Simple Diagnostics for Two-Way Fixed Effects

Pamela Jakiela*

March 8, 2026

Abstract

Difference-in-differences estimation is a widely used method of program evaluation. When treatment is implemented in different places at different times, researchers often use two-way fixed effects (TWFE) to control for location-specific and period-specific shocks. TWFE can yield severely biased estimates of average treatment effects when impacts are heterogeneous, particularly when they evolve over time within treated units. I review the sources of this bias, propose simple diagnostics for assessing its likely severity, and compare TWFE to simple and popular alternative estimators of treatment effects. I illustrate these tools through a case study of free primary education in Sub-Saharan Africa.

JEL codes: C21, O15

Keywords: difference-in-differences, two-way fixed effects, negative weights, event studies, program evaluation, heterogeneous treatment effects, World Development Indicators, free primary education

*Jakiela: Williams College, BREAD, CGD, IZA, and J-PAL, email: pj5@williams.edu. I am grateful to Lillian Bates, Kirill Borusyak, Jessica Goldberg, Oliver Hall, and Owen Ozier for helpful comments. All errors are my own.

1 Introduction

Difference-in-differences (DiD)estimation is a widely used method of program evaluation. Because it does not require explicit knowledge of the rule governing treatment assignment, it can be used for retrospective evaluation in a wide range of settings where pre-treatment data is available. When a treatment of interest is implemented in different places at different times, researchers often use two-way fixed effects (TWFE) to control for location-specific and period-specific shocks in a generalized difference-in-differences framework, estimating a treatment effect using data from multiple locations and time periods (Angrist and Pischke 2009). Recent research demonstrates that such estimates can be severely biased – and may even be incorrectly signed – when treatment effects are heterogeneous (**?**, De Chaisemartin and d'Haultfoeuille 2020).

I review the sources of this bias and propose several simple diagnostics for assessing its likely severity. When a common trends assumption is satisfied, the TWFE estimator is a linear combination of the treatment effects across treated units; however, when most or all units are treated in later periods, some treated observations may receive negative weight in the calculation of the estimated treatment effect. Negative weights are a natural consequence of the TWFE specification, and are not in and of themselves a cause for concern. However, they highlight the extent to whichDiD estimation using TWFE is not "model free" when treatment timing is staggered.[1] Indeed, the model reflects a specific set of assumptions about the structural relationship between outcomes and treatment.

Negative weighting is appropriate when treatment effects are homogeneous across time and across units, in which case the TWFE model is correctly specified. However, TWFE does not necessarily yield an unbiased estimate of any weighted average treatment effect when treated units receive negative weight and treatment effects are heterogeneous across units or over time (De Chaisemartin and d'Haultfoeuille 2020, **?**, Callaway and

[1] Card and Krueger (1995) argue that the benefit of natural experiments is that they allow for "model free" evaluation of policy impacts (Card and Krueger 1995, p. 24). However, as recent analysis of DiD illustrates, the assumption that analysis of natural experiments is model free is not necessarily justified once one moves away from simple comparisons of means between treated and untreated groups. See Gibbons, Suárez Serrato and Urbancic (2019) for a general discussion of mis-specification in fixed effects estimation.

Sant'Anna 2020, Sun and Abraham 2020, Baker, Larcker and Wang 2021). For example, if the treatment effect were zero for all units except one that received negative weight in TWFE estimation, the expected sign of the TWFE coefficient would be opposite that of any weighted average treatment effect. Even in the absence of negative weighting, the TWFE of the treatment effect need not fall between the minimum and maximum treatment effects on treated units.

I use a simple example – the elimination of primary school fees in 15 African countries – to illustrate a set of diagnostics and robustness checks that can be used to address concerns about bias and identify settings where it is unlikely to be a cause for concern. Simple tests can assess the extent to which the TWFE estimator places negative weight on treated observations, and whether the treatment effect homogeneity assumption required for such an estimator to be unbiased is likely to be appropriate. In many settings, the assumptions necessary for TWFE to provide an unbiased estimate of a weighted average treatment effect may be satisfied, and robustness checks help to address concerns about the potential for bias.

When the treatment effect heterogeneity assumption is likely to be violated, a growing list of alternative estimators are available. I compare traditional TWFE to one simple and widely adopted imputation-based estimator (Gardner, Thakral, Tô and Yap 2024, Borusyak, Jaravel and Spiess 2024), but this survey is not meant to provide a comprehensive list of alternative available. In contrast to other recent surveys (cf. Roth, Sant'Anna, Bilinski and Poe 2023), the aim of this paper is to be less comprehensive and more accessible, and in particular to review the econometrics underlying TWFE using language and notation that will be accessible to applied economists.

