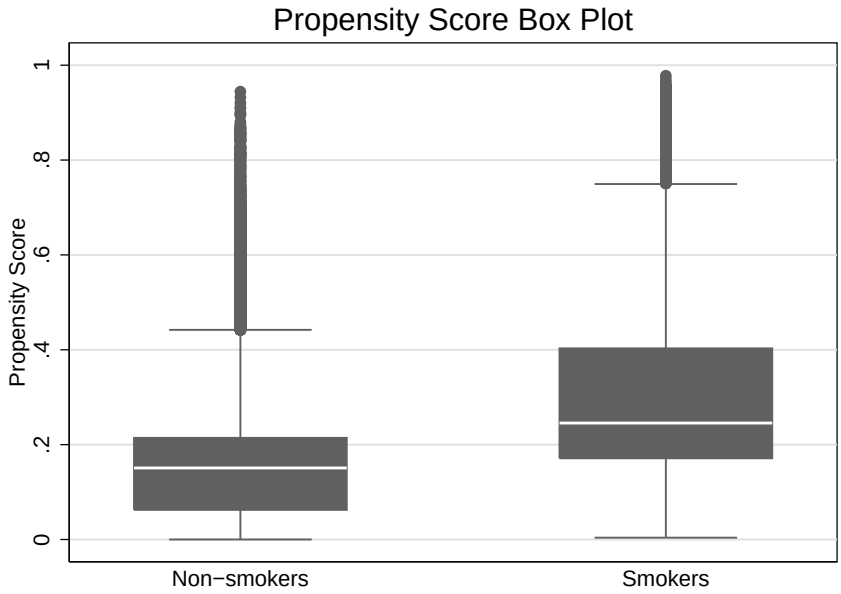

Figure 1: Propensity Score Distribution of Non-smokers and Smokers

with high propensity scores, so can bias estimates by putting a lot of weight on a small number of observations that may have measurement error or imprecisely estimated p-scores.

A plot of average estimated p-scores versus the actual fraction of smokers in blocks of 200 can be seen in Figure 2. Note that the p-scores "track" the 45-degree line, suggesting that the weighting procedure is reasonable.

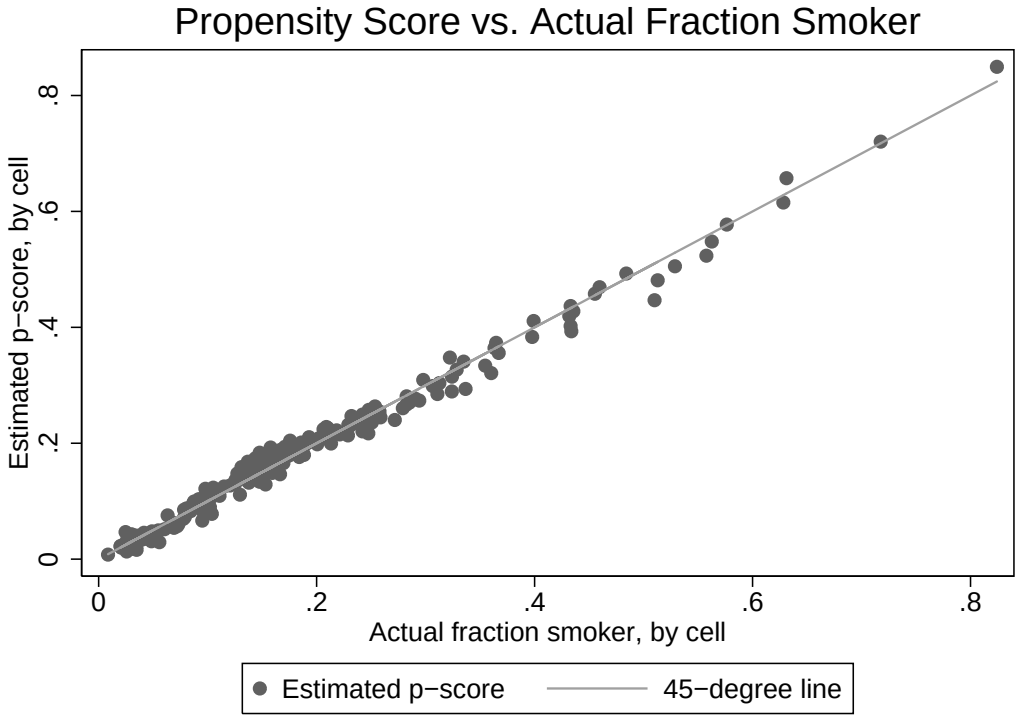

Figure 2: Propensity Score vs. Fraction of Smokers

The fourth column of Table 4 contains an estimate of the average treatment effect on the treated, using the fraction of smokers in each of the 200 blocks to weight the observations. Note that this estimate is not significantly different from those obtained using the estimated p-scores as weights, providing further evidence of the reasonableness of those estimates.

Table 4: OLS and WLS Estimates of Birthweight Equations

	(1) OLS	(2) WLS (ATE)	(3) WLS (TOT)	(4) WLS (TOT_C)
Constant	3482.190***	3417.235***	3377.760***	3376.522***
	(2.881)	(1.803)	(3.263)	(3.056)
Mother Smoked	−209.943***	−211.009***	−209.574***	−208.336***
	(7.135)	(5.916)	(4.869)	(4.732)
Propensity Score	−346.011***			
	(15.230)			
Mother Smoked · P-Score	−5.088			
	(26.033)			
R^2	0.036	0.033	0.031	0.030
F	1665.530	1272.123	1852.727	1937.981
N	138834	138834	138834	138834

Dependent variable is infant birthweight. Robust standard errors in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(g) A plot of average birthweight versus average estimated p-scores in blocks of 100, for non-smokers and smokers separately, can be seen in Figure 3. An analogous plot that combines non-smokers and smokers can be seen in Figure 4. These figures show that mean birthweight declines with propensity to smoke, regardless of smoking status, and that it is systematically lower (by about 200g) for smokers than for non-smokers. This is consistent with the bias in the unadjusted estimates (note especially the steeper slope in Figure 4), and with the robustness of the p-score-adjusted and weighted estimates.

