Title

stata.com

teffects postestimation — Postestimation tools for teffects

Postestimation commands predict Remarks and examples Also see

Postestimation commands

The following postestimation command is of special interest after `teffects`:

Command	Description
`teoverlap`	overlap plots
`tebalance`	check balance of covariates

The following standard postestimation commands are also available:

Command	Description
`estat summarize`	summary statistics for the estimation sample
`estat vce`	variance–covariance matrix of the estimators (VCE)
`estimates`	cataloging estimation results
`etable`	table of estimation results
`hausman`	Hausman's specification test
`lincom`	point estimates, standard errors, testing, and inference for linear combinations of coefficients
`nlcom`	point estimates, standard errors, testing, and inference for nonlinear combinations of coefficients
`predict`	treatment effects, conditional means at treatment, propensity scores, etc.
`predictnl`	point estimates, standard errors, testing, and inference for generalized predictions
`test`	Wald tests of simple and composite linear hypotheses
`testnl`	Wald tests of nonlinear hypotheses

predict

Description for predict

`predict` creates a new variable containing predictions such as treatment effects, potential outcomes, conditional means, propensity scores, linear predictions, nearest-neighbor distances, and log square root of latent variances.

Menu for predict

Statistics > Postestimation

Syntaxes for predict

Syntaxes are presented under the following headings:

Syntax for predict after aipw and ipwra

```
predict [type] {stub*|newvar|newvarlist} [if] [in]
        [, statistic tlevel(treat_level)]

predict [type] stub* [if] [in], scores
```

statistic	Description
Main	
`te`	treatment effect; the default
`cmean`	conditional mean at treatment level
`ps`	propensity score
`xb`	linear prediction
`psxb`	linear prediction for propensity score
`lnsigma`	log square root of conditional latent variance (for outcome model `hetprobit()`) at treatment level
`pslnsigma`	log square root of latent variance (for treatment model `hetprobit()`)

If you do not specify `tlevel()` and only specify one new variable, `te` and `psxb` assume `tlevel()` specifies the first noncontrol treatment level.

If you do not specify `tlevel()` and only specify one new variable, `cmean`, `ps`, `xb`, and `lnsigma` assume `tlevel()` specifies the first treatment level.

You specify one or t new variables with `cmean`, `ps`, `xb`, and `lnsigma`, where t is the number of treatment levels.

You specify one or $t-1$ new variables with `te`, `psxb`, and `pslnsigma`.

Syntax for predict after ipw

```
predict [type] {stub*|newvar|newvarlist} [if] [in]
        [, statistic tlevel(treat_level)]

predict [type] stub* [if] [in], scores
```

statistic	Description
Main	
ps	propensity score; the default
xb	linear prediction for the propensity score
lnsigma	log square root of latent variance (for treatment model hetprobit())

If you do not specify tlevel() and only specify one new variable, ps assumes tlevel() specifies the first treatment level.

If you do not specify tlevel() and only specify one new variable, xb assumes tlevel() specifies the first noncontrol treatment level.

You specify one or t new variables with ps, where t is the number of treatment levels.

You specify one or $t-1$ new variables with xb and lnsigma.

Syntax for predict after nnmatch and psmatch

```
predict [type] {stub*|newvarlist} [, statistic tlevel(treat_level)]
```

statistic	Description
Main	
te	treatment effect; the default
po	potential outcome
distance	nearest-neighbor distance
ps	propensity score (psmatch only)
lnsigma	log square root of latent variance (for treatment model hetprobit())

These statistics are available for the estimation sample only and require the estimation option generate(*stub*). This is because of the nonparametric nature of the matching estimator.

If you do not specify tlevel() and only specify one new variable, po and ps assume tlevel() specifies the first treatment level.

You specify one new variable with te and lnsigma.

You specify one or two new variables with po and ps.

Syntax for predict after ra

```
predict [type] {stub* | newvar | newvarlist} [if] [in]
        [, statistic tlevel(treat_level)]

predict [type] stub* [if] [in], scores
```

statistic	Description
Main	
te	treatment effect; the default
cmean	conditional mean at treatment level
xb	linear prediction
lnsigma	log square root of conditional latent variance (for outcome model hetprobit()) at treatment level

If you do not specify tlevel() and only specify one new variable, te assumes tlevel() specifies the first noncontrol treatment level.

If you do not specify tlevel() and only specify one new variable, cmean, xb, and lnsigma assume tlevel() specifies the first treatment level.

You specify one or t new variables with cmean, xb, and lnsigma, where t is the number of treatment levels.

You specify one or $t-1$ new variables with te.

