

# Estimating treatment effects from observational data using teffects, stteffects, and eteffects

David M. Drukker

Executive Director of Econometrics  
Stata

Belgian Stata Users Group  
6 September 2016

# A question

- Will a mother hurt her child by smoking while she is pregnant?
  - Too vague
- Will a mother reduce the birthweight of her child by smoking while she is pregnant?
  - Less interesting, but more specific
  - There might even be data to help us answer this question
  - The data will be observational, not experimental

# Potential outcomes

- For each treatment level, there is a potential outcome that we would observe if a subject received that treatment level
- Potential outcomes are the data that we wish we had to estimate causal treatment effects
- In the example at hand, the two treatment levels are the mother smokes and the mother does not smoke
  - For each treatment level, there is an outcome (a baby's birthweight) that would be observed if the mother got that treatment level

# Potential outcomes

- Suppose that we could see
  - 1 the birthweight of a child born to each mother when she smoked while pregnant, and
  - 2 the birthweight of a child born to each mother when she did not smoke while pregnant

For example, we wish we had data like

```
. list mother_id bw_smoke bw_nosmoke in 1/5, abbreviate(10)
```

	mother_id	bw_smoke	bw_nosmoke
1.	1	3183	3509
2.	2	3060	3316
3.	3	3165	3474
4.	4	3176	3495
5.	5	3241	3413

# Average treatment effect

- If we had data on each potential outcome, the sample-average treatment effect would be the sample average of `bw_smoke` minus `bw_nosmoke`

```
. mean bw_smoke bw_nosmoke
Mean estimation                Number of obs   =       4,642
```

	Mean	Std. Err.	[95% Conf. Interval]	
bw_smoke	3171.72	.9088219	3169.938	3173.501
bw_nosmoke	3402.599	1.529189	3399.601	3405.597

```
. lincom _b[bw_smoke] - _b[bw_nosmoke]
(1) bw_smoke - bw_nosmoke = 0
```

Mean	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
(1)	-230.8791	1.222589	-188.84	0.000	-233.276	-228.4823

- In population terms, the average treatment effect is

$$ATE = \mathbf{E}[bw_{smoke} - bw_{nosmoke}] = \mathbf{E}[bw_{smoke}] - \mathbf{E}[bw_{nosmoke}]$$

# Missing data

- The “fundamental problem of causal inference” (Holland (1986)) is that we only observe one of the potential outcomes
  - The other potential outcome is missing
    - ① We only see  $bw_{smoke}$  for mothers who smoked
    - ② We only see  $bw_{nosmoke}$  for mothers who did not smoked
- We can use the tricks of missing-data analysis to estimate treatment effects
- For more about potential outcomes Rubin (1974), Holland (1986), Heckman (1997), Imbens (2004), (Cameron and Trivedi, 2005, chapter 2.7), Imbens and Wooldridge (2009), and (Wooldridge, 2010, chapter 21)

# Random-assignment case

- Many questions require using observational data, because experimental data would be unethical
  - We could not ask a random selection of pregnant women to smoke while pregnant
- The random-assignment methods used with experimental data are useful, because observational-data methods build on them
- When the treatment is randomly assigned, the potential outcomes are independent of the treatment
- If smoking were randomly assigned to mothers, the missing potential outcome would be missing completely at random
  - 1 The average birthweight of babies born to mothers who smoked would be a good estimator for mean of the smoking potential outcome of all mothers in the population
  - 2 The average birthweight of babies born to mothers who did not smoke would be a good estimator for mean of the not-smoking potential outcome of all mothers in the population

# As good as random

- Instead of assuming that the treatment is randomly assigned, we assume that the treatment is as good as randomly assigned after conditioning on covariates
- Formally, this assumption is known as conditional independence
- Even more formally, we only need conditional mean independence which says that after conditioning on covariates, the treatment does not affect the means of the potential outcomes



# Assumptions used with observational data

- The assumptions we need vary over estimator and effect parameter, but some version of the following assumptions are required for the exogenous treatment estimators discussed here
  - CMI** The conditional mean-independence CMI assumption restricts the dependence between the treatment model and the potential outcomes
  - Overlap** The overlap assumption ensures that each individual could get any treatment level
  - IID** The independent-and-identically-distributed (IID) sampling assumption ensures that the potential outcomes and treatment status of each individual are unrelated to the potential outcomes and treatment statuses of all the other individuals in the population
- Endogenous treatment effect models replace CMI with a weaker assumption
- In practice, we assume independent observations, not IID

# Some references for assumptions

## For Reference Only

- Versions of the CMI assumption are also known as unconfoundedness and selection-on-observables in the literature; see Rosenbaum and Rubin (1983), Heckman (1997), Heckman and Navarro-Lozano (2004), (Cameron and Trivedi, 2005, section 25.2.1), (Tsiatis, 2006, section 13.3), (Angrist and Pischke, 2009, chapter 3), Imbens and Wooldridge (2009), and (Wooldridge, 2010, section 21.3)
- Rosenbaum and Rubin (1983) call the combination of conditional independence and overlap assumptions strong ignorability; see also (Abadie and Imbens, 2006, pp 237-238) and Imbens and Wooldridge (2009).
- The IID assumption is a part of what is known as the stable unit treatment value assumption (SUTVA); see (Wooldridge, 2010, p.905) and Imbens and Wooldridge (2009)

# Choice of auxiliary model

- Recall that the potential-outcomes framework formulates the estimation of the ATE as a missing-data problem
- We use the parameters of an auxiliary model to solve the missing-data problem
  - The auxiliary model is how we condition on covariates so that the treatment is as good as randomly assigned