2 The TWFE Estimator

Consider a researcher interested in estimating the impact of treatment D_{it} on outcome Y_{it}, where i denotes a geographic unit of observation (e.g. country) and t indicates a time period (e.g. year). I will refer to units i as countries and time periods t as years throughout. Treatment D_{it} varies at the country-year level; once treatment starts in country i, it remains "on" in all subsequent periods for that country: if $D_{it} = 1$, then $D_{i\tau} = 1$ for all $\tau > t$.

The researcher wishes to estimate the treatment effect of D_{it} on Y_{it} via TWFE using the regression specification:

$$Y_{it} = \lambda_i + \gamma_t + \beta D_{it} + \epsilon_{it} \tag{1}$$

where λ_i is a vector of country fixed effects and γ_t is a vector of year fixed effects. By applying the Frisch-Waugh-Lovell theorem, we can write the OLS estimate of the treatment effect, β^{twfe}, as

$$\beta^{twfe} = \sum_{it} Y_{it} \left(\frac{\tilde{D}_{it}}{\sum_{it} \tilde{D}_{it}^2} \right) \tag{2}$$

where $\tilde{D}_{it}$ is the residual from a regression of the treatment indicator, D_{it}, on the country and year fixed effects. In a balanced panel,

$$\tilde{D}_{it} = D_{it} - \bar{D}_t - \bar{D}_i + \bar{D}_{it} \tag{3}$$

where $\bar{D}_t$ is the average level of treatment across all observations in year t, $\bar{D}_i$ is the average level of treatment across all observations for country i, and $\bar{D}_{it}$ is the average level of treatment across the entire sample of country-years (**?**). Thus, β^{twfe} is a weighted sum of the values of the outcome variable across all observations in the data set.[2]

Let Y_{0it} and Y_{1it} denote the potential outcomes with and without treatment for country i in year t. Suppose that the expected potential outcome in the absence of treatment can

[2]Equations 2 and 3 represent the mechanical relationship between the OLS estimate of β^{twfe} and the values of Y and D observed in the data (and not the expected value of β^{twfe}). The Frisch-Waugh-Lovell Theorem states that any regression of outcome Y on treatment dummy D plus an additional matrix of controls X is equivalent to a bivariate regression of $\tilde{Y}$ on $\tilde{D}$, where $\tilde{Y}$ on $\tilde{D}$ are the residuals from regressions of Y and D, respectively, on X. Thus, the discussion of TWFE generalizes directly to triple differences, models involving more than two types of fixed effects, and as well as TWFE specifications including additional controls.

be written as a combination of a unit-specific term μ_i and a period-specific shock η_t:

$$E[Y_{0it}] = \mu_i + \eta_t. \tag{4}$$

When this **common trends** assumption holds, country-level pre-treatment means and year-level shocks are effectively differenced out by the fixed effects and β^{twfe} is, in expectation, a linear combination of the treatment effects across country-year observations where $D_{it} = 1$. Importantly, some treated units may receive negative weight, and not all country-years are weighted equally. Intuitively, this occurs because the inclusion of TWFE transforms the binary treatment indicator D_{it} into a continuous measure of treatment intensity not explained by the fixed effects, $\tilde{D}_{it}$. As in any univariate OLS regression of an outcome on a continuous measure of treatment intensity, observations with below mean treatment intensity receive negative weight, and may be thought of as part of the comparison group. However, in the case of TWFE, it is outcomes with below mean levels of *residualized* treatment intensity – after controlling for country and year fixed effects – that receive negative weight.

When negative weights occur among observations in the treatment group (i.e. country-years with $D_{it} = 1$), they will tend to occur in early-adopter countries (where the country-level treatment mean is high) and in later years (when the year-level treatment mean is also high).[3] Treated country-years that receive negative weight are those where the level of treatment predicted by the country and year fixed effects exceeds one.[4] Hence, a sufficiently

[3]As Goodman-Bacon (2021) demonstrates, when treatment timing varies across units that eventually receive treatment, β^{twfe} can also be decomposed into a weighted average of all possible pairwise 2×2 difference-in-differences estimators that can be constructed from the data. β^{twfe} is a weighted average (with weights summing to one) of three types of 2×2 difference-in-differences estimators: (1) comparisons of early adopters with later adopters over periods when the later adopters are not yet treated, (2) comparisons of early adopters with later adopters over the periods when the early adopters are already treated *using the early adopters as the comparison group*, and (3) comparisons of ever-treated groups with the never-treated group, if there is one. The weight placed on an individual country-year observation in calculating β^{twfe} is the sum of the weights it receives across all three types of comparisons, including those where it is used as a comparison group. While untreated country-years are never used as the treatment group in such pairwise comparisons, treated country-years are used a comparison group some of the time. Country-years that receive negative weight in the calculation of β^{twfe} are those that are used primarily as the comparison group in the construction of pairwise 2×2 difference-in-differences estimators.

