Multitreat: Reproducible Research with Stata & Quarto*

Your Name†

Affiliation & School

Your Email

December 7, 2023

*Your acknowledgements. This work was supported by Deutsche Forschungsgemeinschaft (Project-ID 403041268, TRR 266 Accounting for Transparency).

†Send correspondence to Your Name (orcid.org/yourID), Affiliation & School, Address, ZIP code & country.

Multitreat: Reproducible Research with Stata & Quarto

Abstract

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu. In enim justo, rhoncus ut, imperdiet a, venenatis vitae, justo. Nullam dictum felis eu pede mollis pretium. Integer tincidunt.

Keywords: Stata, Quarto, Reproducible Research, Two-Way-Fixed-Effects

Data Availability: Sources identified in the text

Public Code Repository: github.com/trr266/treat

Declaration of Interest: The author(s) declare(s) that they have no conflict of interest

1 Introduction

Hey there & congrats! You managed to skip the blindtext in the abstract. This means that we can dive into the world of reproducible research using coding languages like YAML, R, Stata, Python, Julia and many more… all from within one document: a **Quarto Markdown file (.qmd)**. Additionally, we provide you with two cherries on top of this fudge: a template .qmd-file for your next paper in our public GitHub repository & an exemplary application of one of the most common designs in causal research: a two way fixed effects research design under recent considerations of Callaway and Sant'Anna (2021). All set? Let's get started.

2 Identification Strategy & Results

Writing a paper with Quarto Markdown allows you to switch between coding languages at your will. Easiest is the implementation of R and Python, but access to Stata works too. You can either import Stata output as .tex-files, e.g. by the courtesy of Stata's user written estout package, or you immediately call Stata from within Quarto Markdown with the Statamarkdown package.[1] Anyways, we think that there is value in allocating a task to the section of your project, that is most suitable for it. Therefore, we decided to generate tables and figures within Stata and call them in the .qmd-file.

2.1 Descriptive Statistics

[Table 1 about here.]

The .do-files for Stata contain many valuable information for our toy-project & we share them with a good reason. Following open sciences practices makes your research FAIR: findable, accessible, interoperable and reusable as well as reproducible. This allows interested individuals to comprehend and verify identification strategies and analyses of your research and to immediately utilize the respective code. For instance, the .do-file *connect_wrds.do* showcases how to establish a connection to the data providers WRDS & Compustat from within Stata. You need *one* additional variable for your analyses? Simply adjust the code and build dependencies from thereon. For our

[1] www.ssc.wisc.edu/ñemken/Stataworkshops/Statamarkdown/stata-and-r-markdown.html.

exemplary analysis, we pull raw data on dynamic and static information about US publicly listed firms from Compustat and then tidy the data. We perform a simple quality check and exclude firm-year observations with missing or negative sales from the sample. For our subsequent analysis, we simulate a stable treatment effect on firms' sales (i.e. a one-standard-deviation increase). Whether or not a firm is designated as "treated" depends on a random draw. The year of treatment is also randomized, which leaves us with (i) never treated firms, (ii) always treated firms, and (iii) firms which receive treatment at some point throughout the sample (aka "staggered adoption"). Table 1 contains aggregated information on our outcome variable and firms' treatment-status. Please note that you can easily access additional information on our research design by looking into the .do-files. One general advice on this end: make sure that your code is easy to understand for everyone by using explanatory notes within the code.

2.2 Difference-in-Differences with Multiple Time Periods

[Figure 1 about here.]

[Table 2 about here.]

The simulated treatment effect can empirically be identified in the data by employing the recent advances by Callaway and Sant'Anna (2021).[2] Before you jump to the empirical analyses, however, it's good advice to visualize your data first. While there are many approaches to this, we showcase one. We plot the Average Treatment Effect on the Treated (ATT) for the group of firms which received their treatment in 2008 (aka "2008 Treatment Cohort") in Figure 1. In contrast to the presented coefficients in 2, the figure is not limited to a window of five years around the treatment. One can easily identify the (stable) treatment effect in the year of treatment and the subsequent years. One can also observe that there is no considerable fluctuation in the pre-periods of the treatment. Table 2 generalizes these visual inferences to our whole sample of treated and control firms.