Model	Estimator
outcome	→ Regression adjustment (RA)
treatment	→ Inverse-probability weighted (IPW)
outcome and treatment	→ Augmented IPW (AIPW)
outcome and treatment	→ IPW RA (IPWRA)
outcome (nonparametrically)	→ Nearest-neighbor matching (NNMATCH)
treatment	→ Propensity-score matching (PSMATCH)

# Regression adjustment estimators

- Regression adjustment (RA) estimators:
  - RA estimators run separate regressions for each treatment level, then
    - means of predicted outcomes using all the data and the estimated coefficients for treatment level  $i$  all the data estimate  $POM_i$
    - use differences of POMs, or conditional on the treated POMs, to estimate ATEs or ATETs
  - Formally, the CMI assumption implies that our regressions of observed  $y$  for a given treatment level directly estimate  $\mathbf{E}[y_t|\mathbf{x}_i]$ 
    - $y_t$  is the potential outcome for treatment level  $t$
    - $\mathbf{x}_i$  are the covariates on which we condition
    - Averages of predicted  $\mathbf{E}[y_t|\mathbf{x}_i]$  yield estimates of the POM  $\mathbf{E}[y_t]$  because  $1/N \sum_{i=1}^N \widehat{\mathbf{E}}[y_t|\mathbf{x}_i] \rightarrow_p \mathbf{E}_x[\widehat{\mathbf{E}}[y_t|\mathbf{x}_i]] = \mathbf{E}[y_t]$
- See (Cameron and Trivedi, 2005, chapter 25), (Wooldridge, 2010, chapter 21), and (Vittinghoff et al., 2012, chapter 9)

## RA example

```

. use cattaneo2
(Excerpt from Cattaneo (2010) Journal of Econometrics 155: 138-154)
. teffects ra (bweight mmarried prenatal1 fbaby medu) (mbsmoke)
Iteration 0: EE criterion = 2.336e-23
Iteration 1: EE criterion = 5.702e-26
Treatment-effects estimation          Number of obs      =      4,642
Estimator       : regression adjustment
Outcome model   : linear
Treatment model : none

```

bweight		Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
ATE							
mbsmoke (smoker vs nonsmoker)		-230.9541	24.34012	-9.49	0.000	-278.6599	-183.2484
POmean							
mbsmoke nonsmoker		3402.548	9.546721	356.41	0.000	3383.836	3421.259

- When all pregnant women smoke the average baby birthweight is estimated to be 231 grams less than when no pregnant women smoke
- The average birthweight when no pregnant women smoke is estimated to be 3403 grams

## RA exponential-mean example

```
. teffects ra (bweight mmarried prenatal1 fbaby medu, poisson) (mbsmoke)
Iteration 0: EE criterion = 3.926e-17
Iteration 1: EE criterion = 1.666e-23
Treatment-effects estimation          Number of obs      =      4,642
Estimator      : regression adjustment
Outcome model  : Poisson
Treatment model: none
```

bweight	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
ATE						
mbsmoke (smoker vs nonsmoker)	-230.7723	24.41324	-9.45	0.000	-278.6213	-182.9232
POmean						
mbsmoke nonsmoker	3402.497	9.547989	356.36	0.000	3383.783	3421.211

- RA using exponential mean  $E[y_t|\mathbf{x}] = \exp(\mathbf{x}\beta_t)$  because birthweights are greater than 0
- `teffects ra` can also model the outcome using probit, logit, heteroskedastic probit, exponential mean, or poisson

# Why are the standard errors always robust?

- have a multistep estimator
  - ① Regress  $y$  on  $\mathbf{x}$  for not treated observations
  - ② Regress  $y$  on  $\mathbf{x}$  for treated observations
  - ③ Mean of all observations of predicted  $y$  given  $\mathbf{x}$  from not-treated regression estimates
  - ④ Mean of all observations of predicted  $y$  given  $\mathbf{x}$  from treated regression estimates
- Each step can be obtained by solving moment conditions yielding a method of moments estimator known as an estimating equation (EE) estimator
  - $\mathbf{m}_i(\boldsymbol{\theta})$  is vector of moment equations and  $\mathbf{m}(\boldsymbol{\theta}) = 1/N \sum_{i=1}^N \mathbf{m}_i(\boldsymbol{\theta})$
- The estimator for the variance-covariance matrix of the estimator has the form  $1/N(DMD')$  where  $D = \left( \frac{1}{N} \frac{\partial \mathbf{m}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^{-1}$  and
 
$$M = \frac{1}{N} \sum_{i=1}^N \mathbf{m}_i(\boldsymbol{\theta}) \mathbf{m}_i(\boldsymbol{\theta})'$$
- Stacked moments do not yield a symmetric  $D$ , so no simplification under correct specification

# Inverse-probability-weighted estimators

- Inverse-probability-weighted (IPW) estimators:
  - IPW estimators weight observations on the outcome variable by the inverse of the probability that it is observed to account for the missingness process
  - Observations that are not likely to contain missing data get a weight close to one; observations that are likely to contain missing data get a weight larger than one, potentially much larger
  - IPW estimators model the probability of treatment without any assumptions about the functional form for the outcome model
  - In contrast, RA estimators model the outcome without any assumptions about the functional form for the probability of treatment model
- See Horvitz and Thompson (1952) Robins and Rotnitzky (1995), Robins et al. (1994), Robins et al. (1995), Imbens (2000), Wooldridge (2002), Hirano et al. (2003), (Tsiatis, 2006, chapter 6), Wooldridge (2007) and (Wooldridge, 2010, chapters 19 and 21)



```

. teffects ipw (bweight ) (mbsmoke mmarried prenatal1 fbaby medu)
Iteration 0:  EE criterion = 1.701e-23
Iteration 1:  EE criterion = 6.343e-27
Treatment-effects estimation      Number of obs      =      4,642
Estimator      : inverse-probability weights
Outcome model  : weighted mean
Treatment model: logit

