Title

stata.com

teffects ipwra — Inverse-probability-weighted regression adjustment

Description

`teffects ipwra` estimates the average treatment effect (ATE), the average treatment effect on the treated (ATET), and the potential-outcome means (POMs) from observational data by inverse-probability-weighted regression adjustment (IPWRA). IPWRA estimators use weighted regression coefficients to compute averages of treatment-level predicted outcomes, where the weights are the estimated inverse probabilities of treatment. The contrasts of these averages estimate the treatment effects. IPWRA estimators have the double-robust property. `teffects ipwra` accepts a continuous, binary, count, fractional, or nonnegative outcome and allows a multivalued treatment.

See [CAUSAL] **teffects intro** or [CAUSAL] **teffects intro advanced** for more information about estimating treatment effects from observational data.

Quick start

ATE of binary treatment `treat2` estimated by IPWRA using a linear model for outcome `y1` on `x1` and `x2` and a logistic model for `treat2` on `x1` and `w`

```
teffects ipwra (y1 x1 x2) (treat2 x1 w)
```

Same as above, but estimate the ATET

```
teffects ipwra (y1 x1 x2) (treat2 x1 w), atet
```

Probit model for binary outcome `y3`

```
teffects ipwra (y3 x1 x2, probit) (treat2 x1 w)
```

Same as above, but use a heteroskedastic probit model for `y3` and a probit model for `treat2`

```
teffects ipwra (y3 x1 x2, hetprobit(x1 x2)) (treat2 x1 w, probit)
```

Same as above, but use a fractional heteroskedastic probit model for `y4` and a probit model for `treat2`

```
teffects ipwra (y4 x1 x2, fhetprobit(x1 x2)) (treat2 x1 w, probit)
```

ATE for each level of a three-valued treatment `treat3`

```
teffects ipwra (y1 x1 x2) (treat3 x1 w)
```

Same as above, and specify that `treat3` = 3 is the control level using the value label "MyControl" for 3

```
teffects ipwra (y1 x1 x2) (treat3 x1 w), control(MyControl)
```

Menu

Statistics > Causal inference/treatment effects > Continuous outcomes > Regression adjustment with IPW

Statistics > Causal inference/treatment effects > Binary outcomes > Regression adjustment with IPW

Statistics > Causal inference/treatment effects > Count outcomes > Regression adjustment with IPW

Statistics > Causal inference/treatment effects > Fractional outcomes > Regression adjustment with IPW

Statistics > Causal inference/treatment effects > Nonnegative outcomes > Regression adjustment with IPW

Syntax

```
teffects ipwra (ovar omvarlist [ , omodel noconstant])
  (tvar tmvarlist [ , tmodel noconstant]) [if] [in] [weight]
  [ , stat options]
```

ovar is a binary, count, continuous, fractional, or nonnegative outcome of interest.

omvarlist specifies the covariates in the outcome model.

tvar must contain integer values representing the treatment levels.

tmvarlist specifies the covariates in the treatment-assignment model.

omodel	Description
Model	
`linear`	linear outcome model; the default
`logit`	logistic outcome model
`probit`	probit outcome model
`hetprobit(`*varlist*`)`	heteroskedastic probit outcome model
`poisson`	exponential outcome model
`flogit`	fractional logistic outcome model
`fprobit`	fractional probit outcome model
`fhetprobit(`*varlist*`)`	fractional heteroskedastic probit outcome model

omodel specifies the model for the outcome variable.

tmodel	Description
Model	
`logit`	logistic treatment model; the default
`probit`	probit treatment model
`hetprobit(`*varlist*`)`	heteroskedastic probit treatment model

tmodel specifies the model for the treatment variable.
For multivalued treatments, only `logit` is available and multinomial logit is used.