(h) Weighted estimates of the average treatment effect of smoking on the probability of low birthweight can be seen in Table 5. A plot of low birthweight versus average estimated p-scores in blocks of 100, for non-smokers and smokers separately, can be seen in Figure 5. An analogous plot that combines non-smokers and smokers can be seen in Figure 6. These figures show that the probability of low birthweight increases with propensity to smoke, regardless of smoking status, and that it is systematically (though marginally) higher for smokers than for non-smokers.

Table 5: WLS Estimates of Low Birthweight Equations

	(1) WLS (ATE)	(2) WLS (TOT)
Constant	0.051***	0.061***
	(0.001)	(0.001)
Mother Smoked	0.037***	0.039***
	(0.003)	(0.002)
R^2	0.005	0.005
F	175.397	268.148
N	138834	138834

Dependent variable is indicator for low birthweight.

Robust standard errors in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(i) Estimates of the effect of smoking on infant death can be seen in Table 6. Note that once observables are controlled for, the effect diminishes by an order of magnitude and is no longer

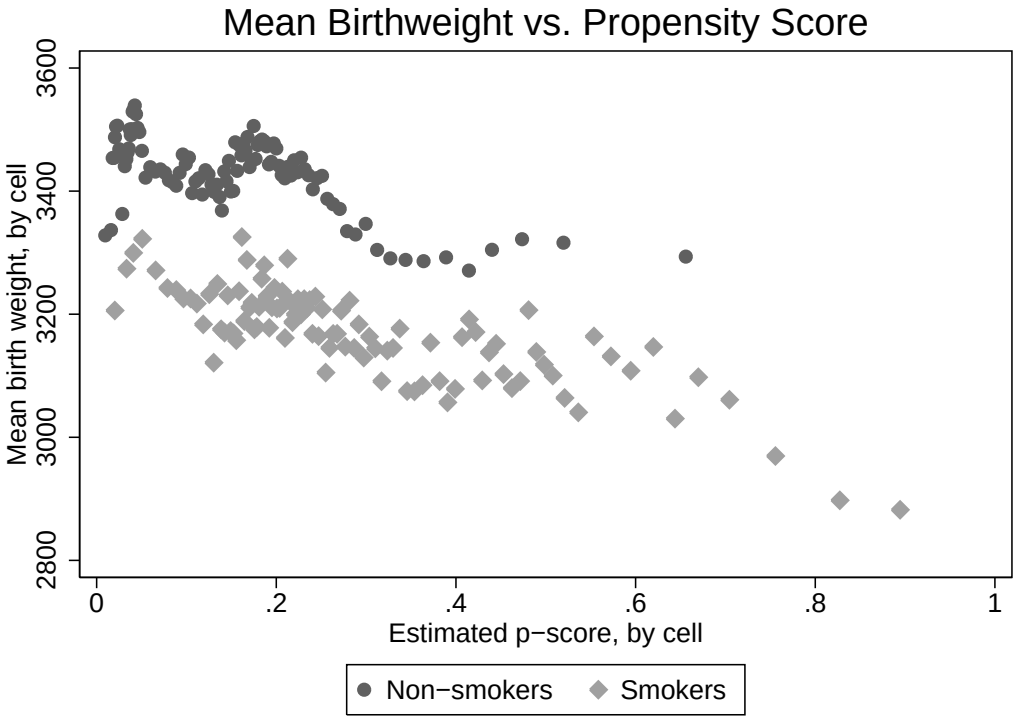

Figure 3: Mean Birthweight vs. Propensity Score

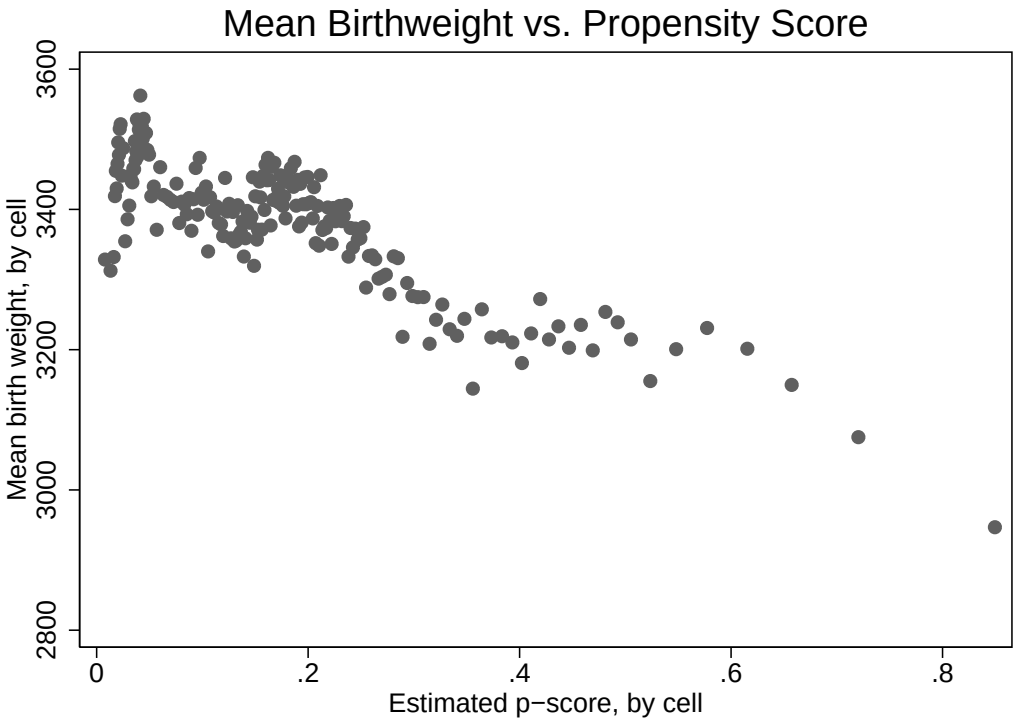

Figure 4: Mean Birthweight vs. Propensity Score

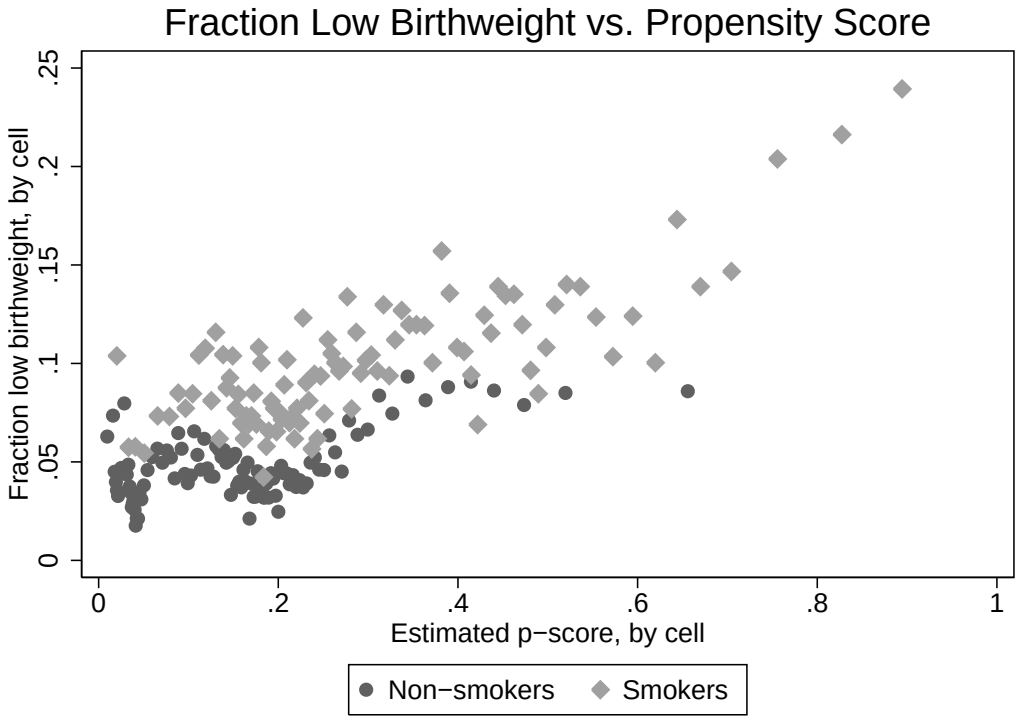

Figure 5: Low Birthweight vs. Propensity Score

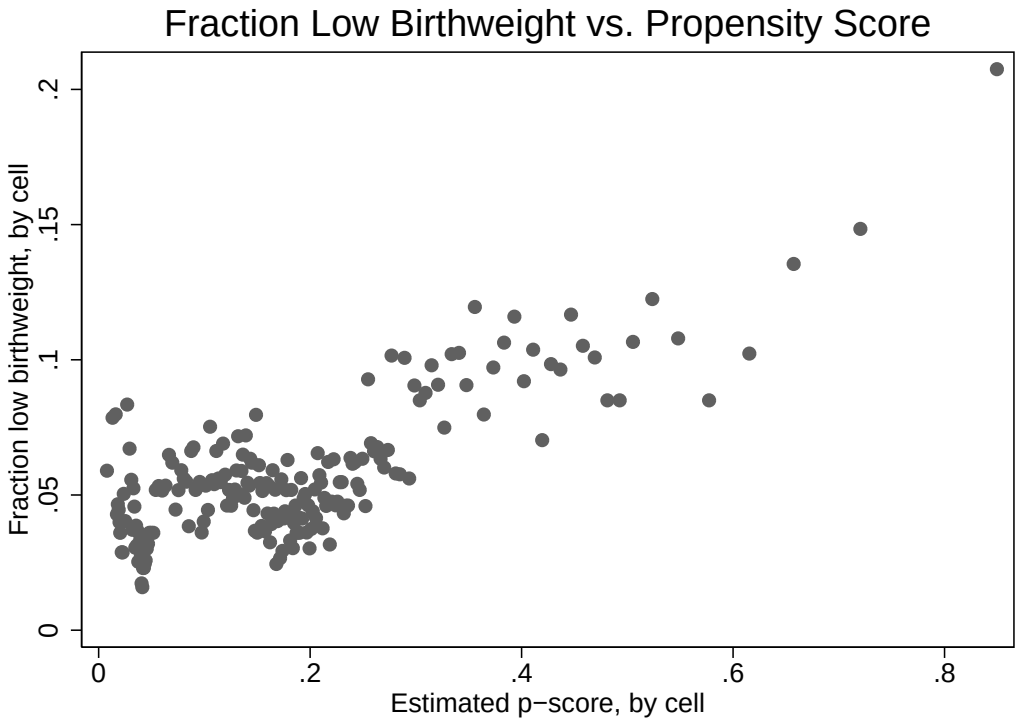

Figure 6: Low Birthweight vs. Propensity Score

significant. Neither is it significant when observations are weighted using their estimated p-scores, as in columns (3) and (4). This suggests that infant deaths are not caused by smoking-related reductions in birthweight, but instead caused by other characteristics that also make a woman more likely to smoke.