```

bweight	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
ATE mbsmoke (smoker vs nonsmoker)	-231.1516	24.03183	-9.62	0.000	-278.2531	-184.0501
POmean mbsmoke nonsmoker	3402.219	9.589812	354.77	0.000	3383.423	3421.015

- IPW with logit to model treatment
- Could have used probit or heteroskedastic probit to model treatment
- Estimator has stacked moment structure; score equations from first-stage maximum-likelihood estimators are now moment equations

# Augmented IPW estimators

- Augmented IPW (AIPW) estimators
  - Augmented-inverse-probability-weighted (AIPW) estimators model both the outcome and the treatment probability
  - The estimating equation that combines both models is essentially an IPW estimating equation with an augmentation term
  - AIPW estimator have the double-robust property
    - only one of the two models must be correctly specified to consistently estimate the treatment effects
  - AIPW estimators can be more efficient than IPW or RA estimators
- See Robins and Rotnitzky (1995), Robins et al. (1995), Lunceford and Davidian (2004), Bang and Robins (2005), (Tsiatis, 2006, chapter 13), Cattaneo (2010), Cattaneo, Drukker, and Holland (2013)

## AIPW example I

```
. teffects aipw (bweight mmarried prenatal1 fbaby medu) ///
> (mbsmoke mmarried prenatal1 fbaby medu)
Iteration 0: EE criterion = 4.031e-23
Iteration 1: EE criterion = 2.180e-26
Treatment-effects estimation          Number of obs   =       4,642
Estimator      : augmented IPW
Outcome model  : linear by ML
Treatment model: logit
```

bweight	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
ATE mbsmoke (smoker vs nonsmoker)	-229.7809	24.96839	-9.20	0.000	-278.718	-180.8437
POmean mbsmoke nonsmoker	3403.122	9.564165	355.82	0.000	3384.376	3421.867

- AIPW with linear model for outcome and logit for treatment

```
. teffects aipw (bweight mmarried prenatal1 fbaby medu, poisson) ///
>      (mbsmoke mmarried prenatal1 fbaby medu, hetprobit(medu))
Iteration 0:  EE criterion = 7.551e-16
Iteration 1:  EE criterion = 8.767e-24
Treatment-effects estimation          Number of obs   =      4,642
Estimator      : augmented IPW
Outcome model  : Poisson by ML
Treatment model: heteroskedastic probit
```

bweight		Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
ATE	mbsmoke (smoker vs nonsmoker)	-220.496	28.30292	-7.79	0.000	-275.9687	-165.0233
POmean	mbsmoke nonsmoker	3402.429	9.557345	356.00	0.000	3383.697	3421.161

- AIPW with exponential conditional mean model for outcome and heteroskedastic probit for treatment
- Could have used linear, poisson, logit, probit, or heteroskedastic probit to model the outcome and probit, logit, or heteroskedastic logit to model the treatment

# Balance: As good as random

- In the unobtainable case of a randomly assigned treatment, the distribution of the covariates among those that get the treatment is the same as the distribution of the covariates among those that do not get the treatment
  - The distribution of the covariates is said to be “balanced” over the treatment/control status
- The estimators implemented in `teffects` use a model or matching method to make the outcome conditionally independent of the treatment by conditioning on covariates
  - If this model or matching method is well specified, it should balance the covariates
  - Balance diagnostic techniques and tests check the specification of the conditioning method used by a `teffects`

# Balance with IPW

- Rosenbaum and Rubin (1983) showed that the propensity score is a balancing score
  - In particular, the treatment is conditionally independent of the covariates after conditioning on the propensity score
  - Among the many applications of this result is the implication that IPW means of covariates will be the same for treated and controls
  - The raw means of covariates will differ over treated and control observations, but the IPW means will be similar

# tebalance

- `tebalance` implements diagnostics and a test for balance after `teffects`
  - Diagnostics are statistics and graphical methods for which we do not know the distribution under the null
  - A test is a statistic for which we know the distribution under the null
- `tebalance` is new to Stata 14

# An example using the Cattaneo data

- Let's look for evidence against balancing using the simple model

```
. clear all
. use cattaneo2
(Excerpt from Cattaneo (2010) Journal of Econometrics 155: 138-154)
. quietly teffects ipw (bweight) (mbsmoke mmarried mage prenatal1 fbaby medu)
>
. tebalance summarize
Covariate balance summary
```

	Raw	Weighted
Number of obs =	4,642	4,642.0
Treated obs =	864	2,280.4
Control obs =	3,778	2,361.6

	Standardized differences		Variance ratio	
	Raw	Weighted	Raw	Weighted
mmarried	-.5953009	-.0258113	1.335944	1.021696
mage	-.300179	-.0803657	.8818025	.8127244
prenatal1	-.3242695	-.0228922	1.496155	1.034023
fbaby	-.1663271	.0221042	.9430944	1.005032
medu	-.5474357	-.1373455	.7315846	.4984786