stat	Description
Stat	
`ate`	estimate average treatment effect in population; the default
`atet`	estimate average treatment effect on the treated
`pomeans`	estimate potential-outcome means

options	Description
SE/Robust	
`vce(`*vcetype*`)`	*vcetype* may be `robust`, `cluster` *clustvar*, `bootstrap`, or `jackknife`
Reporting	
`level(#)`	set confidence level; default is `level(95)`
`aequations`	display auxiliary-equation results
display_options	control columns and column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling
Maximization	
maximize_options	control the maximization process; seldom used
Advanced	
`pstolerance(#)`	set tolerance for overlap assumption
`osample(`*newvar*`)`	*newvar* identifies observations that violate the overlap assumption
`control(#` | *label*`)`	specify the level of *tvar* that is the control
`tlevel(#` | *label*`)`	specify the level of *tvar* that is the treatment
`coeflegend`	display legend instead of statistics

omvarlist and *tmvarlist* may contain factor variables; see [U] **11.4.3 Factor variables**.
`bootstrap`, `by`, `collect`, `jackknife`, and `statsby` are allowed; see [U] **11.1.10 Prefix commands**.
Weights are not allowed with the `bootstrap` prefix; see [R] **bootstrap**.
`fweights`, `iweights`, and `pweights` are allowed; see [U] **11.1.6 weight**.
`coeflegend` does not appear in the dialog box.
See [U] **20 Estimation and postestimation commands** for more capabilities of estimation commands.

Options

Model

`noconstant`; see [R] **Estimation options**.

Stat

stat is one of three statistics: `ate`, `atet`, or `pomeans`. `ate` is the default.

`ate` specifies that the average treatment effect be estimated.

`atet` specifies that the average treatment effect on the treated be estimated.

`pomeans` specifies that the potential-outcome means for each treatment level be estimated.

SE/Robust

`vce(`*vcetype*`)` specifies the type of standard error reported, which includes types that are robust to some kinds of misspecification (`robust`), that allow for intragroup correlation (`cluster` *clustvar*), and that use bootstrap or jackknife methods (`bootstrap`, `jackknife`); see [R] ***vce_option***.

Reporting

`level(#)`; see [R] **Estimation options**.

`aequations` specifies that the results for the outcome-model or the treatment-model parameters be displayed. By default, the results for these auxiliary parameters are not displayed.

display_options: `noci`, `nopvalues`, `noomitted`, `vsquish`, `noemptycells`, `baselevels`, `allbaselevels`, `nofvlabel`, `fvwrap(#)`, `fvwrapon(`*style*`)`, `cformat(%`*fmt*`)`, `pformat(%`*fmt*`)`, `sformat(%`*fmt*`)`, and `nolstretch`; see [R] **Estimation options**.

Maximization

maximize_options: `iterate(#)`, `[no]log`, and `from(`*init_specs*`)`; see [R] **Maximize**. These options are seldom used.

init_specs is one of

matname `[ , skip copy ]`

`[ , # ... ], copy`

Advanced

`pstolerance(#)` specifies the tolerance used to check the overlap assumption. The default value is `pstolerance(1e-5)`. `teffects` will exit with an error if an observation has an estimated propensity score smaller than that specified by `pstolerance()`.

`osample(`*newvar*`)` specifies that indicator variable *newvar* be created to identify observations that violate the overlap assumption.

`control(#|`*label*`)` specifies the level of *tvar* that is the control. The default is the first treatment level. You may specify the numeric level *#* (a nonnegative integer) or the label associated with the numeric level. `control()` may not be specified with statistic `pomeans`. `control()` and `tlevel()` may not specify the same treatment level.

`tlevel(#|`*label*`)` specifies the level of *tvar* that is the treatment for the statistic `atet`. The default is the second treatment level. You may specify the numeric level *#* (a nonnegative integer) or the label associated with the numeric level. `tlevel()` may only be specified with statistic `atet`. `tlevel()` and `control()` may not specify the same treatment level.

The following option is available with `teffects ipwra` but is not shown in the dialog box:

`coeflegend`; see [R] **Estimation options**.