Options for predict

Options are presented under the following headings:

Options for predict after aipw and ipwra
Options for predict after ipw
Options for predict after nnmatch and psmatch
Options for predict after ra

Options for predict after aipw and ipwra

Main

te, the default, calculates the treatment effect for each noncontrol treatment level or the treatment level specified in tlevel(). If you specify the tlevel() option, you need to specify only one new variable; otherwise, you must specify a new variable for each treatment level (except the control level).

cmean calculates the conditional mean for each treatment level or the treatment level specified in tlevel(). If you specify the tlevel() option, you need to specify only one new variable; otherwise, you must specify a new variable for each treatment level.

ps calculates the propensity score of each treatment level or the treatment level specified in tlevel(). If you specify the tlevel() option, you need to specify only one new variable; otherwise, you must specify a new variable for each treatment level.

xb calculates the linear prediction at each treatment level or the treatment level specified in tlevel(). If you specify the tlevel() option, you need to specify only one new variable; otherwise, you must specify a new variable for each treatment level.

`psxb` calculates the linear prediction for the propensity score at each noncontrol level of the treatment or the treatment level specified in `tlevel()`. If you specify the `tlevel()` option, you need to specify only one new variable; otherwise, you must specify a new variable for each treatment level (except the control level).

`lnsigma` calculates the log square root of the conditional latent variance for each treatment level or the treatment level specified in `tlevel()`. This option is valid when outcome model `hetprobit()` was used. If you specify the `tlevel()` option, you need to specify only one new variable; otherwise, you must specify a new variable for each treatment level.

`pslnsigma` calculates the log square root of the latent variance for the propensity score. This option is only valid when treatment model `hetprobit()` was used. Specify only one new variable.

`tlevel(`*treat_level*`)` specifies the treatment level for prediction.

`scores` calculates the score variables. Parameter-level scores are computed for the treatment mean and average treatment-effect equations. Equation-level scores are computed for the propensity-score equations.

The jth new variable will contain the scores for the jth parameter in the coefficient table if $j \leq t$, where t is the number of treatment levels. Otherwise, it will contain the scores for fitted equation $j - t$ following the first t parameters in the coefficient table.

Options for predict after ipw

Main

`ps`, the default, calculates the propensity score of each treatment level or the treatment level specified in `tlevel()`. If you specify the `tlevel()` option, you need to specify only one new variable; otherwise, you must specify a new variable for each treatment level.

`xb` calculates the linear prediction for the propensity score at each noncontrol level of the treatment or the treatment level specified in `tlevel()`. If you specify the `tlevel()` option, you need to specify only one new variable; otherwise, you must specify a new variable for each treatment level (except the control level).

`lnsigma` calculates the log square root of the latent variance. This option is only valid when treatment model `hetprobit()` was used. Specify only one new variable.

`tlevel(`*treat_level*`)` specifies the treatment level for prediction.

`scores` calculates the score variables. Parameter-level scores are computed for the treatment mean and average treatment-effect equations. Equation-level scores are computed for the propensity-score equations.

The jth new variable will contain the scores for the jth parameter in the coefficient table if $j \leq t$, where t is the number of treatment levels. Otherwise, it will contain the scores for fitted equation $j - t$ following the first t parameters in the coefficient table.

Options for predict after nnmatch and psmatch

Main

`te`, the default, calculates the treatment effect.

`po` calculates the predicted potential outcomes for each observation and treatment level or the treatment level specified in `tlevel()`. If you specify the `tlevel()` option, you need to specify only one new variable; otherwise, you must specify new variables for the control and treated groups.

`distance` calculates the distances of the nearest neighbors for each observation. The number of variables generated is equal to the maximum number of nearest-neighbor matches. This is equal to the number of index variables generated by the estimation option `generate(`*stub*`)`. You may use the *stub*`*` syntax to set the distance variable prefix: *stub*`1`, *stub*`2`, ...

`ps` calculates the propensity score of each treatment level or the propensity score of the treatment level specified in `tlevel()`. If you specify the `tlevel()` option, you need to specify only one new variable; otherwise, you must specify new variables for the control and treated groups.

`lnsigma` calculates the log square root of the latent variance. This option is only valid when treatment model `hetprobit()` was used. Specify only one new variable.

`tlevel(`*treat_level*`)` restricts potential-outcome estimation to either the treated group or the control group. This option may only be specified with options `po` and `ps`.

Options for predict after ra

Main

`te`, the default, calculates the treatment effect for each noncontrol treatment level or the treatment level specified in `tlevel()`. If you specify the `tlevel()` option, you need to specify only one new variable; otherwise, you must specify a new variable for each treatment level (except the control level).