# Standardized differences

- Group differences scaled by the average the group variances are known as standardized differences
- The raw standardized differences between treatment levels  $t_1$  and  $t_0$  are

$$\delta(t_1, t_0) = \frac{\hat{\mu}_x(t_1) - \hat{\mu}_x(t_0)}{\sqrt{\hat{\sigma}_x^2(t_1) + \hat{\sigma}_x^2(t_0)}}$$

where

$$\hat{\mu}_x(t) = \frac{1}{N_t} \sum_{i=1}^N (t_i == t) x_i$$

$$\hat{\sigma}_x(t) = \frac{1}{N_t - 1} \sum_{i=1}^N (t_i == t) (x_i - \hat{\mu}_x(t))^2$$

# IPW standardized differences

- If the model for the treatment is correctly specified, the IPW standardized differences will be zero
- The IPW standardized differences between treatment levels  $t_1$  and  $t_0$  are

$$\delta(t_1, t_0) = \frac{\tilde{\mu}_x(t_1) - \tilde{\mu}_x(t_0)}{\sqrt{\tilde{\sigma}_x^2(t_1) + \tilde{\sigma}_x^2(t_0)}}$$

where

$$\tilde{\mu}_x(t) = \frac{1}{M_t} \sum_{i=1}^N \omega_i(t_i == t) x_i$$

$$\tilde{\sigma}_x(t) = \frac{1}{M_t - 1} \sum_{i=1}^N (t_i == t) \omega_i (x_i - \tilde{\mu}_x(t))^2$$

and  $\omega_i$  are the normalized predicted treatment probabilities and

$$M_t = \sum_{i=1}^N (t_i == t) \omega_i$$

# Test for balance

- Imai and Ratkovic (2014) derived a test for balance by viewing the restrictions imposed by balance as overidentifying conditions.
  - Scores for ML estimator of propensity score are moment conditions
  - Moment conditions for equality of means are over-identifying conditions
  - Estimate over-identified parameters by generalized method of moments (GMM)
  - Under the null of covariate balance GMM criterion statistic has  $\chi^2(J)$  distribution, where  $J$  is the number of over-identifying moment conditions imposed by covariate balance

```

. quietly teffects ipw (bweight) (mbsmoke mmarried mage prenatal1 fbaby medu)
>
. tebalance override
Iteration 0: criterion = .01513068
Iteration 1: criterion = .01514951 (backed up)
Iteration 2: criterion = .01521006
Iteration 3: criterion = .01539644
Iteration 4: criterion = .01542377
Iteration 5: criterion = .01550797
Iteration 6: criterion = .01553409
Iteration 7: criterion = .01558564
Iteration 8: criterion = .01568553
Iteration 9: criterion = .01569184
Iteration 10: criterion = .01572741
Iteration 11: criterion = .01573404
Iteration 12: criterion = .01573406
Overidentification test for covariate balance
      H0: Covariates are balanced:
      chi2(6)      = 62.5564
      Prob > chi2  = 0.0000

```

Reject null hypothesis that IPW model/weights balance covariates

# Model selection

- How to selection the model for the outcome or the treatment?
- Use theory to decide the set of covariates
  - Do not condition on variables that are affected by the treatment, Wooldridge (2005)
- What functional form of a set or super set of the correct covariates should I use?

# Minimizing an information criterion

- The idea is to fit a bunch of models and select the model with smallest information criterion
  - An information criterion is  $-LL + \text{penalty term}$ 
    - The better the estimator fits the data, the smaller is the negative of the log-likelihood ( $-LL$ )
    - The more parameters are added to the model, the larger is the penalty term
- Choosing the model that minimizes an information criteria has a long history in statistics and econometrics
  - Claeskens and Hjort (2008), (Cameron and Trivedi, 2005, Section 8.5.1)

# Minimizing an information criterion

- Minimizing the Bayesian information criterion (BIC) can be a consistent model selection technique
  - Selecting the model that minimizes the BIC is an estimator of which model to select
  - The model selected by this estimator converges to the true model as the sample size gets larger
  - $BIC = -2LL + 2\ln(N)q$ , where  $N$  is the sample size and  $q$  is the number of parameters
- Minimizing the Akaike information criterion (AIC) tends to select a model with too many terms
  - The model selected by this estimator converges to a model that overfits as the sample size gets larger
  - $AIC = -2LL + 2q$

