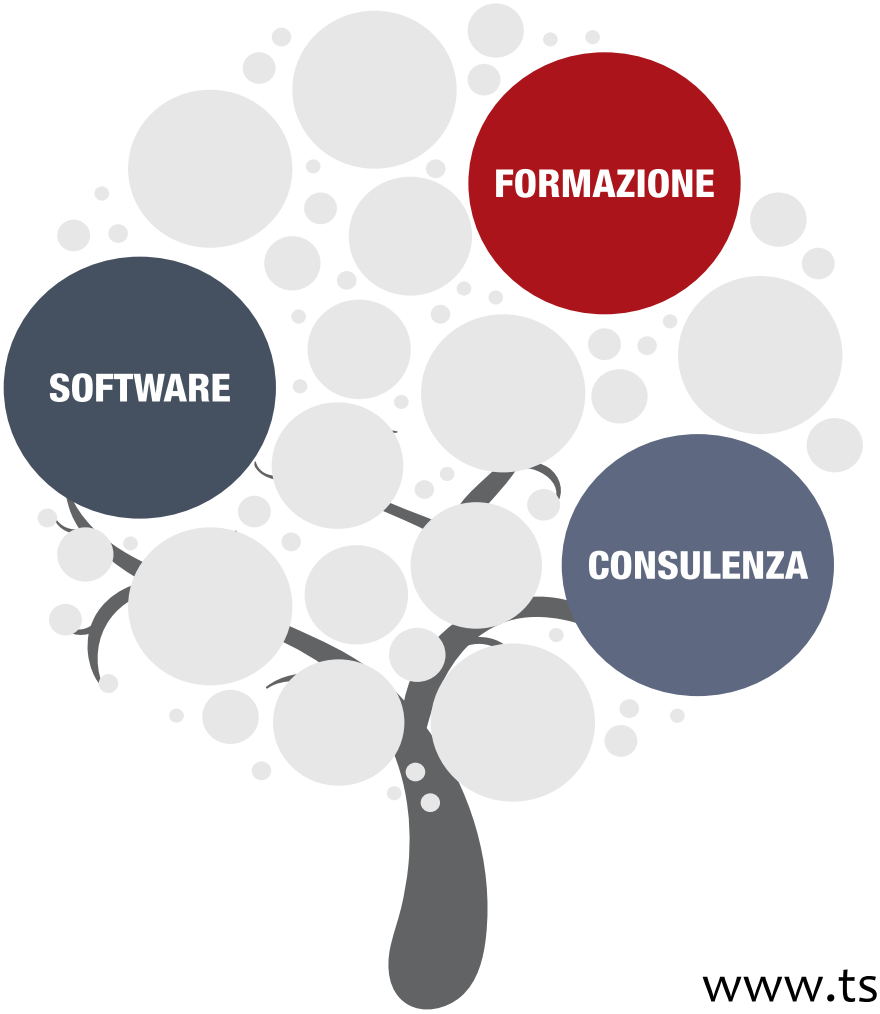
ESTIMATING SURVIVAL-TIME TREATMENT EFFECTS FROM OBSERVATIONAL DATA

David M. Drukker
Executive Director of Econometrics, StataCorp.

XII Italian Stata Users Group Meeting • Florence, 12 November 2015



Estimating survival-time treatment effects from observational data

David M. Drukker

Executive Director of Econometrics
Stata

Italian Stata Users Group meeting
12 November 2015

A question

- Is smoking bad for men who have already had a heart attack?
  - Too vague
- Will smoking reduce the time to a second heart-attack among men aged 45–55 who have already had a heart attack?
  - Less interesting, but more specific
  - There might even be data to help us answer this question
  - The data will be observational, not experimental
  - 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

The data

```
. use sheart2
(Time to second heart attack (fictional))
. describe
Contains data from sheart2.dta
  obs:         5,000                          Time to second heart attack
                                                (fictional)
 vars:             6                          11 Aug 2015 15:28
 size:       120,000
-------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
-------------------------------------------------------------------------------
age             float   %9.0g                 Age (in decades, demeaned)
exercise        float   %9.0g                 Exercise index
diet            float   %9.0g                 Diet index
smoke           float   %9.0g      lsmoke     Smoking indicator
fail            float   %9.0g      lfail      Failure indicator
atime           float   %9.0g                 Time to second attack
-------------------------------------------------------------------------------
Sorted by:
```