Remarks and examples

stata.com

Remarks are presented under the following headings:

Overview

IPWRA estimators use probability weights to obtain outcome-regression parameters that account for the missing-data problem arising from the fact that each subject is observed in only one of the potential outcomes. The adjusted outcome-regression parameters are used to compute averages of treatment-level predicted outcomes. The contrasts of these averages provide estimates of the treatment effects.

IPWRA estimators use a model to predict treatment status, and they use another model to predict outcomes. Because IPWRA estimators have the double-robust property, only one of the two models must be correctly specified for the IPWRA estimator to be consistent.

IPWRA estimators use a three-step approach to estimating treatment effects:

1. They estimate the parameters of the treatment model and compute inverse-probability weights.
2. Using the estimated inverse-probability weights, they fit weighted regression models of the outcome for each treatment level and obtain the treatment-specific predicted outcomes for each subject.
3. They compute the means of the treatment-specific predicted outcomes. The contrasts of these averages provide the estimates of the ATEs. By restricting the computations of the means to the subset of treated subjects, we can obtain the ATETs.

These steps produce consistent estimates of the effect parameters because the treatment is assumed to be independent of the potential outcomes after conditioning on the covariates. The overlap assumption ensures that predicted inverse-probability weights do not get too large. The standard errors reported by `teffects ipwra` correct for the three-step process. See [CAUSAL] **teffects intro** or [CAUSAL] **teffects intro advanced** for more information about this estimator.

We will illustrate the use of `teffects ipwra` by using data from a study of the effect of a mother's smoking status during pregnancy (`mbsmoke`) on infant birthweight (`bweight`) as reported by Cattaneo (2010). This dataset also contains information about each mother's age (`mage`), education level (`medu`), marital status (`mmarried`), whether the first prenatal exam occurred in the first trimester (`prenatal1`), and whether this baby was the mother's first birth (`fbaby`).

▷ Example 1: Estimating the ATE

We begin by using `teffects ipwra` to estimate the average treatment effect of smoking on birthweight. We will use a probit model to predict treatment status as a function of `mmarried`, `mage`, and `fbaby`; to maximize the predictive power of this model, we use factor-variable notation to incorporate quadratic effects of the mother's age, the only continuous covariate in our model. We will use linear regression (the default) to model birthweight, using `prenatal1`, `mmarried`, `mage`, and `fbaby` as explanatory variables. We type

```
. use https://www.stata-press.com/data/r18/cattaneo2
(Excerpt from Cattaneo (2010) Journal of Econometrics 155: 138-154)
. teffects ipwra (bweight prenatal1 mmarried mage fbaby)
> (mbsmoke mmarried c.mage##c.mage fbaby medu, probit)
Iteration 0:  EE criterion =  9.885e-21
Iteration 1:  EE criterion =  7.847e-26
Treatment-effects estimation                    Number of obs     =      4,642
Estimator      : IPW regression adjustment
Outcome model  : linear
Treatment model: probit
```

bweight	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
mbsmoke (Smoker vs Nonsmoker)	-229.9671	26.62668	-8.64	0.000	-282.1544	-177.7798
POmean						
mbsmoke Nonsmoker	3403.336	9.57126	355.58	0.000	3384.576	3422.095

The average birthweight if all mothers were to smoke would be 230 grams less than the average of 3,403 grams that would occur if none of the mothers had smoked.

◁

By default, `teffects ipwra` displays the ATE and untreated POM. We can specify the `pomeans` option to display both the treated and untreated POMs, and we can use the `aequations` option to display the regression model coefficients used to predict the POMs as well as the coefficients from the model used to predict treatment.