[4]Hence, there is a parallel with linear probability models, where predicted probabilities outside the unit

large never-treated group combined with enough pre-treatment data will guarantee that negative weights do not occur in the treatment group. However, in data sets with a limited number of pre-treatment periods, or with periods in which all or most units are treated, TWFE estimation will often put negative weight on the treatment effects in later periods for early-adopter units.

Negative weights on treated country-year observations are not, by themselves, a cause for concern. When treatment effects are homogeneous, the TWFE model is correctly specified because the dose response relationship between the residualized outcome variable $\tilde{Y}_{it}$ and the residualized treatment variable $\tilde{D}_{it}$ is linear. OLS correctly adjusts for the fact that the estimated fixed effects associated with high-treatment units and high-treatment periods are capturing some of the true treatment effect. Hence, negative weights are not a pathology, but a desired and natural consequence of the (implicit) modeling assumption we make when we difference out country and year means and estimate a single (implicitly homogeneous) treatment effect.

Though negative weights are not a cause for concern when treatment effects are homogeneous, the TWFE estimator can be severely biased when treatment effects are heterogeneous — particularly when they change over time within treated units (De Chaisemartin and d'Haultfoeuille 2020, Goodman-Bacon 2021). In these cases, β^{twfe} will not necessarily fall between the minimum and maximum treatment effects on any individual country-years.[5] For example, Baker et al. (2021) show that estimates of the impact of banking deregulation on inequality in the United States are biased because the impacts of deregulation appear to grow larger over time. Hence, it is important to test whether difference-in-differences estimates derived from TWFE estimation are influenced by the inclusion of later country-years receiving negative weight in the calculation of the average treatment effect, and – if so – whether the assumption of treatment effect homogeneity is plausible.

interval may be viewed as an indication of mis-specification.

[5] This is true even in cases where no treated observations receive negative weight, as the TWFE coefficient is effectively re-scaled to fit the linear relationship between outcomes and residualized treatment. However, when treatment effects are heterogeneous but are all of the same sign, the expected value of the TWFE estimator can only be signed incorrectly if treated observations receive negative weight in calculating the estimated effect.

3 Simple Diagnostics for TWFE

When assessing whether difference-in-differences estimates derived from TWFE estimation are likely to be biased, it is important to answer two questions. First, do any treated units receive negative weight in the calculation of β^{twfe}? Answering this question is straightforward since, by the Frisch-Waugh-Lovell Theorem, the weights are proportional to the residuals from a regression of the treatment dummy, D_{it}, on country and year fixed effects. Second, can one reject the hypothesis that treatment effects are homogeneous? To answer this second question, one can exploit the fact that, if the assumption of common trends holds, pre-treatment data can be used to to construct unbiased estimates of $E[Y_{0,it}]$, the expected value of the potential outcome without treatment in unit i at time t, for treated observations (Gardner et al. 2024, Borusyak et al. 2024). For treated observations, estimates of the difference between the actual values of Y_{it} and $E[Y_{0,it}]$ provide proxy measures of the observation-level treatment effect.

In what follows, I illustrate how these diagnostics can be used in practice. I present an empirical example, assessing the impact of the elimination of primary school fees in Sub-Saharan Africa on schooling outcomes. I use data on gross enrollment in and completion of primary school from 18 African countries that eliminated primary school fees between 1994 and 2019. These outcomes are ideal for illustrating the properties of the TWFE estimator because the elimination of primary school fees is likely to have had a large and immediate impact on primary school enrollment, while the impact on primary school completion likely started small (since only students who had not dropped out by the last year of primary school could complete it) and grew over time (across successive cohorts of children) – suggesting that the risk of bias in TWFE is larger for the latter outcome.

Data on schooling outcomes comes from the World Bank's World Development Indicators database. My analysis includes data from 1981 through 2019. Data on the timing of the elimination of school fees comes from Koski, Kaufman, Frank, Heymann and Nandi (2018); Latif and Adelman (2021); and Filmer (2023). The countries included in the sample and the years that each country eliminated primary school tuition fees are listed in Table 1. The data set contains 18 countries, but only 14 distinct timing groups – i.e. groups of countries that eliminated school fees in the same year – since Kenya, Madagascar, and Rwanda all eliminated primary school fees in 2003, Burundi and Mozambique both eliminated fees in 2005, and Benin and Lesotho both eliminated fees in 2006.