The data

```
. stset atime, failure(fail)
     failure event:  fail != 0 & fail < .
obs. time interval:  (0, atime]
 exit on or before:  failure
-------------------------------------------------------------------------------
       5000  total observations
          0  exclusions
-------------------------------------------------------------------------------
       5000  observations remaining, representing
       2969  failures in single-record/single-failure data
  10972.843  total analysis time at risk and under observation
                                                at risk from t =         0
                                     earliest observed entry t =         0
                                          last observed exit t =  40.96622
. save sheart2, replace
file sheart2.dta saved
```

- 2,969 of the 5,000 observations record actual time to a second heart attack; remainder were censored

A Cox model for the treatment

- Many researchers would start by fitting a Cox model

```
. stcox smoke age exercise diet , nolog noshow
Cox regression -- no ties
No. of subjects =        5,000                 Number of obs    =       5,000
No. of failures =        2,969
Time at risk    =  10972.84266
                                               LR chi2(4)       =      271.77
Log likelihood  =   -21963.163                 Prob > chi2      =      0.0000
```

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
smoke	1.540071	.0764791	8.70	0.000	1.397239	1.697505
age	2.024237	.1946491	7.33	0.000	1.676527	2.444062
exercise	.5473001	.0454893	-7.25	0.000	.465026	.6441304
diet	.4590354	.0379597	-9.42	0.000	.3903521	.5398037

- Smoking increases the hazard of a second heart attack by a factor of 1.5

A Cox model for the treatment

- The Cox model models the probability that the event will occur in the next moment given that it has not yet happened as a function of covariates
  - The probability that the event will occur in the next moment given that it has not yet happened and given covariates is known as the conditional hazard function denoted by $\lambda(t|\mathbf{x})$
  - The Cox model specifies that

$$\lambda(t|\mathbf{x}) = \lambda_0(t)\exp(\mathbf{x}\boldsymbol{\beta})$$

    and only estimates $\boldsymbol{\beta}$
- Leaving $\lambda_0(t)$ unspecified increases the flexibility of the model

A Cox model for the treatment

- Does the binary treatment `smoke` affect the time to second heart attack?
- The hazard ratio reported by `stcox` indicates that smoking raises the hazard of a second heart attack by a factor of 1.5 relative to not smoking

$$\frac{\lambda(t|\mathbf{x}, \texttt{smoke} = 1)}{\lambda(t|\mathbf{x}, \texttt{smoke} = 0)} = \frac{\lambda_0(t)\exp(\beta_{smoke} + \mathbf{x}_o\boldsymbol{\beta}_o)}{\lambda_0(t)\exp(\mathbf{x}_o\boldsymbol{\beta}_o)} = \exp(\beta_{smoke})$$

where $\mathbf{x}_0\boldsymbol{\beta}_o = \texttt{age}\beta_{age} + \texttt{exercise}\beta_{exercise} + \texttt{diet}\beta_{diet}$

The effect varies

```
. stcox ibn.smoke#c.(age exercise diet) , nolog noshow
Cox regression -- no ties
No. of subjects =        5,000                 Number of obs    =       5,000
No. of failures =        2,969
Time at risk    =  10972.84266
                                               LR chi2(6)       =      223.11
Log likelihood  =   -21987.493                 Prob > chi2      =      0.0000
```

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
smoke#c.age						
Nonsmoker	1.714749	.1751413	5.28	0.000	1.403655	2.094791
Smoker	3.979649	1.110035	4.95	0.000	2.303673	6.874936
smoke#c.exercise						
Nonsmoker	.5514891	.0476827	-6.88	0.000	.4655224	.6533309
Smoker	.2839313	.0822003	-4.35	0.000	.1609844	.5007752
smoke#c.diet						
Nonsmoker	.4461597	.0389598	-9.24	0.000	.3759769	.5294433
Smoker	.6908017	.1785842	-1.43	0.152	.416201	1.146578

- The ratio of the smoking hazard to the nonsmoking hazard varies by age, `exercise`, and `diet`

Problems with the Cox model

- Two problems with the Cox model
  1. It is hard to understand the units of the hazard ratio
     - How bad is it that smoking raises the hazard ratio by 1.5?
  2. This interpretation is only useful if the treatment enters the $\mathbf{x}\beta$ term linearly
     - If the treatment is interacted with other covariates, the effect of the treatment varies over individuals
- The average difference in time to second heart attack when everyone smokes instead of when no one smokes
  1. is easier to interpret
  2. is easier to estimate

Doctors versus policy analysts

- What can we do when the estimated effects vary over covariate values?
- When an effect varies over the values of other covariates, you can estimate the effect for a particular type of person or estimate a population-level effect
  - Doctors use covariate specific estimates
    (They ask you many questions to learn your covariates.)
  - Policy analysts need to account for the how a policy will effect different people in the population
    The discipline of the population distribution of the effects keeps them from picking winners or losers

