The Stata Journal (*yyyy*) *vv*, Number *ii*, pp. 1–8

Greed is Good: Estimating Forward Difference-in-Differences in Stata

Jared Greathouse
Georgia State University
Atlanta, GA
jgreathouse3@student.gsu.edu

Jason Coupet
Georgia State University
Atlanta, GA
jcoupet@gsu.edu

Eric Sevigny
Georgia State University
Atlanta, GA
esevigny@gsu.edu

Abstract. Difference-in-differences designs build counterfactuals by invoking a parallel trend assumption, but this may be violated in the presence of invalid control units. Thus, selecting a control group is vital to ensure proper identification. We introduce `fdid` based on (Li, Kathleen T. "A Simple Forward Difference-in-Differences Method." *Marketing Science* 43, no. 2 (2024): 267-279). We discuss estimation and inference, document `fdid`'s syntax, and apply it empirically.

Keywords: difference-in-differences, synthetic control methods, causal inference

1 Introduction

For identification, Difference-in-Differences designs (DiD) make some form of a parallel trends assumption (PTA), assuming a constant difference between the average of the control group and treated outcome trajectories absent treatment. Unfortunately, DiD's PTA is invalid in many realistic scenarios, such as retail Costa et al. (2023), where the control group average may differ substantially from the treatment group due to inappropriate controls. Control group selection has become of interest recently to researchers. Shi and Huang (2023) extend Hsiao et al. (2012) by developing a forward selected panel data approach. Synthetic control methods (SCMs, Abadie [2021]) typically rely on a (usually) convex average of some controls to impute the counterfactual.

To better justify DiD's PTA, Li (2024) develops the forward DiD method (FDID), advocating for forward-selection to select the control group. We introduce the `fdid` method for Stata. `fdid` fits in with Stata's pantheon of program evaluation tools. Like `rcm` and `synth2` by Yan and Chen (2022, 2023), `scul` by Greathouse (2022), and `sdid` by Clarke et al. (n.d.), `fdid` uses a subset of controls to estimate the causal impact. Also, `fdid` returns the list of selected controls, graphics, and fit statistics. However, `fdid` is more user-friendly. `rcm`, `synth2`, and `allsynth` by Wiltshire (2021) all require users to specify the panel id for the treatment unit and treatment date, whereas `fdid` simply requires a dummy variable. `fdid` has more flexible data requirements, only requiring outcome data. This is in contrast to SCMs, for example, which frequently depends on covariates for acceptable pre-treatment fit (Yan and Chen 2022; Amjad et al. 2018) . `fdid` is also fast, relying on bivariate OLS for estimation. In contrast, methods such as `fect` by Liu et al. (2024) or `scul` by Greathouse (2022) employ cross validation or LASSO penalization.

 st0001

2 Forward Difference-in-Differences

Algorithm 1: Forward Difference-in-Differences

$\widehat{U}_0 \leftarrow \emptyset$;
for $k = 0$ ***to*** $N_0 - 1$ **do**
  for $i \in \mathcal{N}_0 \setminus \widehat{U}_k$ **do**
    Estimate $y_{1t} = \hat{\alpha}_{\mathcal{N}_0} + \bar{y}_{\widehat{U}_k \cup \{i\}} \quad t \in \mathcal{T}_1$ and calculate $R^2_k(\widehat{U}_k \cup \{i\})$;
  end
  Update $\widehat{U}_{k+1} \leftarrow \widehat{U}_k \cup \left\{ \text{argmax}_{i \in \mathcal{N}_0 \setminus \widehat{U}_k} R^2_k(\widehat{U}_k \cup \{i\}) \right\}$;
end
Set $\widehat{U}^* \leftarrow \text{argmax}_{k \in \{1,\dots,N_0\}} R^2_k(\widehat{U}_k)$;
Compute $\hat{y}^0_{1t} = \hat{\alpha}_{\widehat{U}^*} + \bar{y}_{\widehat{U}^*}$;
return $\widehat{U}^*$ *and* $\widehat{ATT}_{\widehat{U}^*} = \frac{1}{T_2} \sum_{t \in \mathcal{T}_2} \left(y_{1t} - \hat{y}^0_{1t}(\widehat{U}^*) \right)$

