Causal inference commands — Introduction to causal inference commands

Description Remarks and examples Also see

Description

In this entry, we give you an overview of the estimation commands in Stata that are designed for causal inference. We provide important details about each command so that readers can select the one that best fits their data and research needs.

Here we assume that you are familiar with causal inference and the most common assumptions. For an introduction to these concepts, see [CAUSAL] **Intro**.

Remarks and examples

Below, we introduce Stata commands that are specifically designed for causal inference. For each command, we provide information on the type of data required and the necessary assumptions. In addition, we outline the type of statistics that can be estimated—typically one or more of the average treatment effect (ATE), the average treatment effect on the treated (ATET), or the potential-outcome means (POM). We also indicate the type of outcome variable (continuous, binary, count, fractional, or nonnegative) and the type of treatment (binary, multivalued, or continuous) that each command supports. Finally, we note which models must be specified: a model for the outcome, a model for the treatment, both, or none.

Remarks are presented under the following headings:

teffects

The `teffects` suite of commands is useful for estimating treatment effects from cross-sectional data. These commands rely on the stable unit treatment value assumption (SUTVA), unconfoundedness (conditional-independence) assumption, and overlap assumption.

The commands in the `teffects` suite and the type of estimator provided by each are as follows:

`teffects ra`	Regression adjustment
`teffects ipw`	Inverse-probability weighting
`teffects ipwra`	Inverse-probability-weighted regression adjustment
`teffects aipw`	Augmented inverse-probability weighting
`teffects nnmatch`	Nearest-neighbor matching
`teffects psmatch`	Propensity-score matching

Details on the available estimands, types of outcomes and treatments supported, and the models to be specified are given below:

Command	Estimand	Outcome types	Treatment types	Models specified
`teffects ra`	ATE ATET POM	continuous binary count fractional nonnegative	binary multivalued	outcome
`teffects ipw`	ATE ATET POM	continuous binary count fractional nonnegative	binary multivalued	treatment
`teffects ipwra`	ATE ATET POM	continuous binary count fractional nonnegative	binary multivalued	outcome treatment
`teffects aipw`	ATE ATET POM	continuous binary count fractional nonnegative	binary multivalued	outcome treatment
`teffects psmatch`	ATE ATET	continuous binary count fractional nonnegative	binary	treatment
`teffects nnmatch`	ATE ATET	continuous binary count fractional nonnegative	binary	outcome*

*`nnmatch` includes covariates for modeling the outcome but does not require specification of a functional form for the outcome model.

For further information on these commands and the properties of the estimators that they implement, see [CAUSAL] **teffects intro**.

stteffects

The stteffects suite of commands is useful for estimating treatment effects from survival-time data. These commands rely on the SUTVA, unconfoundedness (conditional-independence) assumption, and overlap assumption. They also rely on an assumption that the correct adjustment is made for censoring.

The commands in the stteffects suite and the type of estimator provided by each are as follows:

stteffects ra	Survival-time regression adjustment
stteffects wra	Survival-time weighted regression adjustment
stteffects ipw	Survival-time inverse-probability weighting
stteffects ipwra	Survival-time inverse-probability-weighted regression adjustment

Details on the available estimands, types and treatments supported, and the models to be specified are given below:

Command	Estimand	Treatment types	Models specified
stteffects ra	ATE	binary	outcome
	ATET	multivalued	
	POM		
stteffects wra	ATE	binary	outcome
	ATET	multivalued	censoring
	POM		
stteffects ipw	ATE	binary	treatment
	ATET	multivalued	censoring
	POM		
stteffects ipwra	ATE	binary	outcome
	ATET	multivalued	treatment
	POM		censoring (optional)

For further information on these commands and the properties of the estimators that they implement, see [CAUSAL] **stteffects intro**.

telasso

The telasso command is useful for estimating treatment effects from cross-sectional data and using lasso to select from among many potential control variables to be included in the model. This estimator relies on the SUTVA, unconfoundedness (conditional-independence) assumption, and overlap assumption.

telasso allows a continuous, binary, count, or nonnegative outcome and requires a binary treatment variable. Models are specified for both the outcome and the treatment. The ATE, ATET, or POM may be requested.

For further information on this command and the properties of the estimator that it implements, see [CAUSAL] **telasso**.

cate

The cate suite of commands is useful for estimating the average treatment effects conditional on a set of variables, known as conditional average treatment effects (CATEs). cate provides three different CATE estimates: individualized average treatment effects (IATEs), group average treatment effects (GATEs), and sorted group average treatment effects (GATESs). IATEs are treatment effects conditional on observation-level characteristics. There is one IATE for each observation in the data. GATEs are treatment effects conditional on prespecified groups. There is a treatment effect for each group. GATESs are average treatment effects for a prespecified number of groups. The groups are determined by quantiles of individual-level treatment-effects values. Estimating CATEs allows us to study the treatment-effect heterogeneity and evaluate the treatment-assignment policy.

These commands rely on the stable unit treatment value assumption (SUTVA), unconfoundedness (conditional-independence) assumption, and overlap assumption.

cate allows a continuous outcome and requires a binary treatment variable. cate estimates a CATE function, an outcome model, and a treatment-assignment model. The CATE function is estimated by the partialing-out (PO) estimator or the augmented inverse-probability weighting (AIPW) estimator via random forest or parametric regression. The outcome and the treatment models can be estimated using cross-fitting via lasso, random forest, or parametric regression.

For further information on this command and the properties of the estimator that it implements, see [CAUSAL] **cate**.