In Table 2, I report the estimated impact of eliminating school fees on primary school enrollment and completion while controlling for country and year fixed effects. Estimates suggest that introducing free primary education increased gross enrollment in primary school by approximately 20 percentage points (p-value 0.01).[6]. Using the TWFE specification, I do not find evidence that free primary education led to a statistically significant increase in primary school completion: the estimated coefficient is suggests a seven percentage point increase in completion, with a p-value of 0.13.

3.1 Do Treated Observations Receive Negative Weight?

Figure 1 plots histograms of the weights placed on country-year level observations in calculating the TWFE estimate of the treatment effect of eliminating primary school fees. As discussed above, these weights are proportional to the residuals from a regression of the treatment dummy on the set of country and year fixed effects. As expected, the weights sum to zero (across the treatment and control observations), but some *treated* country-year observations receive negative weight, and some untreated country-year observations receive positive weight – so the weights on country-years in the treatment group do not necessarily sum to one.[7] Approximately 22 percent of all treated country-year observations receive negative weight in the estimation of the treatment effect.[8]

[6]These results are consistent with existing evidence from specific implementing countries demonstrating that the elimination of schools fees increased enrollment (cf. Lucas and Mbiti 2012, Njeru et al. 2014)

[7]As in a standard 2×2 DiD specification with no variation in treatment timing, some untreated country-years receive positive weight. This is expected since pre-treatment observations from untreated units always receive positive weight in DiD estimation.

[8]Because data availability differs across country-years for the two outcome variables, the residualized treatment variable D_{it} is not identical in the two specifications (though the treatment variable is the same in both cases). In the analysis of impacts on primary school enrollment 66 out of 277 non-missing treated country-year observations receive negative weight in the calculation of the treatment effect. In the analysis of

Figure 2 illustrates the distribution of negative weights across country-year observations. The unbalanced nature of the panel impacts which treated country-years receive negative weight in the estimation of the treatment effect, and a country-year may be negatively weighted when subsequent years of data from the same country are not. Nevertheless, the country-years receiving negative weight tend to be the later years of data from early-adopter countries. No treated country-year observation from before 2006 receives negative weight, and most observations receiving negative weight (44 of 66 negatively weighted country-years in the case of gross primary enrollment and 40 of 60 negatively weighted country-years in the case of primary school completion) are concentrated among the four countries that implemented free primary in the 1990s.

Of course, the absence of negative weighting does not mean that conventional TWFE estimation will produce the average treatment effect of interest from a policy perspective. As Figure 1 illustrates, the positive weight placed on an individual treatment country-year ranges from close to zero to approximately 0.2 – so some observations are receiving considerably more weight than others in the calculation of the TWFE estimate of the treatment effect. This will not matter if treatment effects are homogeneous across country-years. Moreover, all country-year-level treatment effects have a common sign, then one only runs the risk of producing an estimate of the average treatment effect that is incorrectly signed when some treated observations receive negative weight. Nevertheless, any researcher implementing TWFE should be aware of both the level of variation in the weights placed on treated observations and the evidence for against the hypothesis that treatment effects are heterogeneous.

3.2 Testing the Homogeneity Assumption Directly

We have seen that the TWFE estimate of the impact of free primary education is a linear combination of country-year outcomes, and that in our sample some treated country years receive negative weight in this calculation. As discussed above, negatively weighting

impacts on primary school completion, 65 out of 230 non-missing treated country-year observations receive negative weight in the calculation of the treatment effect.

observations in the treatment group is appropriate if treatment effects are homogeneous, but can lead to bias when treatment effects change over time within treated units. If the homogeneity assumption holds, the relationship between the residualized outcome variable $\tilde{Y}_{it}$ and the residualized treatment variable $\tilde{D}_{it}$ is linear. To see this, consider a balanced panel. Let μ_i denote the value of the outcome variable Y_{it} in country i when $t = 1$, and let η_t be the change in the outcome variable that would occur between period $t-1$ and t in the absence of treatment (which is assumed to be constant across countries under common trends).[9] Let δ denote the homogeneous treatment effect. The value of the outcome variable for country i in year t can be written as

$$Y_{it} = \mu_i + \sum_{\tau=1}^{t} \eta_t + \delta D_{it} \tag{5}$$

and the residualized outcome variable, $\tilde{Y}_{it}$ is equal to $\delta \tilde{D}_{it}$. Hence, under the assumption of homogeneous treatment effects (and common trends), $\tilde{Y}_{it}$ is a linear function of $\tilde{D}_{it}$, and the slope does not differ between the treatment group and the comparison group.