The Model We follow Li (2024)'s exposition, observing $\mathcal{N} = \{1, 2, \dots, N\}$ units where $\mathcal{N}$ has cardinality $N = |\mathcal{N}|$. $j = 1$ is treated and controls are $\mathcal{N}_0 = \mathcal{N} \setminus \{1\}$. Time is indexed by t. Denote pre-post-policy periods as $\mathcal{T}_1 = \{1, 2, \dots, T_0\}$ and $\mathcal{T}_2 = \{T_0 + 1, \dots, T\}$, where $\mathcal{T} = \mathcal{T}_1 \cup \mathcal{T}_2$. We use Algorithm 1 to select $\widehat{U} \subset \mathcal{N}_0$, or the subset of controls. Example 1 offers a stylized explanation of Algorithm 1, but we also summarize it below by quoting almost verbatim from `fdid`'s help file.

DiD is estimated like $y_{1t} = \hat{\alpha}_{\mathcal{N}_0} + \bar{y}_{\mathcal{N}_0 t}$ $t \in \mathcal{T}_1$, where $\bar{y}_{\mathcal{N}_0 t} \coloneqq \frac{1}{N_0} \sum_{j \in \mathcal{N}_0} y_{jt}$. The estimated least-squares intercept is computed like $\hat{\alpha}_{\mathcal{N}_0} \coloneqq T_1^{-1} \sum_{t \in \mathcal{T}_1} (y_{1t} - \bar{y}_{\mathcal{N}_0 t})$. We first estimate N_0 one-unit DiD submodels using each control, calculating the R-squared statistic for each submodel. The submodel with the highest R-squared is the first selected control, $\widehat{U}_1$. This is also the first "candidate" DiD model. We remove this selected control from $\mathcal{N}_0$. Next, we estimate $N_0 - 1$ two control unit DiD submodels, where we use the first selected control along with each one of the remaining $N_0 - 1$ controls. Whichever of these $N_0 - 1$ submodels has the highest R-squared statistic is the second candidate DiD model, with this second selected control being added to U_1 to form $\widehat{U}_2$. We now remove this selected control from $\mathcal{N}_0$. We continue until there are N_0 candidate DiD models/R^2 statistics. FDID uses the candidate model with the highest R-squared statistic, $\widehat{U}^*$ (we omit the asterisk for simplicity). Post-selection, Li (2024) estimates FDID like

$$y_{1t} = \hat{\alpha}_{\widehat{U}} + \bar{y}_{\widehat{U} t} \quad t \in \mathcal{T}_1 \tag{1}$$

where we now exchange $\mathcal{N}_0$ for $\widehat{U}$. Denote the FDID predictions as $\hat{y}^0_{1t} = \hat{\alpha}_{\widehat{U}} + \bar{y}_{\widehat{U} t}$, where the pre-treatment periods corresponds to the in-sample fit and the opposite denotes the out-of-sample counterfactual. Our causal estimand is: $\widehat{ATT}_{\widehat{U}} = \frac{1}{T_2} \sum_{t \in \mathcal{T}_2} (y_{1t} - \hat{y}^0_{1t})$, or the average treatment effect on the treated. From Assumption 2.1 of Li (2024) and

Arkhangelsky et al. (2021, 4094), FDID assumes *parallel* trends, $\hat{y}^0_{1t} - \bar{y}_{\widehat{U}t} = \hat{\alpha}_{\widehat{U}} + \epsilon$.[1]