# bfit does model selection

- `bfit` is a user written command documented in Cattaneo et al. (2013)
- `bfit` will find the model that minimizes either the BIC or the AIC within a subset of all possible models



```
. bfit logit mbsmoke mmarried mage prenatal1 fbaby medu
bfit logit results sorted by bic
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
_bfit_32	4642	-2230.748	-2002.985	9	4023.97	4081.956
_bfit_30	4642	-2230.748	-2012.263	7	4038.525	4083.626
_bfit_31	4642	-2230.748	-2008.151	8	4032.302	4083.845
_bfit_33	4642	-2230.748	-1995.658	12	4015.316	4092.631
_bfit_34	4642	-2230.748	-1989.613	18	4015.225	4131.197
_bfit_19	4642	-2230.748	-2033.762	8	4083.524	4135.067
_bfit_18	4642	-2230.748	-2040.745	7	4095.49	4140.591
_bfit_25	4642	-2230.748	-2039.028	8	4094.056	4145.6
_bfit_16	4642	-2230.748	-2053.041	5	4116.081	4148.296
_bfit_17	4642	-2230.748	-2049.147	6	4110.294	4148.952
_bfit_26	4642	-2230.748	-2033.566	10	4087.132	4151.561
_bfit_12	4642	-2230.748	-2051.069	6	4114.138	4152.796
_bfit_23	4642	-2230.748	-2051.658	6	4115.316	4153.974
_bfit_24	4642	-2230.748	-2047.907	7	4109.815	4154.915
_bfit_20	4642	-2230.748	-2027.135	12	4078.271	4155.585
_bfit_14	4642	-2230.748	-2029.388	12	4082.776	4160.091
_bfit_11	4642	-2230.748	-2055.651	6	4123.303	4161.96
_bfit_13	4642	-2230.748	-2044.248	9	4106.496	4164.482
_bfit_35	4642	-2230.748	-1983.735	24	4015.469	4170.099
_bfit_21	4642	-2230.748	-2017.789	16	4067.577	4170.664
_bfit_9	4642	-2230.748	-2072.867	4	4153.733	4179.505
_bfit_27	4642	-2230.748	-2026.593	15	4083.186	4179.83
_bfit_10	4642	-2230.748	-2069.425	5	4148.85	4181.065
_bfit_7	4642	-2230.748	-2060.093	8	4136.187	4187.73
_bfit_28	4642	-2230.748	-2016.621	20	4073.242	4202.1
_bfit_4	4642	-2230.748	-2082.388	5	4174.776	4206.99
_bfit_6	4642	-2230.748	-2079.501	6	4171.003	4209.66
_bfit_5	4642	-2230.748	-2088.62	4	4185.241	4211.012
_bfit_29	4642	-2230.748	-2085.159	6	4182.317	4220.975
_bfit_3	4642	-2230.748	-2102.649	4	4213.297	4239.069
_bfit_2	4642	-2230.748	-2109.805	3	4225.61	4244.939
_bfit_15	4642	-2230.748	-2133.02	4	4274.041	4299.812
_bfit_22	4642	-2230.748	-2130.327	5	4270.653	4302.868
_bfit_8	4642	-2230.748	-2138.799	3	4283.598	4302.926
_bfit_1	4642	-2230.748	-2200.161	2	4404.322	4417.207

Note: N= used in calculating BIC

(results \_bfit\_32 are active now)

```
. display "r(bvlist)"
```

```
i. (mmarried prenatal1 fbaby) mage medu c.mage#c.mage c.mage#c.medu c.medu#c.med
```

```
> u
```

## Over-identification test with selected model

```
. teffects ipw (bweight) (mbsmoke i.(mmarried prenatal1 fbaby) mage medu ///
>          c.mage#c.mage c.mage#c.medu c.medu#c.medu), nolog
Treatment-effects estimation      Number of obs      =      4,642
Estimator      : inverse-probability weights
Outcome model  : weighted mean
Treatment model: logit
```

bweight		Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
ATE	mbsmoke (smoker vs nonsmoker)	-220.7592	28.47705	-7.75	0.000	-276.5732	-164.9452
POmean	mbsmoke nonsmoker	3403.625	9.544666	356.60	0.000	3384.917	3422.332

```
. tebalance overid, nolog
Overidentification test for covariate balance
      H0: Covariates are balanced:
      chi2(9)      =   9.38347
      Prob > chi2  =   0.4027
```

# Survival-time example

Does smoking decrease the time to a second heart attack in the population of men aged 45–55 who have had one heart attack?

- 1 For ethical reasons, these data will be observational.
- 2 This question is about the time to an event, and such data are commonly known as survival-time data or time-to-event data. These data are nonnegative and, frequently, right-censored.
- 3 Many researchers and practitioners want an effect estimate in easy-to-understand units of time.

Much of the survival-time literature uses a hazard ratio as the effect of interest. The ATE has three advantages over the hazard ratio as an effect measure.

- 1 The ATE measures the effect in the same time units as the outcome instead of in relative conditional probabilities.
- 2 The ATE is much easier to explain to nontechnical audiences.
- 3 The models used to estimate the ATE can be much more flexible.

Hazard ratios are useful for population effects when they are constant, which occurs when the treatment enters linearly and the distribution of the outcome has a proportional-hazards form.

Neither linearity in treatment nor proportional-hazards form is required for the ATE, and neither is imposed on the models fit by the estimators implemented in `stteffects`.

# Estimators in stteffects

- Regression adjustment (RA)
  - Model outcome
  - Treatment assignment is handled by estimating separate models for each treatment level
  - Censoring handled in log-likelihood function for outcome
- Inverse-probability weighting
  - Model treatment assignment
  - Outcome is not modeled; estimated is weighted average of observed outcomes
  - Censoring handled by modeling time to censoring, which must be random
- Inverse-probability weighted regression adjustment (IPWRA)
  - Model outcome and treatment
  - Censoring handled in one of two ways
    - Censoring handled in log-likelihood function for outcome, or
    - Censoring handled by modeling time to censoring, which must be random
- stteffects is new Stata 14

## stset the data

```

. use sheart
(Time to second heart attack (fictional))
. stset atime, failure(fail)
      failure event:  fail != 0 & fail < .
obs. time interval:  (0, atime]
exit on or before:  failure

```

---

```

      2000  total observations
         0  exclusions

```

---

```

      2000  observations remaining, representing
      1208  failures in single-record/single-failure data
3795.226  total analysis time at risk and under observation
                                     at risk from t =          0
                                     earliest observed entry t =      0
                                     last observed exit t = 34.17743

```

- 1,208 of the 2,000 observations record actual time to a second heart attack; remainder were censored

## stteffects ra

```

. stteffects ra (age exercise diet education) (smoke)
      failure _d:  fail
      analysis time _t:  atime
Iteration 0:  EE criterion = 1.525e-19
Iteration 1:  EE criterion = 3.127e-31
Survival treatment-effects estimation      Number of obs      =      2,000
Estimator      :  regression adjustment
Outcome model  :  Weibull
Treatment model:  none
Censoring model:  none

```

_t	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
ATE smoke (Smoker vs Nonsmoker)	-1.956657	.3331787	-5.87	0.000	-2.609676	-1.303639
POmean smoke Nonsmoker	4.243974	.2620538	16.20	0.000	3.730358	4.75759

- The time to second heart attack is 1.96 years sooner when all the men smoke instead of when none of them smoke

## stteffects ipw

