A SAS macro for estimation and inference in differences-in-differences applications with only a few treated groups

Nicolas Moreau

DOCUMENTS DE TRAVAIL CEMOI / CEMOI WORKING PAPERS

A SAS macro for estimation and inference in differences-in-differences applications with only a few treated groups

Nicolas Moreau[1]

http://cemoi.univ-reunion.fr

Centre d'Economie et de Management de l'Océan Indien
Université de La Réunion

May 2017

Abstract

This paper presents a SAS macro written for estimation and inference in differences-in-differences applications with only a few treated groups following Conley and Taber's (2011) methodology.

JEL: C210, C810

[1] E-mail: nicolas.moreau@univ-reunion.fr
We would like to thank Christopher R. Taber for his helpful clarifications. The usual disclaimer applies.

Introduction

Here we introduce the SAS macro *DiD_CT* written for estimation and inference in differences-in-differences applications with only a few treated groups following Conley and Taber's (2011) methodology.

Conley and Taber (2011) show that the point estimator of the treatment effect is not consistent when the number of treatment groups is small, even for a large number of control groups. However, it is still possible to conduct inference about treatment effect using information in the error term from the control groups.

The SAS macro *DiD_CT* considers both group×time and individual-level data. It mixes SAS procedures and lines of code written with SAS/IML. The complete code is available at http://cemoi.univ-reunion.fr/econometrie-avec-r-et-sas/.

Differences-in-differences estimates with the SAS macro *DiD_CT*

For group×time level data, the model is:

$$y_{gt} = \alpha d_{gt} + X_{gt}\beta + \theta_g + \gamma_t + \eta_{gt}, \quad (1)$$

where d_{gt} is the policy variable, X_{gt} a vector of regressors, θ_g a time-invariant fixed effect for group g, γ_t a fixed time effect that is common across all groups but varies across time $t = 1, \ldots, T$, and η_{gt} a group-by-time error term.

For individual-level data, the model specification is:

$$y_{igt} = \alpha d_{gt} + X_{gt}\beta + Z_i\delta + \theta_g + \gamma_t + \eta_{gt} + \varepsilon_i,$$

where Z_i is a vector of individual-specific regressors and ε_i another error term. Note that because the policy variable d_{gt} is only observed at the group×time level, it is common first to aggregate the individual-level data to the group×time level and then to OLS regress the model:

$$\tilde{y}_{gt} = \alpha d_{gt} + X_{gt}\beta + \theta_g + \gamma_t + \eta_{gt}, \quad (2)$$

where $\tilde{y}_{gt}$ is the average of the outcome variable y in group g and time period t after partialling out individual-specific covariates. To compute $\tilde{y}_{gt}$, *DiD_CT* uses a two-step approach advocated by Hansen (2007), Conley and Taber (2011), and Cameron and Miller (2015). First, *DiD_CT* runs the OLS regression of the dependent variable y_{igt} on all group-by-time dummies and individual-specific covariates Z_i, with no constant. Second, $\tilde{y}_{gt}$ equals the estimated coefficients of the group-by-time dummies. If the model does not include any individual-specific covariates, $\tilde{y}_{gt}$ is only the average of y_{igt} within the group-time cell.

To estimate the treatment effect α and the vector of parameters β, *DiD_CT* allows the user to directly estimate equations (1) and (2) or to additionally partial out the group and time dummies by applying the Frisch-Waugh-Lovell theorem. In the latter case, *DiD_CT* regresses y_{gt}, $\tilde{y}_{gt}$, d_{gt}, and X_{gt} on the group and year dummies, with no constant. Then α and β are estimated by regressing the residuals of $\tilde{y}_{gt}$ on the residuals of d_{gt} and X_{gt}, with no constant.

Inference with the SAS macro *DiD_CT*[2]

DiD_CT tests the null hypothesis of the no treatment effect $\alpha = 0$. To make the inference valid, the approach of Conley and Taber (2011) is to use $\hat{\alpha}$ as a test statistic and construct an acceptance region from a distribution of simulated values for the test statistic under the null. The null hypothesis is rejected if $\hat{\alpha}$ falls outside the acceptance region.

Conley and Taber (2011) consider two such empirical distributions $\hat{\Gamma}(w)$ and $\hat{\Gamma}^*(w)$. Both are based on a ratio of cross products of deviations from the mean of the policy variable from treatment groups with residuals to the sum of squared deviations from the mean of the policy variable from treatment groups. Let N_0 and N_1 be the number of control and treatment groups, respectively. Let $\tilde{\eta}^c$ be the vector of constrained residuals from the constrained estimation of equation (1) under the null hypothesis of no treatment effect. Let $\tilde{\eta}_l^c$ be the vector of constrained residuals from control group l.