Example 2: Displaying the POMs and equations

```
. use https://www.stata-press.com/data/r18/cattaneo2
(Excerpt from Cattaneo (2010) Journal of Econometrics 155: 138-154)
. teffects ipwra (bweight prenatal1 mmarried mage fbaby)
> (mbsmoke mmarried c.mage##c.mage fbaby medu, probit), pomeans aequations
Iteration 0:  EE criterion =  9.885e-21
Iteration 1:  EE criterion =  6.922e-26
Treatment-effects estimation                    Number of obs     =      4,642
Estimator      : IPW regression adjustment
Outcome model  : linear
Treatment model: probit
```

bweight	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
POmeans						
mbsmoke						
Nonsmoker	3403.336	9.57126	355.58	0.000	3384.576	3422.095
Smoker	3173.369	24.86997	127.60	0.000	3124.624	3222.113
OME0						
prenatal1	67.98549	28.78428	2.36	0.018	11.56933	124.4017
mmarried	155.5893	26.46903	5.88	0.000	103.711	207.4677
mage	2.893051	2.134788	1.36	0.175	-1.291056	7.077158
fbaby	-71.9215	20.39317	-3.53	0.000	-111.8914	-31.95162
_cons	3194.808	55.04911	58.04	0.000	3086.913	3302.702
OME1						
prenatal1	34.76923	43.18534	0.81	0.421	-49.87248	119.4109
mmarried	124.0941	40.29775	3.08	0.002	45.11193	203.0762
mage	-5.068833	5.954425	-0.85	0.395	-16.73929	6.601626
fbaby	39.89692	56.82072	0.70	0.483	-71.46966	151.2635
_cons	3175.551	153.8312	20.64	0.000	2874.047	3477.054
TME1						
mmarried	-.6484821	.0554173	-11.70	0.000	-.757098	-.5398663
mage	.1744327	.0363718	4.80	0.000	.1031452	.2457202
c.mage#						
c.mage	-.0032559	.0006678	-4.88	0.000	-.0045647	-.0019471
fbaby	-.2175962	.0495604	-4.39	0.000	-.3147328	-.1204595
medu	-.0863631	.0100148	-8.62	0.000	-.1059917	-.0667345
_cons	-1.558255	.4639691	-3.36	0.001	-2.467618	-.6488926

As is well known, the standard probit model assumes that the error terms in the latent-utility framework are homoskedastic; the model is not robust to departures from this assumption. An alternative is to use the heteroskedastic probit model, which explicitly models the error variance as a function of a set of variables.

Example 3: Heteroskedastic probit treatment model

Here we use the variables as before, but we use a heteroskedastic probit model to predict treatment status, modeling the heteroskedasticity as a quadratic function of the mother's age:

```
. teffects ipwra (bweight prenatal1 mmarried fbaby c.mage)
> (mbsmoke mmarried c.mage##c.mage fbaby medu, hetprobit(c.mage##c.mage)),
> aequations
Iteration 0:  EE criterion =  3.904e-10
Iteration 1:  EE criterion =  4.647e-11
Treatment-effects estimation                    Number of obs     =      4,642
Estimator      : IPW regression adjustment
Outcome model  : linear
Treatment model: heteroskedastic probit
```

bweight	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
mbsmoke (Smoker vs Nonsmoker)	-229.6322	26.33451	-8.72	0.000	-281.2468	-178.0175
POmean						
mbsmoke Nonsmoker	3403.74	9.545798	356.57	0.000	3385.03	3422.449
OME0						
prenatal1	64.95123	28.6216	2.27	0.023	8.853921	121.0485
mmarried	154.2297	26.45867	5.83	0.000	102.3717	206.0878
fbaby	-71.61131	20.33774	-3.52	0.000	-111.4725	-31.75007
mage	3.010149	2.133812	1.41	0.158	-1.172046	7.192344
_cons	3195.355	55.05451	58.04	0.000	3087.45	3303.26
OME1						
prenatal1	38.55275	43.57023	0.88	0.376	-46.84334	123.9488
mmarried	126.3378	40.7398	3.10	0.002	46.48921	206.1863
fbaby	45.43551	56.44827	0.80	0.421	-65.20106	156.0721
mage	-6.069917	5.952507	-1.02	0.308	-17.73662	5.596783
_cons	3195.795	152.3978	20.97	0.000	2897.101	3494.489
TME1						
mmarried	-.0295517	.0238767	-1.24	0.216	-.0763492	.0172459
mage	.015789	.0105411	1.50	0.134	-.0048711	.0364492
c.mage# c.mage	-.0002837	.0001899	-1.49	0.135	-.000656	.0000886
fbaby	-.0093304	.0079946	-1.17	0.243	-.0249996	.0063387
medu	-.0036772	.0030294	-1.21	0.225	-.0096147	.0022602
_cons	-.1822172	.117975	-1.54	0.122	-.413444	.0490095
TME1_lnsigma						
mage	-.2211507	.0631072	-3.50	0.000	-.3448385	-.097463
c.mage# c.mage	.0037613	.001243	3.03	0.002	.0013251	.0061976