```

. stteffects ipw (smoke age exercise diet education)    ///
>      (age exercise diet education)
      failure _d: fail
      analysis time _t: atime
Iteration 0:  EE criterion = 2.042e-18
Iteration 1:  EE criterion = 3.796e-31
Survival treatment-effects estimation          Number of obs    =      2,000
Estimator      : inverse-probability weights
Outcome model  : weighted mean
Treatment model: logit
Censoring model: Weibull

```

_t	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
ATE smoke (Smoker vs Nonsmoker)	-2.187297	.6319837	-3.46	0.001	-3.425962	-.9486314
POmean smoke Nonsmoker	4.225331	.517501	8.16	0.000	3.211047	5.239614



## stteffects ipwra: likelihood adjustment for censoring

```

. stteffects ipwra (age exercise diet education)      ///
>      (smoke age exercise diet education)
      failure _d: fail
      analysis time _t: atime
Iteration 0:  EE criterion = 2.153e-16
Iteration 1:  EE criterion = 9.051e-30
Survival treatment-effects estimation                Number of obs    =      2,000
Estimator      : IPW regression adjustment
Outcome model  : Weibull
Treatment model: logit
Censoring model: none

```

_t	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
ATE smoke (Smoker vs Nonsmoker)	-1.592494	.4872777	-3.27	0.001	-2.54754	-.637447
POmean smoke Nonsmoker	4.214523	.2600165	16.21	0.000	3.7049	4.724146

## stteffects ipwra: Weighted adjustment for censoring

```

. stteffects ipwra (age exercise diet education)          ///
>   (smoke age exercise diet education)                ///
>   (age exercise diet)
      failure _d: fail
      analysis time _t: atime
Iteration 0:   EE criterion = 1.632e-16
Iteration 1:   EE criterion = 9.890e-31
Survival treatment-effects estimation      Number of obs   =      2,000
Estimator      : IPW regression adjustment
Outcome model  : Weibull
Treatment model: logit
Censoring model: Weibull

```

_t	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
ATE						
smoke (Smoker vs Nonsmoker)	-2.037944	.6032549	-3.38	0.001	-3.220302	-.855586
POmean						
smoke Nonsmoker	4.14284	.4811052	8.61	0.000	3.199891	5.085789

# Endogenous treatment effects

- Allow an unobserved component to affect treatment assignment and each potential outcome
  - Violates CMI even though covariates are unrelated to error terms
- View the estimators implemented in `eteffects` as extensions to RA for a type of endogenous treatment
- `eteffects` is new to Stata 14

# Endogenous treatment effects

- Here are the equations, when the outcome is linear

$$y_0 = \mathbf{x}\beta_0 + \epsilon_0 + \gamma_0\nu$$

$$y_1 = \mathbf{x}\beta_1 + \epsilon_1 + \gamma_1\nu$$

$$t = (\mathbf{z}\alpha + \nu > 0)$$

$$y = ty_1 + (1 - t)y_0$$

- $\mathbf{x}$  and  $\mathbf{z}$  are unrelated to  $\nu$  and  $\epsilon$
- $\nu \sim N(0, 1)$
- The endogeneity is caused by the presence of  $\nu$  in all the equations

# Endogenous treatment effects: Method

- Estimate probit of treatment on  $\mathbf{z}$ , and get residuals  $\hat{\nu}$
- Regress  $y$  on  $\mathbf{x}$  and  $\hat{\nu}$ , when  $t=0$  to get  $\hat{\mu}_{0i} = \hat{\mathbf{E}}[y_0|\mathbf{x}_i, \nu_i]$
- Regress  $y$  on  $\mathbf{x}$  and  $\hat{\nu}$ , when  $t=1$  to get  $\hat{\mu}_{1i} = \hat{\mathbf{E}}[y_1|\mathbf{x}_i, \nu_i]$
- ATE is average of  $\hat{\mu}_{1i} - \hat{\mu}_{0i}$
- Correct standard errors by stacking the moment conditions

## RA estimates

```

. use pschool
. teffects aipw (gpa hgpa pedu) (private i.religious pincome i.squality)
Iteration 0: EE criterion = 2.190e-15
Iteration 1: EE criterion = 8.081e-27
Treatment-effects estimation          Number of obs    =    10,000
Estimator      : augmented IPW
Outcome model  : linear by ML
Treatment model: logit

```

gpa		Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
ATE							
private	(Yes vs No)	.5856043	.0071606	81.78	0.000	.5715697	.5996389
POmean							
private	No	3.114636	.003141	991.60	0.000	3.10848	3.120792

## eteffects estimates

```
. eteffects (gpa hgpa pedu) (private i.religious pincome i.squality)
Iteration 0: EE criterion = 2.029e-22
Iteration 1: EE criterion = 1.040e-31
Endogenous treatment-effects estimation      Number of obs      =      10,000
Outcome model : linear
Treatment model: probit
```

gpa		Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
ATE							
private		.1295686	.0225492	5.75	0.000	.0853729	.1737642
(Yes vs No)							
POmean							
private		3.181094	.0048958	649.75	0.000	3.171498	3.19069
No							

# Testing for endogeneity

- There is no endogeneity if the coefficients on the control term, the generalized residuals, are zero
- A Wald test that these coefficients are jointly zero is a test of the null hypothesis of no endogeneity



# Testing for endogeneity