Effects that vary over individuals

- For each individual, the effect of the treatment is a contrast of what would happen if the individual received the treatment versus what would happen if the individual did not receive the treatment
  - A potential outcome is the outcome an individual would receive if given a specific treatment level
  - For each treatment level, there is a potential outcome for each individual

```
. use sheart2_po
(Potential outcome time to second heart attack)
. list id atime_ns atime_s smoke atime in 21/25
```

	id	atime_ns	atime_s	smoke	atime
21.	21	1.44135	.7616374	Nonsmoker	1.44135
22.	22	1.422631	1.422631	Smoker	1.422631
23.	23	4.264108	.3285356	Nonsmoker	4.264108
24.	24	1.533371	1.246619	Nonsmoker	1.533371
25.	25	.1929609	.1929609	Nonsmoker	.1929609

Ratio of unconditional hazards

- The hazard-ratio measure of the treatment effect is the ratio of the hazard of the smoking potential outcome to the hazard nonsmoking potential outcome
  - The hazard-ratio measure of the treatment effect is the ratio of the hazard from the distribution when everyone smokes to the hazard from the distribution when no one smokes
  - This ratio hazards of unconditional distributions is not the same as an average of conditional hazard ratios (See Appendix 1)

Average treatment effect

- Ratios of unconditional hazards are harder to estimate and more difficult to interpret than the average difference in time to second heart attack when everyone smokes instead of no one smokes
  - The average difference in time to second heart attack when everyone smokes instead of no one smokes is an average treatment effect (ATE)
  - $ATE = \mathbf{E}[t_i(\texttt{smoke}) - t_i(\texttt{notsmoke})]$
    $t_i(\texttt{smoke})$ is the time to event when person i smokes
    and
    $t_i(\texttt{notsmoke})]$ is the time to event when person i does not smoke
- The ATE provides a measure of the effect in the units of time in which the time to event is measured
  - In our example, the ATE is measured in years

Average treatment effect

- Recall that one of the two potential outcomes is always missing

```
. use sheart2_po
(Potential outcome time to second heart attack)
. list id atime_ns atime_s smoke atime in 21/25
```

	id	atime_ns	atime_s	smoke	atime
21.	21	1.44135	.7616374	Nonsmoker	1.44135
22.	22	1.422631	1.422631	Smoker	1.422631
23.	23	4.264108	.3285356	Nonsmoker	4.264108
24.	24	1.533371	1.246619	Nonsmoker	1.533371
25.	25	.1929609	.1929609	Nonsmoker	.1929609

- Potential outcomes are the data that we wish we had to estimate causal treatment effects
- Estimating treatment effects can be viewed as a missing-data problem

Average treatment effect

- If we had data on each potential outcome, the average difference in the (observed) potential outcomes would estimate the population average treatment effect
- The average of a potential outcome in the population is known as the potential-outcome mean (POM) for a treatment level
  - The ATE is a difference in POMs

$$\begin{aligned} ATE &= POM_{smoke} - POM_{nonsmoke} \\ &= \mathbf{E}[t_i(\texttt{smoke})] - \mathbf{E}[t_i(\texttt{notsmoke})] \end{aligned}$$

$t_i(\texttt{smoke})$ is the time to event when person i smokes
and
$t_i(\texttt{notsmoke})$ is the time to event when person i does not smoke

Missing data

- The "fundamental problem of causal inference" (Holland (1986)) is that we only observe one of the potential outcomes
- 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

- If smoking were randomly assigned, the missing potential outcome would be missing completely at random
  - If the time to second heart attack was never censored and smoking was randomly assigned
    1. The average time to second heart attack among smokers would estimate the smoking POM
    2. The average time to second heart attack among nonsmokers would estimate the nonsmoking POM
    3. The difference in these estimated POMs would estimate the ATE

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 (CMI) which says that after conditioning on covariates, the treatment does not affect the means of the potential outcomes

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
  - The auxiliary model also handles the data lost to censoring

Model		Estimator
outcome	$\longrightarrow$	Regression adjustment (RA)
treatment	$\longrightarrow$	Inverse-probability weighted (IPW)
outcome and treatment	$\longrightarrow$	IPW RA (IPWRA)