The estimated ATE and base-level POM are essentially the same as those produced by the model that used a homoskedastic probit.

◁

Video example

Treatment effects: Inverse-probability-weighted regression adjustment

Stored results

`teffects ipwra` stores the following in `e()`:

Scalars

`e(N)`	number of observations
`e(n`*j*`)`	number of observations for treatment level *j*
`e(N_clust)`	number of clusters
`e(k_eq)`	number of equations in `e(b)`
`e(k_levels)`	number of levels in treatment variable
`e(treated)`	level of treatment variable defined as treated
`e(control)`	level of treatment variable defined as control
`e(converged)`	`1` if converged, `0` otherwise

Macros

`e(cmd)`	`teffects`
`e(cmdline)`	command as typed
`e(depvar)`	name of outcome variable
`e(tvar)`	name of treatment variable
`e(subcmd)`	`ipwra`
`e(tmodel)`	`logit`, `probit`, or `hetprobit`
`e(omodel)`	`linear`, `logit`, `probit`, `hetprobit`, `poisson`, `flogit`, `fprobit`, or `fhetprobit`
`e(stat)`	statistic estimated, `ate`, `atet`, or `pomeans`
`e(wtype)`	weight type
`e(wexp)`	weight expression
`e(title)`	title in estimation output
`e(clustvar)`	name of cluster variable
`e(tlevels)`	levels of treatment variable
`e(vce)`	*vcetype* specified in `vce()`
`e(vcetype)`	title used to label Std. err.
`e(properties)`	`b V`
`e(estat_cmd)`	program used to implement `estat`
`e(predict)`	program used to implement `predict`
`e(marginsnotok)`	predictions disallowed by `margins`
`e(asbalanced)`	factor variables `fvset` as `asbalanced`
`e(asobserved)`	factor variables `fvset` as `asobserved`

Matrices

`e(b)`	coefficient vector
`e(V)`	variance–covariance matrix of the estimators

Functions

`e(sample)`	marks estimation sample

In addition to the above, the following is stored in `r()`:

Matrices

`r(table)`	matrix containing the coefficients with their standard errors, test statistics, p-values, and confidence intervals

Note that results stored in `r()` are updated when the command is replayed and will be replaced when any r-class command is run after the estimation command.

Methods and formulas

`teffects ipwra` implements a smooth treatment-effects estimator. All smooth treatment-effects estimators are documented in *Methods and formulas* of [CAUSAL] **teffects aipw**.

References

Cattaneo, M. D. 2010. Efficient semiparametric estimation of multi-valued treatment effects under ignorability. *Journal of Econometrics* 155: 138–154. https://doi.org/10.1016/j.jeconom.2009.09.023.

Huber, C. 2015. Introduction to treatment effects in Stata: Part 1. *The Stata Blog: Not Elsewhere Classified.* http://blog.stata.com/2015/07/07/introduction-to-treatment-effects-in-stata-part-1/.

Also see

[CAUSAL] **teffects postestimation** — Postestimation tools for teffects

[CAUSAL] **teffects** — Treatment-effects estimation for observational data

[CAUSAL] **teffects aipw** — Augmented inverse-probability weighting+

[U] **20 Estimation and postestimation commands**