Table 6: OLS and WLS Estimates of Infant Death Equations

	(1) OLS	(2) OLS	(3) WLS (ATE)	(4) WLS (TOT)
Constant	0.006448***	0.028405***	0.006970***	0.009245***
	(0.000238)	(0.007453)	(0.000266)	(0.000503)
Mother Smoked	0.002768***	0.000470	0.000898	−0.000029
	(0.000639)	(0.000673)	(0.000947)	(0.000778)
Mother Age		−0.000902*		
		(0.000429)		
Mother Age^2		0.000018*		
		(0.000008)		
Mother Education		0.001243		
		(0.000813)		
Mother Education^2		−0.000049		
		(0.000030)		
Mother Unmarried		0.002414**		
		(0.000796)		
Mother Black		0.001064		
		(0.001014)		
Mother Hispanic		−0.001608		
		(0.001632)		
Mother Other Race		−0.000078		
		(0.001945)		
Mother Drank		−0.001940		
		(0.001887)		
Prenatal Visits		−0.001770***		
		(0.000106)		
Mo. Since Last		−0.000010		
		(0.000008)		
Prev. Birth Pre-Term		0.018875***		
		(0.003510)		
Prev. Birth > 4kg		−0.002420		
		(0.001527)		
Twins or Greater		0.025695***		
		(0.003757)		
Hypertension		0.002502		
		(0.001514)		
Mother Diabetic		0.003915*		
		(0.001864)		
R^2	0.000168	0.008884	0.000027	0.000000
F	18.734363	22.251426	0.898461	0.001419
N	138834	138834	138834	138834

Dependent variable is indicator for infant death. Robust standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(j) Sample sizes and smoking rates versus age can be seen in Figure 7. Average birthweight versus age can be seen in Figure 8. Evidently younger women are more likely to smoke. Moreover, younger women (and smokers, regardless of age) are more likely to have low birthweight infants.

(k) In summary, these results suggest that smoking reduces birthweight by roughly 210g, and that smoking does not affect infant deaths. Mothers who smoke are already more likely to have low birthweight infants, regardless of their smoking status. It is likely this propensity for low birthweight that is driving infant mortality among smoking mothers. The robustness of the estimated average treatment effect of maternal smoking to alternative estimation methods indicates its credibility.