```
. estat endogenous
Test of endogeneity
Ho: treatment and outcome unobservables are uncorrelated
      chi2( 2) = 418.18
      Prob > chi2 = 0.0000
```

# Other functional forms

- Outcome model in `eteffects` could be fractional, probit, or exponential-mean, in addition to linear

# Now what?

- Go to <http://www.stata.com/manuals14/te.pdf> entry `teffects` intro advanced for more information and lots of links to literature and examples

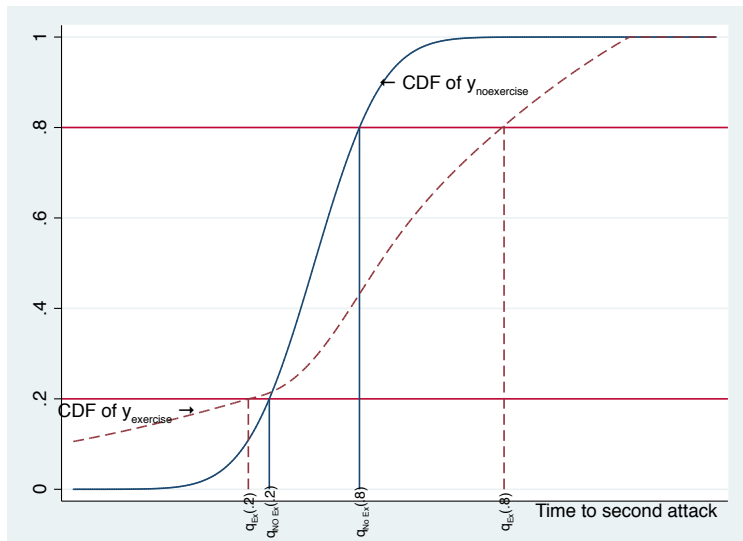
# QTEs for survival data

- Imagine a study that followed middle-aged men for two years after suffering a heart attack
  - Does exercise affect the time to a second heart attack?
  - Some observations on the time to second heart attack are censored
  - Observational data implies that treatment allocation depends on covariates
  - We use a model for the outcome to adjust for this dependence

# QTEs for survival data

- Exercise could help individuals with relatively strong hearts but not help those with weak hearts
- For each treatment level, a strong-heart individual is in the .75 quantile of the marginal, over the covariates, distribution of time to second heart attack
  - $QTE(.75)$  is difference in .75 marginal quantiles
- Weak-heart individual would be in the .25 quantile of the marginal distribution for each treatment level
  - $QTE(.25)$  is difference in .25 marginal quantiles
- our story indicates that the  $QTE(.75)$  should be significantly larger than the  $QTE(.25)$

# What are QTEs?



# Quantile Treatment effects

- We can easily estimate the marginal quantiles, but estimating the quantile of the differences is harder
- We need a rank preservation assumption to ensure that quantile of the differences is the difference in the quantiles
  - The  $\tau$ (th) quantile of  $y_1$  minus the  $\tau$ (th) quantile of  $y_0$  is not the same as the  $\tau$ (th) quantile of  $(y_1 - y_0)$  unless we impose a rank-preservation assumption
  - Rank preservation means that the random shocks that affect the treated and the not-treated potential outcomes do not change the rank of the individuals in the population

The rank of an individual in  $y_1$  is the same as the rank of that individual in  $y_0$

- Graphically, the horizontal lines must intersect the CDFs “at the same individual”

# A regression-adjustment estimator for QTEs

- Estimate the  $\theta_1$  parameters of  $F(y|\mathbf{x}, t = 1, \theta_1)$  the CDF conditional on covariates and conditional on treatment level
  - Conditional independence implies that this conditional on treatment level CDF estimates the CDF of the treated potential outcome
- Similarly, estimate the  $\theta_0$  parameters of  $F(y|\mathbf{x}, t = 0, \theta_0)$
- At the point  $y$ ,

$$1/N \sum_{i=1}^N F(y|\mathbf{x}_i, \hat{\theta}_1)$$

estimates the marginal distribution of the treated potential outcome

- The  $\hat{q}_{1,.75}$  that solves

$$1/N \sum_{i=1}^N F(\hat{q}_{1,.75}|\mathbf{x}_i, \hat{\theta}_1) = .75$$

estimates the .75 marginal quantile for the treated potential outcome



# A regression-adjustment estimator for QTEs

- The  $\hat{q}_{0,.75}$  that solves

$$1/N \sum_{i=1}^N F(\hat{q}_{0,.75} | \mathbf{x}_i, \hat{\boldsymbol{\theta}}_0) = .75$$

estimates the .75 marginal quantile for the control potential outcome

- $\hat{q}_1(.75) - \hat{q}_0(.75)$  consistently estimates QTE(.75)
- See Drukker (2014) for details

## mqgamma example

- mqgamma is a user-written command documented in Drukker (2014)
- . ssc install mqgamma

```
. use exercise, clear
. mqgamma t active, treat(exercise) fail(fail) lns(health) quantile(.25 .75)
Iteration 0: EE criterion = .7032254
Iteration 1: EE criterion = .05262105
Iteration 2: EE criterion = .00028553
Iteration 3: EE criterion = 6.892e-07
Iteration 4: EE criterion = 4.706e-12
Iteration 5: EE criterion = 1.604e-22
Gamma marginal quantile estimation      Number of obs      =      2000
```

t	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
q25_0						
_cons	.2151604	.0159611	13.48	0.000	.1838771	.2464436
q25_1						
_cons	.2612655	.0249856	10.46	0.000	.2122946	.3102364
q75_0						
_cons	1.591147	.0725607	21.93	0.000	1.44893	1.733363
q75_1						
_cons	2.510068	.1349917	18.59	0.000	2.245489	2.774647

## mqgamma example