`cmean` calculates the conditional mean for each treatment level or the treatment level specified in `tlevel()`. If you specify the `tlevel()` option, you need to specify only one new variable; otherwise, you must specify a new variable for each treatment level.

`xb` calculates the linear prediction at each treatment level or the treatment level specified in `tlevel()`. If you specify the `tlevel()` option, you need to specify only one new variable; otherwise, you must specify a new variable for each treatment level.

`lnsigma` calculates the log square root of the conditional latent variance for each treatment level or the treatment level specified in `tlevel()`. This option is valid when outcome model `hetprobit()` was used. If you specify the `tlevel()` option, you need to specify only one new variable; otherwise, you must specify a new variable for each treatment level.

`tlevel(`*treat_level*`)` specifies the treatment level for prediction.

`scores` calculates the score variables. Parameter-level scores are computed for the treatment mean and average treatment-effect equations. Equation-level scores are computed for the regression equations.

The jth new variable will contain the scores for the jth parameter in the coefficient table if $j \leq t$, where t is the number of treatment levels. Otherwise, it will contain the scores for fitted equation $j - t$ following the first t parameters in the coefficient table.

Remarks and examples

stata.com

Checking model specification is the most frequent reason for postestimation computation after `teffects`. `teoverlap` provides a graphical method for checking the overlap assumption; see [CAUSAL] **teoverlap**. Summarizing the estimated probabilities provides another check. Recall that the reciprocals of these estimated probabilities are used as weights by some of the estimators. If the estimated probabilities are too small, the weights blow up.

We estimate the ATE of maternal smoking on infant birthweight by inverse-probability weighting; see example 1 of [CAUSAL] **teffects ipw** for background.

```
. use https://www.stata-press.com/data/r18/cattaneo2
(Excerpt from Cattaneo (2010) Journal of Econometrics 155: 138-154)
. teffects ipw (bweight) (mbsmoke mmarried c.mage##c.mage fbaby medu, probit)
Iteration 0:  EE criterion =  4.621e-21
Iteration 1:  EE criterion =  7.358e-26
Treatment-effects estimation                   Number of obs     =      4,642
Estimator      : inverse-probability weights
Outcome model  : weighted mean
Treatment model: probit
```

bweight	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
mbsmoke (Smoker vs Nonsmoker)	-230.6886	25.81524	-8.94	0.000	-281.2856	-180.0917
POmean						
mbsmoke Nonsmoker	3403.463	9.571369	355.59	0.000	3384.703	3422.222

Below we compute and summarize the estimated treatment probabilities.

```
. predict pr1
(option ps assumed; propensity score)
. summarize pr1 if mbsmoke==1, detail
            propensity score, mbsmoke=Nonsmoker
      Percentiles      Smallest
 1%     .2991634       .2196947
 5%      .544155       .2258079
10%     .5973879       .2258079       Obs                 864
25%       .63777       .2409025       Sum of wgt.         864
50%     .7601717                      Mean           .7456264
                       Largest        Std. dev.      .1276102
75%     .8453946       .9533503
90%     .8943686       .9596144       Variance       .0162844
95%     .9096801        .961022       Skewness      -.7701643
99%     .9367017       .9665684       Kurtosis       3.858214
```

The smallest values do not imply very large weights.

Below we compute and summarize the estimated probabilities of not getting the treatment.

```
. generate pr0 = 1 -pr1
. summarize pr0 if mbsmoke==0, detail
```

pr0

Percentiles		Smallest		
1%	.0351884	.0074551		
5%	.0578012	.0079309		
10%	.0674359	.0106305	Obs	3,778
25%	.0950869	.0106305	Sum of wgt.	3,778
50%	.1372589		Mean	.1698913
		Largest	Std. dev.	.1059434
75%	.2211142	.7547572		
90%	.3242757	.774192	Variance	.011224
95%	.3883457	.7803053	Skewness	1.514456
99%	.501537	.7816764	Kurtosis	6.151114

Although there are two small probabilities, overall the small values do not imply large weights.

Also see

[CAUSAL] **teoverlap** — Overlap plots

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

[CAUSAL] **teffects ipw** — Inverse-probability weighting

[CAUSAL] **teffects ipwra** — Inverse-probability-weighted regression adjustment

[CAUSAL] **teffects nnmatch** — Nearest-neighbor matching

[CAUSAL] **teffects psmatch** — Propensity-score matching

[CAUSAL] **teffects ra** — Regression adjustment

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