Title

scdata —— Data Preparation for Synthetic Control Methods.

Syntax

scdata *features* [*if*] [*in*] **, id(***idvar***) time(***timevar***) outcome(***outcomevar***)**
treatment(*treatmentvar***) dfname(***string***)** [**covadj(***string***) anticipation(#)**
cointegrated constant pypinocheck]

Description

scdata prepares the data to be used by scest or scpi to implement estimation and inference procedures for Synthetic Control (SC) methods. It allows the user to specify the outcome variable, the features of the treated unit to be matched, and covariate-adjustment feature by feature. The command follows the terminology proposed in Cattaneo, Feng, and Titiunik (2021). The command is a wrapper of the companion Python package. As such, the user needs to have a running version of Python with the package installed. A tutorial on how to install Python and link it to Stata can be found here.

Companion R and Python packages are described in Cattaneo, Feng, Palomba and Titiunik (2022).

Companion commands are: scest for point estimation, scpi for inference procedures, and scplot for SC plots.

Related Stata, R, and Python packages useful for inference in SC designs are described in the following website:

https://nppackages.github.io/scpi/

For an introduction to synthetic control methods, see Abadie (2021) and references therein.

Options

Variables

id(*idvar***)** specifies the variable containing the identifier for each unit.

time(*timevar***)** specifies the variable containing the time period of each observation.

outcome(*outcomevar***)** specifies the outcome variable of interest. Note that *outcomevar* may not be among the *features* specified.

treatment(*treatmentvar***)** specifies the treatment indicator.

Estimator

covadj(*string***)** specifies the variables to be used for adjustment for each feature. If the user wants to specify the same set of covariates for all features, a string should be provided according to the following format: **covadj("***cov1*, *cov2***").** If instead a different set of covariates per feature has to be specified, then the following format should be used **covadj("***cov1*, *cov2; cov1*, *cov3***").** Note that in this latter case the number of sub-lists delimited by ";" must be equal to the number of *features*. Moreover, the order of the sub-lists matters, in the sense that the first sub-list is interpreted as the set of covariates used for adjustment for the first feature, and so on. Finally, the user can specify 'constant' and 'trend' as covariates even if they are not present in the loaded dataset, the former includes a constant, whilst the latter a linear deterministic trend.

anticipation(#) specifies the number of periods of potential anticipation effects. Default is **anticipation(0).**

cointegrated if specified indicates that there is a belief the features form a cointegrated system.

constant if specified includes a constant term across features.

Others

dfname(*string***)** specifies the name of the Python object that is saved and that will be passed to scest or scpi.

pypinocheck) if specified avoids to check that the version of scpi_pkg in Python is the one required by **scdata** in Stata. When not specified performs the check and stores a macro called to avoid checking it multiple times.{p_end

---

Example: Germany Data

Setup

```
. use scpi_germany.dta
```

Prepare data

```
. scdata gdp, dfname("python_scdata") id(country) outcome(gdp) time(year)
treatment(status) cointegrated
```

Stored results

scdata stores the following in **e():**

Scalars

e(J)	number of donors
e(KM)	total number of covariates used for adjustment

Macros

e(features)	name of features
e(outcomevar)	name of outcome variable
e(constant)	logical indicating the presence of a common constant across features
e(cointegrated_data)	logical indicating cointegration

Matrices

e(A)	pre-treatment features of the treated unit
e(B)	pre-treatment features of the control units
e(C)	covariates used for adjustment
e(P)	predictor matrix

References

Abadie, A. 2021. Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, 59(2), 391–425.

Cattaneo, M. D., Feng, Y., and Titiunik, R. 2021. Prediction intervals for synthetic control methods. *Journal of the American Statistical Association*, 116(536), 1865–1880.

Cattaneo, M. D., Feng, Y., Palomba F., and Titiunik, R. 2022. scpi: Uncertainty Quantification for Synthetic Control Estimators, *arXiv*:2202.05984.

Cattaneo, M. D., Feng, Y., Palomba F., and Titiunik, R. 2022. Uncertainty Quantification in Synthetic Controls with Staggered Treatment Adoption, *arXiv*:2210.05026.

Authors

Matias D. Cattaneo, Princeton University, Princeton, NJ. cattaneo@princeton.edu.

Yingjie Feng, Tsinghua University, Beijing, China. fengyj@sem.tsinghua.edu.cn.

Filippo Palomba, Princeton University, Princeton, NJ. fpalomba@princeton.edu.

Rocio Titiunik, Princeton University, Princeton, NJ. titiunik@princeton.edu.