```
. nlcom (_b[q25_1:_cons] - _b[q25_0:_cons])      ///
>      (_b[q75_1:_cons] - _b[q75_0:_cons])
      _nl_1:  _b[q25_1:_cons] - _b[q25_0:_cons]
      _nl_2:  _b[q75_1:_cons] - _b[q75_0:_cons]
```

t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
_nl_1	.0461051	.0295846	1.56	0.119	-.0118796	.1040899
_nl_2	.9189214	.1529012	6.01	0.000	.6192405	1.218602

# poparms also estimates QTEs

- `poparms` is a user-written command documented in Cattaneo, Drukker, and Holland (2013)
- `poparms` estimates mean and quantiles of the potential-outcome distributions
  - `poparms` implements an IPW and an AIPW derived in Cattaneo (2010)
  - Cattaneo (2010) and Cattaneo, Drukker, and Holland (2013) call the AIPW estimator an efficient-influence function (EIF) estimator because EIF theory is what produces the augmentation term

- Abadie, Alberto and Guido W. Imbens. 2006. "Large sample properties of matching estimators for average treatment effects," *Econometrica*, 235–267.
- Angrist, J. D. and J.-S. Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton, NJ: Princeton University Press.
- Bang, Heejung and James M. Robins. 2005. "Doubly robust estimation in missing data and causal inference models," *Biometrics*, 61(4), 962–973.
- Cameron, A. Colin and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications*, Cambridge: Cambridge University Press.
- Cattaneo, Matias D., David M. Drukker, and Ashley D. Holland. 2013. "Estimation of multivalued treatment effects under conditional independence," *Stata Journal*, 13(3), 407–450.
- Cattaneo, M.D. 2010. "Efficient semiparametric estimation of multi-valued treatment effects under ignorability," *Journal of Econometrics*, 155(2), 138–154.
- Claeskens, Gerda and Nils Lid Hjort. 2008. *Model selection and model averaging*, Cambridge, UK: Cambridge University Press.

- Drukker, David M. 2014. "Quantile treatment effect estimation from censored data by regression adjustment," Tech. rep., Under review at the *Stata Journal*, <http://www.stata.com/ddrukker/mqgamma.pdf>.
- Heckman, James and Salvador Navarro-Lozano. 2004. "Using matching, instrumental variables, and control functions to estimate economic choice models," *Review of Economics and statistics*, 86(1), 30–57.
- Heckman, James J. 1997. "Instrumental variables: A study of implicit behavioral assumptions used in making program evaluations," *Journal of Human Resources*, 32(3), 441–462.
- Hirano, Keisuke, Guido W. Imbens, and Geert Ridder. 2003. "Efficient estimation of average treatment effects using the estimated propensity score," *Econometrica*, 71(4), 1161–1189.
- Holland, Paul W. 1986. "Statistics and causal inference," *Journal of the American Statistical Association*, 945–960.
- Horvitz, D. G. and D. J. Thompson. 1952. "A Generalization of Sampling Without Replacement From a Finite Universe," *Journal of the American Statistical Association*, 47(260), 663–685.

- Imai, Kosuke and Marc Ratkovic. 2014. "Covariate balancing and propensity score," *Journal of the Royal Statistical Society: Series B*, 76(1), 243–263.
- Imbens, Guido W. 2000. "The role of the propensity score in estimating dose-response functions," *Biometrika*, 87(3), 706–710.
- . 2004. "Nonparametric estimation of average treatment effects under exogeneity: A review," *Review of Economics and statistics*, 86(1), 4–29.
- Imbens, Guido W. and Jeffrey M. Wooldridge. 2009. "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature*, 47, 5–86.
- Lunceford, Jared K and Marie Davidian. 2004. "Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study," *Statistics in medicine*, 23(19), 2937–2960.
- Robins, James M. and Andrea Rotnitzky. 1995. "Semiparametric Efficiency in Multivariate Regression Models with Missing Data," *Journal of the American Statistical Association*, 90(429), 122–129.

- Robins, James M., Andrea Rotnitzky, and Lue Ping Zhao. 1994. "Estimation of Regression Coefficients When Some Regressors Are Not Always Observed," *Journal of the American Statistical Association*, 89(427), 846–866.
- . 1995. "Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data," *Journal of the American Statistical Association*, 90(429), 106–121.
- Rosenbaum, P. and D. Rubin. 1983. "Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41–55.
- Rubin, Donald B. 1974. "Estimating causal effects of treatments in randomized and nonrandomized studies." *Journal of educational Psychology*, 66(5), 688.
- Tsiatis, Anastasios A. 2006. *Semiparametric theory and missing data*, New York: Springer Verlag.
- Vittinghoff, E., D. V. Glidden, S. C. Shiboski, and C. E. McCulloch. 2012. *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models*, New York: Springer, 2 ed.



- Wooldridge, Jeffrey M. 2002. "Inverse probability weighted M-estimators for sample selection, attrition, and stratification," *Portuguese Economic Journal*, 1, 117–139.
- Wooldridge, Jeffrey M. 2005. "Violating ignorability of treatment by controlling for too many factors," *Econometric Theory*, 21(5), 1026.
- Wooldridge, Jeffrey M. 2007. "Inverse probability weighted estimation for general missing data problems," *Journal of Econometrics*, 141(2), 1281–1301.
- . 2010. *Econometric Analysis of Cross Section and Panel Data*, Cambridge, Massachusetts: MIT Press, second ed.