Example 1. *Let $\mathcal{N}_0 = \{i_1$ (Chicago)$, i_2$ (Miami)$, i_3$ (Phoenix)$\}$ be the controls for a generic treated unit. For ($k = 1$), we estimate DiD for each control unit in $\mathcal{N}_0$ individually, yielding pre-treatment R^2 values: $R^2_{1,1} = 0.60$, $R^2_{2,1} = 0.50$, and $R^2_{3,1} = 0.23$. Since $R^2_{1,1} = 0.60$ is the highest, we update the control set to $\widehat{U}_1 = \{i_1\}$ and $R^2_k = 0.60$. For ($k = 2$), we estimate two DiD models using i_1 with the remaining controls from $\{i_2, i_3\}$, yielding $R^2_{2,2} = 0.88$ and $R^2_{3,2} = 0.68$. We select i_2 (Miami) and update the control set to $\widehat{U}_2 = \{i_1, i_2\}$ since $R^2_{2,2} = 0.88$ is the highest. For ($k = 3$), using all controls, we get $R^2_{3,3} = 0.55$. The final control set is $\widehat{U}_2 = \{i_1, i_2\}$, as $\max_k R^2_k = 0.88$.*

Inference Per Li (2024), our default standard error for the ATT is:

$$\hat{\Omega} = \left[\left(\frac{T_2}{T_1} \right) \cdot T_1^{-1} \sum_{t \in \mathcal{T}_1} \hat{v}^2_{1t} \right]^{0.5}, \quad \hat{v} = y_{1t} - \bar{y}_{\widehat{U}} - \hat{\alpha}_{\widehat{U}} \tag{2}$$

Li (2024) establishes the normal inference theory of the FDID method (see appendices B and D for theoretical derivations). In particular, Li (2024) shows that the selection algorithm chooses the correct control set as the number of pre-intervention periods tends to infinity. Li (2024) also allows the number of control units to be very large, allowing the number of controls to increase with T_1. The finite sample properties are also demonstrated in Appendix E of Li (2024).

3 The fdid command

Users need strongly balanced panel data (see [XT] **xtset**). `sdid_event` must be installed. Users also need Stata 16 or later.

3.1 Syntax

`fdid` *depvar* [*if*] [*in*] `treated(`*varname*`)` [`unitnames(`*string*`)` `gr1opts(`*string*`)` `gr2opts(`*string*`)` `placebo`]

where *depvar* is our dependent variable and *treated* is our dummy for treatment.

3.2 Options

`gr1opts`: Edits the display options of the observed versus predicted plot.

`gr2opts`: See the above, except for the plotted pointwise-treatment effect.

1. SCMs generally attempt to *match* the counterfactual to the pre-treatment trajectory.

`unitnames`: The string variable that serves as the value labels (required if the panel id is not already labeled). Note each string value pair must be uniquely identified.

`placebo`: Uses the placebo standard error of the ATT from Arkhangelsky et al. (2021) (500 replications).

3.3 Estimation Results

Matrices

`e(results)`	DID/FDID results	`e(b)`	Coefficients
`e(V)`	variance-covariance matrix	`e(dyneff)`	dynamic effects
`e(series)`	means/counterfactuals	`e(setting)`	pre-treatment periods, treatment date, post-treatment periods, number of time periods

Macros

`e(U)`	selected controls	`e(depvar)`	dependent variable
`e(properties)`	list of properties		

4 Empirical Application

We replicate Abadie et al. (2010) for two reasons: firstly, the basic results of DiD are not in dispute, being quite popular in the econometrics literature for introducing the SCM or shortcomings of DiD. More importantly, Abadie et al. (2010) explicitly say DiD's PTA is invalid. Since the point of FDID is to choose controls such that standard PTA is more credible, Abadie et al. (2010) presents a good avenue to demonstrate how `fdid` is useful for Stata users. We begin with loading in the dataset, obtained from the syntax from section 6.

```
use state year treated cigsale id using smoking, clear
```