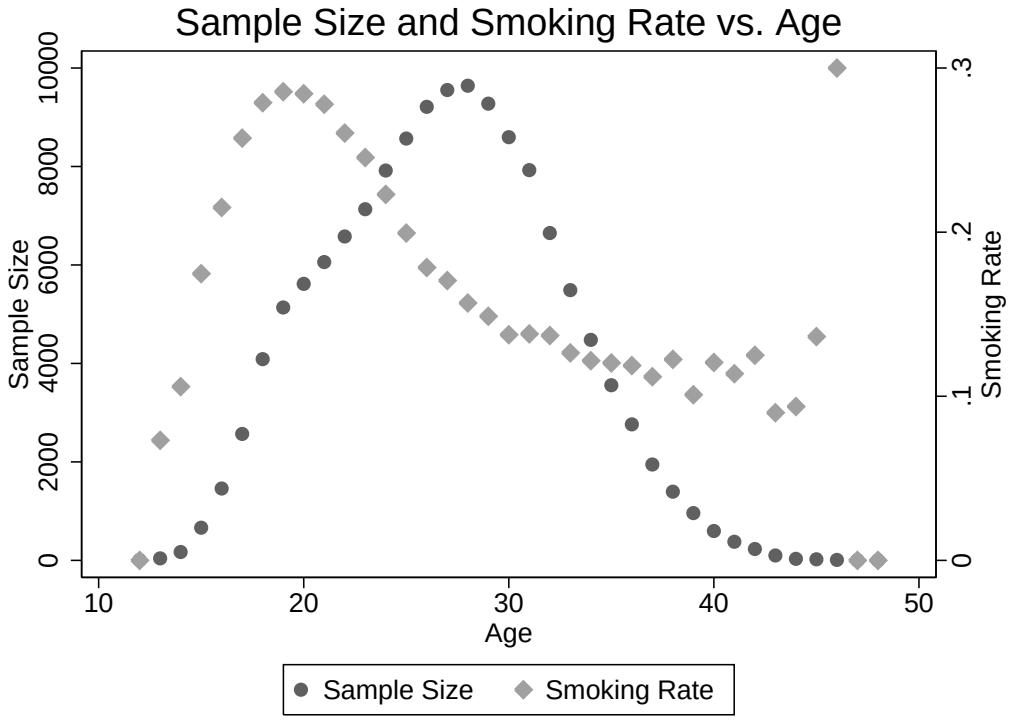

Figure 7: Sample Size and Smoking Rate vs. Age

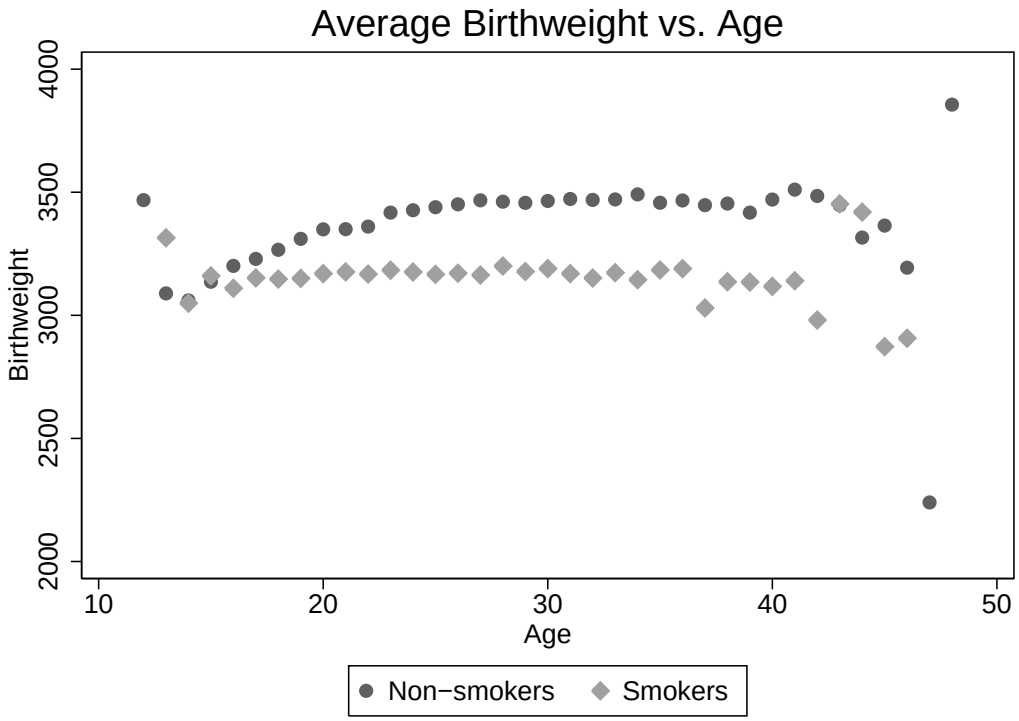

Figure 8: Average Birthweight vs. Age

A STATA Source Code

```
cd "E:\2011 Fall\Labor and Population Economics\Problem Sets\PS2"

local nblock = 5
local level = 0.01

local fcount = 0

set more off

set matsize 11000

cap log close PS2
log using PS2, replace text name(PS2)

cap which estout
if _rc ssc install estout
cap which savesome
if _rc ssc install savesome

set scheme s1mono
graph set eps logo off
graph set eps preview off
graph set eps orientation portrait
graph set eps mag 100

use smoking2.dta, replace
// outsheet using smoking2.txt, nolabel replace

la define tobla 0 "Non-smokers" 1 "Smokers"
la values tobacco tobla

rename mblack dmblack
rename mhispan dmhispan
rename motherr dmotherr

drop if dmblack+dmhispan+dmotherr>1

gen dmwhite = ~(dmblack | dmhispan | dmotherr)
gen dfwhite = ~(fblack | fhispan | fotherr)
gen dmage2 = dmage^2
gen dmage3 = dmage^3
gen dmage4 = dmage^4
gen dmeduc2 = dmeduc^2
gen disllb2 = disllb^2


la var dmage "Mother Age"
la var dmage2 "Mother Age$^2$"
la var dmeduc "Mother Education"
la var dmeduc2 "Mother Education$^2$"
la var dmar "Mother Unmarried"
la var dfage "Father Age"
la var dfeduc "Father Education"
la var dfwhite "Father White"
```

```
la var alcohol "Mother Drank"
la var nprevist "Prenatal Visits"
la var deadkids "Previous Deaths"
la var diabete "Mother Diabetic"
la var anemia "Mother Anemic"
la var tobacco "Mother Smoked"
la var dmwhite "Mother White"
la var dmblack "Mother Black"
la var dmhispan "Mother Hispanic"
la var dmotherr "Mother Other Race"
la var drink "Num. Drinks"
la var dlivord "Birth Order"
la var disllb "Mo. Since Last"
la var disllb2 "Mo. Since Last$^2$"
la var preterm "Prev. Birth Pre-Term"
la var pre4000 "Prev. Birth $>4\text{kg}$"
la var plural "Twins or Greater"
la var phyper "Hypertension"

eststo r_bw_t: reg dbirwt tobacco, vce(robust)

local mtvars dmage dmeduc dmwhite dmar dfage dfeduc dfwhite alcohol nprevist ///
deadkids diabete anemia
qui eststo mt_not: estpost summarize ‘mtvars’ if tobacco==0
qui eststo mt_tob: estpost summarize ‘mtvars’ if tobacco==1
qui eststo mt_ttl: estpost summarize ‘mtvars’ if tobacco~=.
qui eststo tt_tob: estpost ttest ‘mtvars’, by(tobacco)
qui estadd mat mean=e(b)
qui estadd mat sd=e(se)

esttab mt_not mt_tob mt_ttl tt_tob using TEXDocs/tab1.tex, replace ///
cells(mean(fmt(%9.3f) star) sd(fmt(%9.3f) par)) ///
stats(N, fmt(%9.0g) layout("\multicolumn{1}{c}{@}") ///
labels("\ensuremath{N}")) ///
nodepvar ///
nonumber ///
collabels(none) ///
mtitles("Non-Smokers" "Smokers" "Total" "Difference") ///
label ///
legend ///
note("Mean values; standard errors in parentheses.") ///
title("Descriptive Statistics\label{tab:groupm}") ///
substitute(none) ///
gaps ///
alignment(D{.}{.}{-1}) ///
booktabs ///
style(tex)

eststo r_bw_tc: reg dbirwt tobacco dmage dmage2 dmeduc dmeduc2 dmar ///
dmblack dmhispan dmotherr alcohol nprevist disllb preterm pre4000 ///
plural phyper diabete, vce(robust)

egen category = group(dmar dmblack dmhispan dmotherr)
egen dmrace = group(dmblack dmhispan dmotherr)
xi: eststo r_bw_tmvm: reg dbirwt ///
tobacco ///
```

```
alcohol ///
i.category ///
i.dmrace|preterm ///
i.category|pre4000 ///
i.category|plural ///
i.category|phyper ///
i.category|diabete ///
///
i.category|dmage ///
i.category|dmage2 ///
i.category|dmage3 ///
i.category|dmage4 ///
i.category|dmeduc ///
i.category|dmeduc2 ///
i.category|nprevist ///
i.category|disllb ///
i.category|disllb2, vce(robust)

// Regression Table for:
//   dbirwt on tobacco
//   dbirwt on tobacco and covariates
//   dbirwt on tobacco and covariates, interactions, and 2nd order terms
esttab r_bw_t r_bw_tc r_bw_tmvm using TEXDocs/reg1.tex, replace ///
order(_cons tobacco dmage dmage2 dmage3 dmage4 dmeduc dmeduc2 dmar dmblack ///
dmhispan dmotherr alcohol nprevist disllb disllb2 preterm pre4000 ///
plural phyper diabete) ///
varlabels(_cons "Constant") ///
cells(b(star fmt(%9.3f)) se(par fmt(%9.3f))) ///
stats(r2 F N, fmt(%9.3f %9.3f %9.0g) ///
layout("@" "@" "\multicolumn{1}{c}{@}") ///
labels("\ensuremath{R^2}" "\ensuremath{F}" "\ensuremath{N}")) ///
indicate("Int. and $2^{\text{nd}}$ Order"= _I* *3 *4 disllb2, ///