Difference in differences

The difference-in-differences suite of commands is useful for estimating treatment effects from data in which some of the units are observed both before and after a treatment and some units remain untreated. The difference-in-differences suite comprises the following commands:

didregress	Difference in differences
xtdidregress	Difference in differences for panel data
hdidregress	Heterogeneous difference in differences
xthdidregress	Heterogeneous difference in differences for panel data

The didregress and hdidregress commands estimate treatment effects for repeated cross-sectional data, while xtdidregress and xthdidregress estimate treatment effects for panel data. The didregress and xtdidregress commands estimate a single ATET. The hdidregress and xthdidregress commands allow for heterogeneous treatment effects and report separate ATETs for each time and treatment cohort.

These estimators rely on the SUTVA, unconfoundedness (conditional-independence) assumption, and overlap assumption. In addition, they rely on an assumption of parallel trends in the treatment and control groups.

For further information on these commands and the properties of the estimators that they implement, see [CAUSAL] **DID intro**.

Endogenous treatment

The et commands are useful for estimating treatment effects from cross-sectional data in cases where the unconfoundedness (conditional-independence) assumption is violated because treatment assignment is not independent of the potential outcomes. The et commands comprise the following:

eteffects	Endogenous treatment-effects estimation
etpoisson	Poisson regression with endogenous treatment effects
etregress	Linear regression with endogenous treatment effects

Details on the estimands, types of outcomes and treatments supported, and the models to be specified are given below:

Command	Estimand	Outcome types	Treatment types	Models specified
eteffects	ATE ATET POM	continuous binary count fractional nonnegative	binary	outcome treatment
etpoisson	ATE ATET POM	count nonnegative	binary	outcome treatment
etregress	ATE ATET POM	continuous	binary	outcome treatment

Note that eteffects provides the ATE, ATET, and POM directly. etregress estimates the ATE directly, while the ATET and POM can be obtained from margins after estimation. For etpoisson, ATE, ATET, and POM can all be obtained from margins after estimation.

For further information on these commands and the properties of the estimators that they implement, see [CAUSAL] **eteffects**, [CAUSAL] **etpoisson**, and [CAUSAL] **etregress**.

Causal mediation

The mediate command is useful for estimating direct, indirect, and total treatment effects from cross-sectional data in some cases where the treatment may affect an outcome both directly and indirectly. An indirect effect is one in which the treatment affects another variable, called a mediator, and the mediator in turn affects the outcome.

The mediate command allows both outcome and mediator variables to be continuous, binary, count, and nonnegative. The treatment may be binary, multivalued, or continuous. Models may be specified for the treatment and the mediator.

This estimator relies on the SUTVA, unconfoundedness (conditional-independence) assumption, and overlap assumption.

mediate provides estimates of the following statistics:

Estimand	Synonym
average indirect treatment effect (AITE)	natural indirect effect (NIE)
average direct treatment effect (ADTE)	natural direct effect (NDE)
total average treatment effect (ATE)	marginal total effect (MTE)
average indirect treatment effect with respect to controls (AITEC)	pure natural indirect effect (PNIE)
average direct treatment effect with respect to the treated (ADTET)	total natural direct effect (TNDE)

For further information on this command and the properties of the estimator that it implements, see [CAUSAL] **mediate**.

Extended regression models

The extended regression model (ERM) suite of commands is designed to account for treatment (exogenous or endogenous), endogenous covariates, and nonrandom sample selection one at a time or in combination. Commands are available for both cross-sectional and panel data. The following commands are comprised in the ERM suite:

eregress	Extended linear regression
eintreg	Extended interval regression
eprobit	Extended probit regression
eoprobit	Extended ordered probit regression
xteregress	Extended linear regression for panel data
xteintreg	Extended interval regression for panel data
xteprobit	Extended probit regression for panel data
xteoprobit	Extended ordered probit regression for panel data

eregress and xteregress fit models for continuous outcomes. eintreg and xteintreg fit models for interval-censored outcomes. eprobit and xteprobit fit models for binary outcomes. eoprobit and xteoprobit fit models for ordinal outcomes. All commands allow binary and multivalued treatments.

After fitting a model that accounts for endogenous or exogenous treatment with one of the ERM commands, you can use estat teffects to estimate the ATE, ATET, or POM.

For further information on these commands and the properties of the estimators that they implement, see [ERM] **Intro 1**.

Other commands in Stata provide some of the features found in the ERM commands. For instance, when you account only for endogenous covariates, eregress and ivregress provide equivalent parameter estimates. Instrumental-variable commands—ivregress, ivprobit, ivpoisson, and ivtobit—are designed to account for endogeneity (unobserved confounding) and provide consistent parameter estimates in this situation. Thus, these commands are used when the goal is causal inference. In some cases, a parameter estimated by these commands can be directly interpreted as the causal effect of interest, and in other cases, postestimation commands can be used to obtain the ATE, ATET, and POM.

margins

The `margins` command is available after many estimation commands in Stata. When a researcher has determined that appropriate assumptions have been satisfied for performing causal inference, many estimation commands can be used in combination with `margins` to estimate the ATE, ATET, and POM. As a simple example, you may type

```
. regress y c.x##i.trt, vce(robust)
```

to fit a linear regression of `y` on treatment `trt` and adjusted for covariate `x`. To estimate the POM, you could type

```
. margins trt, vce(unconditional)
```

The ATE is a contrast of the POM, and `margins` uses the `r.` operator to request such a contrast:

```
. margins r.trt, vce(unconditional)
```

The `margins` command can be used similarly after other estimation commands, and the results can be interpreted causally when proper assumptions for causal inference have been met.

For more information on `margins`, see [R] **margins**.

Also see

[CAUSAL] **Intro** — Introduction to causal inference and treatment-effects estimation

[CAUSAL] **Glossary**