Regression adjustment estimators

- Regression adjustment (RA) estimators use predicted values from the model for the time to event to solve the missing-data problems
- RA estimators estimate the parameters of separate survival models for the outcome for each treatment level, then
  - The mean of the predicted times to second heart attack using the estimated coefficients from the model for smokers and all the observations estimates the smoking POM
  - The mean of the predicted times to second heart attack using the estimated coefficients from the model for nonsmokers and all the observations estimates the nonsmoking POM
  - The difference between the estimated smoking POM and the estimated nonsmoking POM estimates the ATE
  - Censoring is handled in the log likelihood functions of the survival models

```
. use sheart2
(Time to second heart attack (fictional))
. stteffects ra (age exercise diet) (smoke), nolog noshow
Survival treatment-effects estimation           Number of obs     =      5,000
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)	-1.520671	.2011014	-7.56	0.000	-1.914822	-1.126519
POmean						
smoke Nonsmoker	4.057439	.1028462	39.45	0.000	3.855864	4.259014

- The average time to second heart attack is 1.5 years sooner when everyone in the population smokes instead of no one smokes
- The average time to second heart attack is 4.1 years when no one smokes


```
. stteffects ra (age exercise diet, gamma) (smoke), nolog noshow
Survival treatment-effects estimation           Number of obs     =      5,000
Estimator      : regression adjustment
Outcome model  : gamma
Treatment model: none
Censoring model: none
```

_t	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ATE						
smoke (Smoker vs Nonsmoker)	-1.616514	.177703	-9.10	0.000	-1.964805	-1.268222
POmean						
smoke Nonsmoker	4.014823	.0988662	40.61	0.000	3.821049	4.208598

- Can model the outcome using either a gamma, exponential, or log normal distribution instead of the default Weibull distribution
- Can model the ancillary distribution parameters using `ancillary()` option

Inverse-probability-weighted estimators

- Inverse-probability-weighted (IPW) estimators:
  - IPW estimators weight observations on the observed 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

Inverse-probability-weighted estimators

- IPW estimators use estimates from models for the probability of treatment and the probability of censoring to correct for the missing potential outcome and the observations lost to censoring
- In contrast, RA estimators model the outcome without any assumptions about the functional form for the probability of treatment model
  - RA estimators handle censoring in the log likelihood function
  - Handling censoring in the log likelihood function allows for fixed censoring times
- IPW estimators have a long history in statistics, biostatistics, and econometrics
  - 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)

```
. stteffects ipw (smoke age exercise diet) (age exercise diet), nolog noshow
Survival treatment-effects estimation           Number of obs     =      5,000
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)	-1.689397	.3373219	-5.01	0.000	-2.350536	-1.028258
POmean						
smoke Nonsmoker	4.200135	.2156737	19.47	0.000	3.777423	4.622848

- The average time to second heart attack is 1.7 years sooner when everyone in the population smokes instead of no one smokes
- The average time to second heart attack is 4.2 years when no one smokes


```
. stteffects ipw (smoke age exercise diet, logit)            ///
>               (age exercise diet, gamma), nolog noshow
Survival treatment-effects estimation           Number of obs     =      5,000
Estimator      : inverse-probability weights
Outcome model  : weighted mean
Treatment model: logit
Censoring model: gamma
```

_t	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ATE						
smoke (Smoker vs Nonsmoker)	-1.922143	.4502077	-4.27	0.000	-2.804534	-1.039752
POmean						
smoke Nonsmoker	4.555551	.3345953	13.62	0.000	3.899756	5.211345

- Can model treatment by probit, logit, or heteroskedastic probit
- Can model censoring by Weibull, gamma, or log normal
  Can model ancillary parameters

Combining IPW and RA

- Inverse-probability-weighted regression-adjustment (IPWRA) estimators combine models for the outcome and the treatment to get more efficient estimates
- IPWRA estimators use the inverse of the estimated treatment-probability weights to estimate missing-data-corrected regression coefficients that are subsequently used to estimate the POMs
  - The ATE is estimated by a difference in the estimated POMs
- Censoring can be handled in the log likelihood function or by modeling the censoring process
  - Handling censoring in the log likelihood function allows for fixed censoring times
- See Wooldridge (2007) and (Wooldridge, 2010, section 21.3.4)


```
. steffects ipwra (age exercise diet) (smoke age exercise diet) , nolog noshow
Survival treatment-effects estimation           Number of obs     =      5,000
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)	-1.543315	.2027738	-7.61	0.000	-1.940744	-1.145885
POmean						
smoke Nonsmoker	4.064291	.1032385	39.37	0.000	3.861947	4.266634

- The average time to second heart attack is 1.5 years sooner when everyone in the population smokes instead of no one smokes
- The average time to second heart attack is 4.1 years when no one smokes