labels("\multicolumn{1}{c}{\ensuremath{\text{Yes}}}" ///
"\multicolumn{1}{c}{\ensuremath{\text{No}}}")) ///
nomtitles ///
label ///
collabels(none) ///
legend ///
addnotes("Dependent variable is infant birthweight." ///
"Robust standard errors in parentheses.") ///
title("OLS Estimates of Birthweight Equations\label{reg:ols}") ///
alignment(D{.}{.}{-1}) ///
booktabs ///
style(tex)

// proceed with pscore estimation
capture drop pscore
capture drop block
capture drop block_inf
capture drop block_sup
// estimate the propensity score
xi: logit tobacco ///
i.category ///
alcohol ///
i.dmrace|preterm ///
i.category|pre4000 ///
```

```
i.category|plural ///
i.category|phyper ///
i.category|diabete ///
///
i.category|dmage ///
i.category|dmage2 ///
i.category|dmage3 ///
i.category|dmage4 ///
i.category|dmeduc ///
i.category|dmeduc2 ///
i.category|nprevist ///
i.category|disllb ///
i.category|disllb2 ///

predict double pscore
la var pscore "Propensity Score"
sum pscore, detail

// divide the data into blocks
gen id = _n // store obs. ids so we can restore the original order later
xtile block = pscore, nq(‘nblock’)
by block, sort: egen block_inf = min(pscore)
by block, sort: egen block_sup = max(pscore)
la var block "Block No."
la var block_inf "Block Infimum"
la var block_sup "Block Supremum"

// subdivide blocks until pscores are balanced
tab block tobacco
local iblock = 1
while ‘iblock’<=‘nblock’ {
di "block ‘iblock’ of ‘nblock’..."
// count the number of treated/contols in this block
qui count if block == ‘iblock’ & tobacco==0
local ncontrol = r(N)
qui count if block == ‘iblock’ & tobacco==1
local ntreated = r(N)
// skip blocks with no treatments or no controls in them
if ‘ncontrol’ == 0 | ‘ntreated’ == 0 {
local iblock = ‘iblock’+1
}
else {
local ivar = 1
local nvar = 0
local fail_count = 0
// test for difference in mean values
qui ttest pscore if block==‘iblock’, by(tobacco)
if r(p)<‘level’ {
local fail_count = ‘fail_count’+1
}
// if pscores are different, split the block
if ‘fail_count’>0 {
// shift the remaining blocks up
qui replace block = block+1 if block>‘iblock’ & block!=.
local nblock = ‘nblock’+1
// get the block split point
```

```
sum block_inf if block==‘iblock’, meanonly
local temp_inf = r(mean)
sum block_sup if block==‘iblock’, meanonly
local temp_sup = r(mean)
local split = (‘temp_inf’ + ‘temp_sup’)/2
// split the block
di "splitting block ‘iblock’"
qui replace block = block+1 if block==‘iblock’ & pscore>=‘split’ ///
& pscore<=block_sup
qui replace block_sup = ‘split’ if block==‘iblock’
qui replace block_inf = ‘split’ if block==‘iblock’+1
} // end if
else {
// pscores are not different, iterate to the next block
local iblock = ‘iblock’+1
} // end else
} // end else
} // end while
tab block tobacco

// check that the covariates are balanced
local total_fail = 0
local iblock = 1
while ‘iblock’<=‘nblock’ {
di "block ‘iblock’ of ‘nblock’..."
// count the number of treated/contols in this block
qui count if block == ‘iblock’ & tobacco==0
local ncontrol = r(N)
qui count if block == ‘iblock’ & tobacco==1
local ntreated = r(N)
// skip blocks with no treatments or no controls in them
if ‘ncontrol’ == 0 | ‘ntreated’ == 0 {
local iblock = ‘iblock’+1
}
else {
local ivar = 1
local nvar = 0
local fail_count = 0
// test for difference in mean values
foreach var in tobacco dmage dmeduc dmar dmblack dmhispan ///
dmotherr alcohol nprevist disllb preterm ///
pre4000 plural phyper diabete {
qui ttest ‘var’ if block==‘iblock’, by(tobacco)
if r(p)<‘level’ {
di "‘var’ fail in block ‘iblock’!"
local fail_count = ‘fail_count’+1
}
local nvar = ‘nvar’+1
} // end foreach
// if 10% or more mean values are different, fail
if (‘fail_count’)/‘nvar’>=.10 {
local total_fail = ‘total_fail’+1
} // end if
// iterate to the next block
local iblock = ‘iblock’+1
} // end else
```

```
} // end while
di "`total_fail' (" round(`total_fail'/`nblock'*100) ///
"%) blocks were unbalanced"
// append the pscore data to the file on disk
preserve
sort id
keep pscore block block_inf block_sup
merge 1:1 _n using smoking2.dta, nogen
save smoking2.dta, replace
restore

// graph the pscore boxplot
local fcount = `fcount'+1
graph box pscore, over(tobacco) ///
title(Propensity Score Box Plot) legend(off) ///