The following output from `xtdescribe` displays the panel setup for `smoking.dta`.

```
      id:  1, 2, ..., 39                                 n =         39
    year:  1970, 1971, ..., 2000                         T =         31
           Delta(year) = 1 year
           Span(year)  = 31 periods
           (id*year uniquely identifies each observation)

Distribution of T_i:   min      5%     25%       50%       75%     95%     max
                        31      31      31        31        31      31      31

     Freq.  Percent    Cum. |  Pattern
 ---------------------------+--------------------------------
       39    100.00  100.00 |  1111111111111111111111111111111
 ---------------------------+--------------------------------
       39    100.00         |  XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
```

California is treated in 1989, compared to $N_0 = 38$ states that remain untreated. Time extends from 1970 to 2000, so $T_1 = 19$ and $T_2 = 12$. Our outcome is the rate of tobacco consumption per capita. We estimate `fdid` like

```
fdid cigsale, tr(treated) unitnames(state)

Forward Difference-in-Differences         T0 R2:     0.988     T0 RMSE:     1.282

-----------------------------------------------------------------------------------
     cigsale |      ATT       Std. Err.      t       P>|t|    [95% Conf. Interval]
-------------+---------------------------------------------------------------------
     treated | -13.64671     0.46016      29.66     0.000    -14.54861  -12.74481
-----------------------------------------------------------------------------------
Treated Unit: California
FDID selects Montana, Colorado, Nevada, Connecticut, as the optimal donors.
See Li (2024) for technical details.
```

We plot the in and out of sample predictions from both DiD and FDID as well as their control group means.[2] The results appear in Figure 1. DiD's in-sample prediction

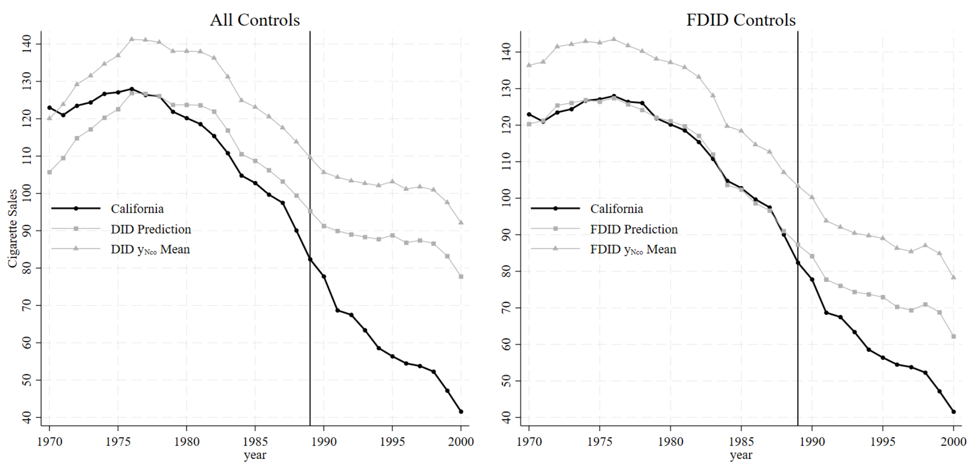

Figure 1: Observed, Predicted, and Average Curves

misses the observed in-sample values of California for the first 5 years of the time series and overestimates them from the mid-1970s until 1989. Abadie et al. (2010) also remark that in 1988, the rest of the United States has a 27% higher consumption rate than California. Given DiD's pre-intervention R^2 is equal to 60%, this comports with Abadie et al. (2010)'s conclusion that that PTA is untenable for all controls. The DiD ATT is $\widehat{ATT}_{\mathcal{N}_0} = -27.349$, a value which is likely overestimated. When we view the results of `xtdidregress` by doing `mat l e(results)` after running `fdid`, we get a 95% CI for DiD of $[-33.02, -21.68]$. Note that `xtdidregress` uses the robust standard error.