The basic building block of $\hat{\Gamma}(w)$ is the following ratio:

[2]We refer the reader to Conley and Taber (2011) for a clear and comprehensive presentation of the methodology at stake.

$$\frac{\sum_{t=1}^{T}(d_{gt}-\bar{d}_g)\tilde{\eta}_{lt}^{c}}{\sum_{g=1}^{N_1}\sum_{t=1}^{T}\left(d_{gt}-\bar{d}_g\right)^2}.$$

For each treatment group g, there are N_0 ratios of the kind. Let S_g be the set that contains these N_0 ratios for $g=1,...,N_1$. Now, select one ratio from each set S_g and add them to obtain one particular element of $\hat{\Gamma}(w)$. As there are ${N_0}^{N_1}$ possible combinations to attain this particular sum, the distribution $\hat{\Gamma}(w)$ contains ${N_0}^{N_1}$ elements. This number tends quickly to infinity even for moderate values of N_0 and N_1.

To avoid this problem, *DiD_CT* allows the user to specify the maximum number of combinations used to approximate the distribution $\hat{\Gamma}(w)$. Let *maxcomb* be this number. Then *DiD_CT* randomly draws K ratios without replacement from each set S_g such that $K^{N_1} = maxcomb$.

To obtain the empirical distribution $\hat{\Gamma}^*(w)$, the residuals under the null hypothesis for the treatment groups are used along with the residuals from controls to form $N_0 + N_1$ vectors of residuals of dimension T. Let $\tilde{\eta}_l$ be the vector of residuals under the null for group $l=1,...,N_1 + N_0$.Then N_1 of these vectors are randomly taken without replacement to form the particular ratio:

$$\frac{\sum_{g=1}^{N_1}\sum_{t=1}^{T}(d_{gt}-\bar{d}_g)\tilde{\eta}_{l_{gt}}}{\sum_{g=1}^{N_1}\sum_{t=1}^{T}\left(d_{gt}-\bar{d}_g\right)^2}.$$

As there are $\frac{(N_0+N_1)!}{N_0!N_1!}$ different possibilities to choose N_1 elements without replacement from a set with $N_1 + N_0$ elements, the empirical distribution $\hat{\Gamma}^*(w)$ includes $\frac{(N_0+N_1)!}{N_0!N_1!}$ ratios. *DiD_CT* allows the user to randomly draw *maxcomb* of these elements to approximate the distribution $\hat{\Gamma}^*(w)$.

DiD_CT provides two estimators for $\hat{\Gamma}^*(w)$. The first is labelled $\hat{\Gamma}_c^*(w)$; it uses the constrained residuals $\tilde{\eta}^c$ for $\tilde{\eta}$. The second is $\hat{\Gamma}_u^*(w)$; it uses the unconstrained residuals $\tilde{\eta}^u$ from the estimation of equation (1) for the controls and forms residuals under the null hypothesis for the treatment groups as $\tilde{\eta}_{gt} = \tilde{\eta}_{gt}^u + \hat{\alpha}d_{gt}$.

For each of the three distributions $\hat{\Gamma}(w)$, $\hat{\Gamma}_c^*(w)$, and $\hat{\Gamma}_u^*(w)$, *DiD_CT* supplies the lower and upper bounds for the 90%, 95%, and 99% acceptance regions. At the 5% level, for instance, and using $\hat{\Gamma}(w)$, the lower and upper bounds are defined as L. bound= $\hat{\Gamma}^{-1}(0.05)$ and U. bound= $\hat{\Gamma}^{-1}(0.95)$, respectively. Reject the null if $\hat{\alpha}$ falls outside the acceptance region.

Syntax of the SAS macro *DiD_CT*

The syntax is:

%*DiD_CT*(data=,depvar=,regtreat=,reg=,regind=,groupid=,timeid=,parout=,maxcomb=).

data specifies the input data set. *depvar* is the outcome variable. *regtreat* is the policy variable. *reg* specifies the list of regressors X_{gt}; this list can be empty. *regind* specifies the list of regressors Z_i; this list can be empty. *groupid* specifies the variable that identifies the group for each observation; it cannot be empty. *timeid* is the variable that identifies the time period for each observation in the input data; it cannot be empty. All variables in *depvar*, *regtreat, reg*,and *regind* must be numeric.

The parameter *parout* is used to additionally partial out the group and time dummies. This is the default option if *parout* is not specified by the user. If *parout* =no or *parout* =none (or whatever character/value except a blank), the direct estimation of equation (1) is provided.