```
. steffects ipwra (age exercise diet)                    ///
>                 (smoke age exercise diet)              ///
>                 (age exercise diet)       , nolog noshow
Survival treatment-effects estimation           Number of obs     =      5,000
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)	-1.782505	.3091845	-5.77	0.000	-2.388495	-1.176514
POmean						
smoke Nonsmoker	4.233607	.2185565	19.37	0.000	3.805244	4.661969

- This example models the censoring process instead handling it in the log likelihood function for the outcome
- Additional model choices as for RA and IPW estimators

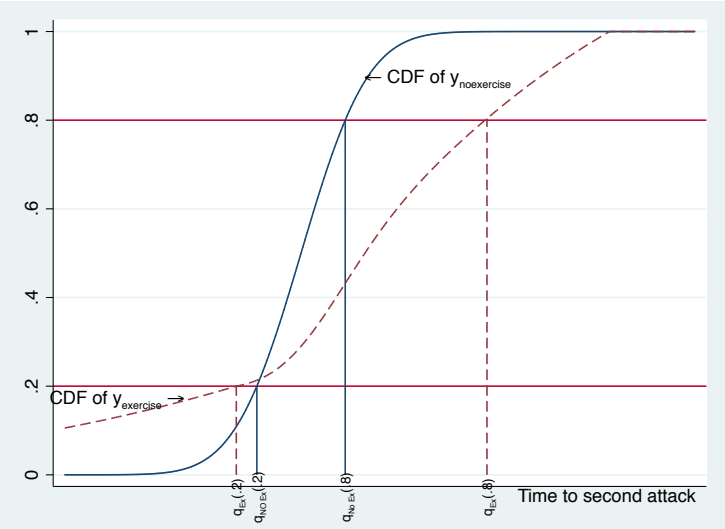
QTEs for survival data

- 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 that 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 τ(th) quantile of y_1 minus the τ(th) quantile of y_0 is not the same as the τ(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 $\boldsymbol{\theta}_1$ parameters of $F(y|\mathbf{x}, t = 1, \boldsymbol{\theta}_1)$ 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 $\boldsymbol{\theta}_0$ parameters of $F(y|\mathbf{x}, t = 0, \boldsymbol{\theta}_0)$
- At the point y,

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

  estimates the marginal distribution of the treated potential outcome
- The $\widehat{q}_{1,.75}$ that solves

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

A regression-adjustment estimator for QTEs

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

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

  estimates the .75 marginal quantile for the control potential outcome
- $\widehat{q}_1(.75) - \widehat{q}_0(.75)$ 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

Appendix 1: Ratio of unconditional hazards

- The ratio hazards of unconditional (marginal) distributions is not the same as an average of conditional hazard ratio

$$\frac{\lambda_{smoke}(t)}{\lambda_{nonsmoke}(t)} = \frac{\frac{f_{smoke}(t)}{S_{smoke}(t)}}{\frac{f_{nonsmoke}(t)}{S_{nonsmoke}(t)}} \neq \mathbf{E}\left[\frac{\lambda_{smoke}(t|\mathbf{x}\beta_{smoke})}{\lambda_{nonsmoke}(t|\mathbf{x}\beta_{nonsmoke})}\right]$$

$\lambda_{smoke}(t)$	is the unconditional hazard when everyone smokes
$\lambda_{nonsmoke}(t)$	is the unconditional hazard when no one smokes
$f_{smoke}(t)$	is the unconditional density when everyone one smokes
$f_{nonsmoke}(t)$	is the unconditional density when no one smokes
$S_{smoke}(t)$	is the unconditional survival function when everyone smokes
$S_{nonsmoke}(t)$	is the unconditional survival function when no one smokes

Appendix 2: Why robust standard errors?

- Have a multistep estimator
  1. Example based on RA, same logic works for IPW and IPWRA
  2. Model outcome conditional on covariates for treated observations
  3. Model outcome conditional on covariates for not treated observations
  4. Estimate predicted mean survival time of all observations given covariates from treated model estimates
  5. Estimate predicted mean survival time of all observations given covariates from not-treated model estimates
  6. Difference in means of predicted means estimates ATE

Appendix 2: Why robust standard errors?

- 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 $D = \left(\frac{1}{N}\frac{\partial m(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^{-1}$ 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

Cameron, A. Colin and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and applications*, Cambridge: 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 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.

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.

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.

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.

Wooldridge, Jeffrey M. 2002. "Inverse probability weighted M-estimators for sample selection, attrition, and stratification," *Portuguese Economic Journal*, 1, 117–139.

———. 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.