Algorithm 1 chooses 4 control units: Montana, Colorado, Nevada, and Connecticut (*all* of which were given weight by the original SCM). The pre-intervention average of these units is obviously parallel to the pre-trends of California. This fact is supported by $R^2_{\widehat{U}}$, which says 98.8% of the pre-intervention variance is explained by the $\alpha_{\widehat{U}}$ shifted

2. We omit the code in order to save space, but see FDID_SJ_Rep.do at the first author's GitHub.

average of the selected controls. For FDID, $\widehat{ATT}_{\widehat{U}} = -13.647$, a reduction of DiD's ATT by half. FDID's 95% CI is $[-14.55, -12.74]$. Another point to note is how FDID's in-sample PTA seems to hold without any covariates or predictors, suggesting that FDID's data requirements are, in some cases, more relaxed compared to SCM whose methods typically rely on predictors for convergence (Amjad et al. 2018; Vives-i-Bastida 2022), or DiD where analysts sometimes make a conditional PTA.

5 Conclusion

We wish to make clear the central limitation of `fdid`: its PTA must still be valid. As per usual, researchers should check if the standard DiD PTA is plausible first. Researchers who have found DiD's PTA to be invalid should then check if PTA holds for FDID in the pre-intervention period. Li (2024) notes that if researchers have a treated unit whose trend is much steeper than control units, for example, then use of `fdid` is invalid. Researchers should consider methods such as factor models or synthetic controls in this case (Li and Shankar 2024). However, even if FDID's PTA is plausible, other methods such as `synth2` may also serve as a robustness check.

While `fdid` is useful, we now highlight FDID's limitations and opportunities for development. For staggered adoption, Li (2024) is silent on whether using the not yet treated controls vs. never treated controls would be preferable, or on how to weight ATTs across multiple units (Wing et al. 2024). `fdid` uses the never treated controls by default and reports Cohort ATTs. We believe more formal investigation of FDID and how it could be extended to a dynamic staggered adoption is warranted. Also, some newer methods invoke *conditional* PTAs where covariates are included (Callaway and Sant'Anna 2021), or allow for heterogeneous treatment effects. FDID does not do either at present. FDID also does not account for settings where units may be treated and then untreated, or where units receive non-binary treatments as discussed in de Chaisemartin and D'Haultfœuille (2024) and D'Haultfœuille et al. (2023). Li (2024) notes other control group selection methods may be used such as the recently user-written `classifylasso` by Huang et al. (2024) (naturally, a comparison is outside the scope of our paper). All of these are potential avenues for extension, practically and theoretically.

We introduced the `fdid` command whose algorithm selects a control group for DiD. We overviewed `fdid`'s syntax and applied it empirically where the classical PTA would not deliver satisfactory results. Given `fdid`'s practical benefits, we believe `fdid` is of use to Stata users who are interested in treatment effect estimation.

6 Program Installation

```
net from "https://raw.githubusercontent.com/jgreathouse9/FDIDTutorial/main"
net install fdid
net get fdid, replace
```

7 References

Abadie, A. 2021. Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects. *Journal of Economic Literature* 59(2): 391–425. 10.1257/jel.20191450.

Abadie, A., A. Diamond, and J. Hainmueller. 2010. Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association* 105(490): 493–505. https://doi.org/10.1198/jasa.2009.ap08746.

Amjad, M., D. Shah, and D. Shen. 2018. Robust synthetic control. *The Journal of Machine Learning Research* 19(1): 802–852.

Arkhangelsky, D., S. Athey, D. A. Hirshberg, G. W. Imbens, and S. Wager. 2021. Synthetic Difference-in-Differences. *American Economic Review* 111(12): 4088–4118. https://doi.org/10.1257/aer.20190159.