The parameter *maxcomb* specifies the maximum number of combinations used to simulate the distributions $\hat{\Gamma}(w)$, $\hat{\Gamma}_c^*(w)$, and $\hat{\Gamma}_u^*(w)$. *maxcomb* is non-binding if the number of all possible combinations is less than *maxcomb*; it cannot be empty.

Note that inference with only a few treatment groups cannot be processed from the empirical distributions $\hat{\Gamma}(w)$ and $\hat{\Gamma}_c^*(w)$ if the parameters *reg* and *parout* are both left empty. It is not possible to form residuals under the null hypothesis of no treatment effect in that case. If the model does not include any regressor X_{gt}, *parout* must be completed.

Output results of the SAS macro *DiD_CT*

Several tables show the number and percentage of observations per group and time period as the distribution of the policy variable per group and time period.
The estimation results are presented in two different tables. The first exhibits the estimate $\hat{\alpha}$ of the treatment parameter followed by the lower and upper bounds of the corresponding 90%, 95%, and 99% acceptance regions for the distributions $\hat{\Gamma}(w)$, $\hat{\Gamma}_c^*(w)$, and $\hat{\Gamma}_u^*(w)$.
The second table presents the estimated parameters $\hat{\alpha}$ and $\hat{\beta}$(if any) and their standard errors from three different cluster robust variance estimators. The first is the typical cluster robust variance estimator. The second (called "Cluster Robust adj1" in the table) uses the correction $\frac{N_0+N_1}{N_0+N_1-1}\frac{N-1}{N-k}$ for the typical cluster robust variance estimator, where N is the number of observations and k the number of parameters to be estimated. The third (called "Cluster Robust adj2") uses the correction $\frac{N_0+N_1}{N_0+N_1-1}$.
T-statistics and corresponding p-values then follow.

An example
We illustrate the use of *DiD_CT* with data taken from a study of the effect of merit scholarship on college attendance as reported by Conley and Taber (2011).

coll=1 if college attendance, 0 otherwise
merit= 1 if recipient of merit scholarship, 0 otherwise
male= 1 if male, 0 otherwise
black= 1 if Black , 0 otherwise
asian= 1 if Asian, 0 otherwise
state: state of residence
year: time period.

The instructions:

```
%DiD  CT(data=lib.regmraw,depvar=coll,regtreat=merit,reg=,regind=male black
        asian,groupid=state,timeid=year, parout=no,maxcomb=9999);
```

give the following results:

AGGREGATED DATA: SUMMARY STATISTICS

Number of observations	612
Number of clusters	51
Number of time periods	12

ESTIMATION RESULTS

INFERENCE FOR TREATMENT EFFECT WITH CONLEY AND TABER'S (2011) SIMULATED ACCEPTANCE REGIONS

The parameter estimate for the treatment effect is: 0.051209

	Gamma(w)		Gamma*(w)c		Gamma*(w)u	
	L. Bound	U. Bound	L. Bound	U. Bound	L. Bound	U. Bound
90% Acceptance region	-0.022134	0.0166082	-0.029465	0.0271678	-0.023567	0.0330664
95% Acceptance region	-0.023687	0.0181432	-0.035221	0.0323397	-0.029322	0.0382383
99% Acceptance region	-0.025631	0.0201046	-0.045411	0.0424672	-0.039513	0.0483657

Number of combinations used for the empirical distribution Gamma(w):	1024
Number of combinations used for the empirical distributions Gamma*(w):	9999

ESTIMATION RESULTS

INFERENCE WITH CLUSTER ROBUST COVARIANCE ESTIMATORS

	Estimate	Std.Error	T.stat	P-value
Parameter: MERIT	0.051209	.	.	.
Standard Errors:	.	.	.	.
Cluster robust	.	0.0106101	4.8264485	0.0000135
Cluster robust adj1	.	0.0113045	4.5299474	0.0000368
Cluster robust adj2	.	0.0107157	4.7788962	0.0000159

References

Cameron, A. Colin, and Douglas L. Miller. 2015. "A Practitioner's Guide to Cluster-Robust Inference." *Journal of Human Resources* 50(2): 317-372.

Conley, Timothy G., and Christopher R. Taber. 2011. "Inference with 'Difference in Differences' with a Small Number of Policy Changes." *Review of Economics and Statistics* 93(1): 113-125.

Hansen Hansen, Christian. 2007. "Generalized Least Squares Inference in Panel and Multi-level Models with Serial Correlation and Fixed Effects." *Journal of Econometrics* 141(2): 597-620.