[2]Please note that the Stata code for the analysis may take some time to run.

3 Conclusion

Open science rocks and so does causal research in accounting! Learn more about writing in Quarto Markdown at quarto.org.

References

Callaway, Brantly, and Pedro H. C. Sant'Anna. 2021. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics* 225: 200–230. https://doi.org/https://doi.org/10.1016/j.econom.2020.12.001.

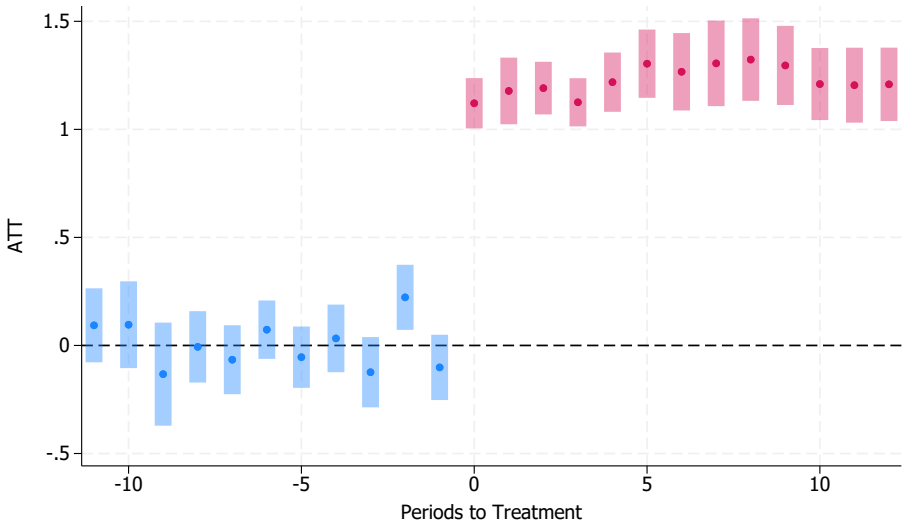
Appendix

Table A1: Variable Definitions

Name	Definition	Source
VarA	tbd	tbd
VarB	tbd	tbd
VarC	tbd	tbd

The .do-files for Stata comprise additional information to the explanations in Table A1. Please refer to the Online Supplement.

Figure 1: ATT - 2008 Treatment Cohort



This figure depicts the ATT for the 2008 treatment cohort. You can add additional information on this plot by editing the .qmd file. Did you notice that the plot is extremely sharp? That happens when you bench .png and .jpeg and use .svg/.pdf as output for figures instead.

Table 1: Descriptive Statistics

	mean	sd	min	p50	max
$\frac{sales}{lagged\,total\,assets}$	1.709746	1.234813	.0077933	1.521451	7.704735
treated	.7500101	.4330086	0	1	1
post	.421357	.4937785	0	0	1
N	124137				

Table 2: ATT (aggregated and relative)

	$\frac{sales}{lagged\,total\,assets}$	$\frac{sales}{lagged\,total\,assets}$
ATT	1.1131	
	(0.0147)	
Pretreatment $t-5$		0.0098
		(0.0148)
Pretreatment $t-4$		0.0181
		(0.0146)
Pretreatment $t-3$		-0.0036
		(0.0136)
Pretreatment $t-2$		0.0116
		(0.0120)
Pretreatment $t-1$		0.0187
		(0.0117)
Treatment		1.1353
		(0.0109)
Posttreatment $t+1$		1.1048
		(0.0139)
Posttreatment $t+2$		1.0854
		(0.0162)
Posttreatment $t+3$		1.1008
		(0.0179)
Posttreatment $t+4$		1.1172
		(0.0200)
Posttreatment $t+5$		1.1052
		(0.0215)

Standard errors in parentheses

You can add more notes.

Online Supplement

The public GitHub repositories at github.com/trr266/treat and at github.com/arndtupb/multitreat contain additional information. "multitreat" contains .do-files for Stata (*.do) and the Quarto markdown file (*paper_stata.qmd*).