name(fig`fcount', replace) saving(Figures\fig`fcount'.gph, replace)
graph export Figures\fig`fcount'.eps, replace

// tabulate the blocks
estpost tab block tobacco
esttab . using TEXDocs/tab2.tex, replace ///
cell(b(fmt(g)) count(fmt(g) par keep(Total))) ///
collabels(none) unstack noobs nonumber nomtitle  ///
eqlabels(, lhs("Block No.")) ///
varlabels(, blist(Total "{hline @width}{break}")) ///
title("Propensity Score Blocks\label{tab:blocks}") ///
alignment(D{.}{.}{-1}) ///
booktabs ///
style(tex)

// regression-adjust using the pscore
local pscore_mean = sum(pscore)/sum(pscore~=.)
disp `pscore_mean'
gen tobacco_pscore = tobacco*(pscore-`pscore_mean')
la var tobacco_pscore "\ensuremath{\text{Mother Smoked}\cdot\text{P-Score}}"
eststo r_bw_tp: reg dbirwt tobacco pscore tobacco_pscore, vce(robust)

// stratify the sample and estimate TOT
drop if pscore == .
// calculate the total number of treatments
egen nT = total(tobacco)
// claulate number of treatments and controls in each block
egen block_nT = total(tobacco), by(block)
egen block_nC = total(~tobacco), by(block)
// not calculate TOT from the stratified sample
gen block_unwtTOT = tobacco*dbirwt/block_nT-(~tobacco)*dbirwt/block_nC
gen block_wt = block_nT/nT
preserve
collapse (sum) block_unwtTOT (first) block_wt, by(block)
gen TOT = block_unwtTOT*block_wt
collapse (sum) TOT
disp "TOT=" TOT
restore

// low birthweight
gen lowbw = (dbirwt<2500)
```

```
// ATE and TOT probability weights
gen ate_weight = (~tobacco)*1/(1-pscore)+tobacco*1/pscore
gen tot_weight = (~tobacco)*pscore/(1-pscore)+tobacco

foreach var in dbirwt lowbw {
if "`var'"=="dbirwt" {
local yvar = "Mean birth weight, by cell"
local gtitle = "Mean Birthweight vs. Propensity Score"
}
else {
local yvar = "Fraction low birthweight, by cell"
local gtitle = "Fraction Low Birthweight vs. Propensity Score"
}

// regress using the pscores as weights
eststo r_bw_tpw_ate_`var': reg `var' tobacco [pw=ate_weight], vce(robust)
eststo r_bw_tpw_tot_`var': reg `var' tobacco [pw=tot_weight], vce(robust)

if "`var'"=="dbirwt" {
// check reasonableness of reweighing procedure
capture drop block200
capture drop block200_nT
capture drop block200_nC
capture drop block200_fracT
xtile block200 = pscore, nq(200)
egen block200_nT = total(tobacco), by(block200)
egen block200_nC = total(~tobacco), by(block200)
gen block200_fracT = block200_nT/(block200_nT+block200_nC)
preserve
collapse (mean) pscore (first) block200_fracT, by(block200)
local fcount = `fcount'+1
twoway scatter pscore block200_fracT || line block200_fracT block200_fracT || , ///
title(Propensity Score vs. Actual Fraction Smoker) ///
xtitle("Actual fraction smoker, by cell") ///
ytitle("Estimated p-score, by cell") ///
legend(on label(1 "Estimated p-score") label(2 "45-degree line")) ///
name(fig`fcount', replace) saving(Figures\fig`fcount'.gph, replace)
graph export Figures\fig`fcount'.eps, replace
restore
// estimate TOT using actual fraction of smokers in each block
capture drop tot_weight200
gen tot_weight200 = (~tobacco)*block200_fracT/(1-block200_fracT)+tobacco
eststo r_bw_tpw_tot200_`var': reg `var' tobacco [pw=tot_weight200], vce(robust)
}

// compare "non-parametric" means
preserve
keep if tobacco==1
capture drop block100
xtile block100 = pscore, nq(100)
collapse (mean) `var' pscore (first) tobacco, by(block100)
save smokers, replace
restore
preserve
keep if tobacco==0
capture drop block100
```

```
xtile block100 = pscore, nq(100)
collapse (mean) `var' pscore (first) tobacco, by(block100)
append using smokers
local fcount = `fcount'+1
twoway scatter `var' pscore if tobacco==0 || scatter `var' pscore if tobacco==1 || , ///
title("`gtitle'") ///
xtitle("Estimated p-score, by cell") ///
ytitle("`yvar'") ///
legend(on label(1 "Non-smokers") label(2 "Smokers")) ///
name(fig`fcount', replace) saving(Figures\fig`fcount'.gph, replace)
graph export Figures\fig`fcount'.eps, replace
erase smokers.dta
restore
preserve
capture drop block200
xtile block200 = pscore, nq(200)
collapse (mean) `var' pscore (first) tobacco, by(block200)
local fcount = `fcount'+1
scatter `var' pscore, ///
title("`gtitle'") ///
xtitle("Estimated p-score, by cell") ///
ytitle("`yvar'") ///
name(fig`fcount', replace) saving(Figures\fig`fcount'.gph, replace)
graph export Figures\fig`fcount'.eps, replace