Callaway, B., and P. H. Sant'Anna. 2021. Difference-in-Differences with multiple time periods. *Journal of Econometrics* 225(2): 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001.

de Chaisemartin, C., and X. D'Haultfœuille. 2024. Difference-in-Differences Estimators of Intertemporal Treatment Effects. *The Review of Economics and Statistics* 1–45. https://doi.org/10.1162/rest_a_01414.

Clarke, D., D. Pailañir, S. Athey, and G. Imbens. n.d. On Synthetic Difference-in-Differences and Related Estimation Methods in Stata. [working paper]. https://doi.org/10.48550/arXiv.2301.11859.

Costa, L., V. F. Farias, P. Foncea, J. D. Gan, A. Garg, I. R. Montenegro, K. Pathak, T. Peng, and D. Popovic. 2023. Generalized Synthetic Control for TestOps at ABI: Models, Algorithms, and Infrastructure. *INFORMS Journal on Applied Analytics* 53(5): 336–349. https://doi.org/10.1287/inte.2023.0028.

D'Haultfœuille, X., S. Hoderlein, and Y. Sasaki. 2023. Nonparametric difference-in-differences in repeated cross-sections with continuous treatments. *Journal of Econometrics* 234(2): 664–690. https://doi.org/10.1016/j.jeconom.2022.07.003.

Greathouse, J. 2022. Scul: Regularized Synthetic Controls in Stata. [working paper]. https://doi.org/10.2139/ssrn.4196189.

Hsiao, C., H. Steve Ching, and S. Ki Wan. 2012. A PANEL DATA APPROACH FOR PROGRAM EVALUATION: MEASURING THE BENEFITS OF POLITICAL AND ECONOMIC INTEGRATION OF HONG KONG WITH MAINLAND CHINA. *Journal of Applied Econometrics* 27(5): 705–740.

Huang, W., Y. Wang, and L. Zhou. 2024. Identify latent group structures in panel data: The classifylasso command. *The Stata Journal* 24(1): 46–71. https://doi.org/10.1177/1536867X241233642.

Li, K. T. 2024. Frontiers: A Simple Forward Difference-in-Differences Method. *Marketing Science* 43(2): 239–468. https://doi.org/10.1287/mksc.2022.0212.

Li, K. T., and V. Shankar. 2024. A Two-Step Synthetic Control Approach for Estimating Causal Effects of Marketing Events. *Management Science* 70(6): 3734–3747.

Liu, L., Y. Wang, and Y. Xu. 2024. A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data. *American Journal of Political Science* 68(1): 160–176. https://doi.org/10.1111/ajps.12723.

Shi, Z., and J. Huang. 2023. Forward-selected panel data approach for program evaluation. *Journal of Econometrics* 234(2): 512–535.

Vives-i-Bastida, J. 2022. Predictor Selection for Synthetic Controls. [working paper]. https://arxiv.org/abs/2203.11576.

Wiltshire, J. C. 2021. allsynth: Synthetic Control Bias-Correction Utilities for Stata. [working paper].

Wing, C., S. M. Freedman, and A. Hollingsworth. 2024. Stacked Difference-in-Differences. Working Paper 32054, National Bureau of Economic Research. http://www.nber.org/papers/w32054.

Yan, G., and Q. Chen. 2022. rcm: A command for the regression control method. *The Stata Journal* 22(4): 842–883. https://doi.org/10.1177/1536867X221140960.

———. 2023. synth2: Synthetic control method with placebo tests, robustness test, and visualization. *The Stata Journal* 23(3): 597–624. https://doi.org/10.1177/1536867X231195278.

About the authors

Jared Greathouse is a PHD Candidate of Public Policy at Georgia State University. He studies causal inference, econometrics, and machine-learning.

Jason Coupet is an Associate Professor of Public Management and Policy in the Andrew Young School of Policy Studies at Georgia State University. His research interests include strategic management, efficiency, applied microeconometrics, organizational economics, public sector management science, and the political economy of organizations.

Eric Sevigny...