restore
} // end for

// effect of smoking on infant death
eststo r_d_t: reg death tobacco, vce(robust)
eststo r_d_tc: reg death tobacco dmage dmage2 dmeduc dmeduc2 dmar ///
dmblack dmhispan dmotherr alcohol nprevist disllb preterm pre4000 ///
plural phyper diabete, vce(robust)
eststo r_d_tpw_ate: reg death tobacco [pw=ate_weight], vce(robust)
eststo r_d_tpw_tot: reg death tobacco [pw=tot_weight], vce(robust)

// smoking lifecycle
preserve
egen ageblock = group(dmage)
collapse (sum) ageblock_nT=tobacco ageblock_nC=~tobacco (first) dmage ///
(count) ageblock_n=dmage, by(ageblock)
gen ageblock_fracT = ageblock_nT/ageblock_n
local fcount = `fcount'+1
twoway scatter ageblock_n dmage || scatter ageblock_fracT dmage, yaxis(2) || , ///
title("Sample Size and Smoking Rate vs. Age") ///
xtitle("Age") ///
ytitle("Sample Size", axis(1)) ///
ytitle("Smoking Rate", axis(2)) ///
legend(on label(1 "Sample Size") label(2 "Smoking Rate")) ///
name(fig`fcount', replace) saving(Figures\fig`fcount'.gph, replace)
graph export Figures\fig`fcount'.eps, replace
restore
preserve
egen agetgroup = group(dmage tobacco)
collapse (mean) dbirwt (first) dmage tobacco, by(agetgroup)
local fcount = `fcount'+1
twoway scatter dbirwt dmage if tobacco==0 || scatter dbirwt dmage if tobacco==1|| , ///
```

```
title("Average Birthweight vs. Age") ///
xtitle("Age") ///
ytitle("Birthweight") ///
legend(on label(1 "Non-smokers") label(2 "Smokers")) ///
name(fig‘fcount’, replace) saving(Figures\fig‘fcount’.gph, replace)
graph export Figures\fig‘fcount’.eps, replace
restore

// Regression Table for:
//   dbirwt on tobacco and pscore
//   dbirwt on tobacco, weighted by pscore ate
//   dbirwt on tobacco, weighted by pscore tot
//   dbirwt on tobacco, weighted by fractional tot
esttab r_bw_tp r_bw_tpw_ate_dbirwt r_bw_tpw_tot_dbirwt r_bw_tpw_tot200_dbirwt ///
using TEXDocs/reg2.tex, replace ///
order(_cons tobacco pscore tobacco_pscore) ///
varlabels(_cons "Constant") ///
cells(b(star fmt(%9.3f)) se(par fmt(%9.3f))) ///
stats(r2 F N, fmt(%9.3f %9.3f %9.0g) ///
layout("@" "@" "\multicolumn{1}{c}{@}") ///
labels("\ensuremath{R^2}" "\ensuremath{F}" "\ensuremath{N}")) ///
mtitles("OLS" "WLS (ATE)" "WLS (TOT)" "WLS (TOT$^{’}$)") ///
label ///
collabels(none) ///
legend ///
addnotes("Dependent variable is infant birthweight." ///
"Robust standard errors in parentheses.") ///
title("OLS and WLS Estimates of Birthweight Equations\label{reg:wls}") ///
alignment(D{.}{.}{-1}) ///
booktabs ///
style(tex)

// Regression Table for:
//   lowbw on tobacco, weighted by pscore ate
//   lowbw on tobacco, weighted by pscore tot
esttab r_bw_tpw_ate_lowbw r_bw_tpw_tot_lowbw ///
using TEXDocs/reg3.tex, replace ///
order(_cons tobacco) ///
varlabels(_cons "Constant") ///
cells(b(star fmt(%9.3f)) se(par fmt(%9.3f))) ///
stats(r2 F N, fmt(%9.3f %9.3f %9.0g) ///
layout("@" "@" "\multicolumn{1}{c}{@}") ///
labels("\ensuremath{R^2}" "\ensuremath{F}" "\ensuremath{N}")) ///
mtitles("WLS (ATE)" "WLS (TOT)") ///
label ///
collabels(none) ///
legend ///
addnotes("Dependent variable is indicator for low birthweight." ///
"Robust standard errors in parentheses.") ///
title("WLS Estimates of Low Birthweight Equations\label{reg:wlslbw}") ///
alignment(D{.}{.}{-1}) ///
booktabs ///
style(tex)

// Regression Table for:
//   death on tobacco
```

```
//   death on tobacco and covariates
//   death on tobacco, weighted by pscore ate
//   death on tobacco, weighted by pscore tot
esttab r_d_t r_d_tc r_d_tpw_ate r_d_tpw_tot using TEXDocs/reg4.tex, replace ///
order(_cons tobacco dmage dmage2 dmeduc dmeduc2 dmar dmblack ///
dmhispan dmotherr alcohol nprevist dislb preterm pre4000 ///
plural phyper diabete) ///
varlabels(_cons "Constant") ///
cells(b(star fmt(%9.6f)) se(par fmt(%9.6f))) ///
stats(r2 F N, fmt(%9.6f %9.6f %9.0g) ///
layout("@" "@" "\multicolumn{1}{c}{@}") ///
labels("\ensuremath{R^2}" "\ensuremath{F}" "\ensuremath{N}")) ///
mtitles("OLS" "OLS" "WLS (ATE)" "WLS (TOT)") ///
label ///
collabels(none) ///
legend ///
addnotes("Dependent variable is indicator for infant death." ///
"Robust standard errors in parentheses.") ///
title("OLS and WLS Estimates of Infant Death Equations\label{reg:death}") ///
alignment(D{.}{.}{-1}) ///
booktabs ///
style(tex)

capture log close PS2
```