4 Conclusion

In this paper, we use a case study of free primary education in Sub-Saharan Africa to illustrate a set of simple diagnostics and robustness checks that can be used to assess the potential for bias in TWFE. When a common trends assumption holds, the TWFE estimator is a linear combination of outcomes in treated country-years, but in some cases a subset of treated country-years are negatively weighted. One can test for the extent of negative weighting, since the weights are proportional to the residuals from a regression of treatment on country and year fixed effects.

Negative weights are not a problem when treatment effects are homogeneous; in fact, they are a natural consequence of the (implicit) assumption of treatment effect homogeneity in two-way fixed effects. When the assumption of homogeneous treatment effects is valid,

[9] Hence, $\eta_1 = 0$.

the residuals from a regression of the outcome variable on country and year fixed effects are linearly related to the residuals from a regression of treatment on country and year fixed effects – and the slope of this linear relationship does not differ between the treatment group and the comparison group.

When negative weights are present or there is evidence of treatment effect heterogeneity, a range of alternative estimators are available (cf. Callaway and Sant'Anna 2020, Sun and Abraham 2020).[10] However, these estimators may have less statistical power than the pooled estimator, so it may be preferable to use traditional TWFE when the underlying identification assumptions seem plausible. I have proposed a simple test of treatment effect heterogeneity, but researchers should be alert to the possibility that the test could lack sufficient power in their data or context. Statistical reasoning should be combined with introspection in assessing whether the assumptions required for TWFE estimation seem appropriate to the setting.

Even when no treated country-years receive negative weight, the two-way fixed effects estimator does not place equal weight on all treated country-years and imposes a linear dose-response relationship that may be mis-specified (Baker et al. 2021, Sun and Abraham 2020). When impacts are heterogeneous, it is not clear that applied researchers will wish to weight early-adopters more heavily than later-adopters – as would happen if all treated country-years received equal weight. When treatment effects are not homogeneous, many different policy-relevant average treatment effects may exist. Researchers may prefer to characterize the extent and nature of heterogeneity – for example, through event study designs – rather than focusing on a single average treatment effect that may or may not be policy relevant or externally valid, depending on the context. As the range of difference-in-differences related approaches expands, researchers have more opportunity to choose a method that identifies their estimand of interest.

Difference-in-differences estimation through TWFE is one of the most widely used approaches for evaluating social policy. Recent research highlights the potential pitfalls of

[10]See Baker et al. (2021) for an accessible introduction to the estimators proposed by (Callaway and Sant'Anna 2020) and (Sun and Abraham 2020).

two-way fixed effects, but these criticisms should not be interpreted as an indication that two-way fixed effects is no longer a credible identification strategy. Instead, these criticisms and the newly-developed tools that have emerged in parallel provide an analytical framework for more nuanced and better understood difference-in-differences estimation via two-way fixed effects.

References

Angrist, Joshua D. and Jörn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, 2009.

Baker, Andrew, David F. Larcker, and Charles C.Y. Wang, "How Much Should We Trust Staggered Difference-In-Differences Estimates?," Finance Working Paperr 7361, European Corporate Governance Institute 2021.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess, "Revisiting Event-Study Designs: Robust and Efficient Estimation," *Review of Economic Studies*, 2024, *91* (6), 3253–3285.

Callaway, Brantly and Pedro H.C. Sant'Anna, "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, 2020.

Card, David and Alan B. Krueger, *Myth and Measurement: The New Economics of the Minimum Wage*, Princeton University Press, 1995.

Chaisemartin, Clément De and Xavier d'Haultfoeuille, "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, 2020, *110* (9), 2964–96.

Filmer, Deon, "Long-lived consequences of rapid scale-up? The case of free primary education in six Sub-Saharan African countries," *International Journal of Educational Development*, 2023, *102*, 102872.

Gardner, John, Neil Thakral, Linh T. Tô, and Luther Yap, "Two-Stage Difference in Differences," working paper 2024.

Gibbons, Charles E., Juan Carlos Suárez Serrato, and Michael B. Urbancic, "Broken or fixed effects?," *Journal of Econometric Methods*, 2019, *8* (1), 1–12.

Goodman-Bacon, Andrew, "Difference-in-Differences with Variation in Treatment Timing," *Journal of Econometrics*, 2021, *225* (2), 254–277.

Latif, Scherezad and Melissa Adelman, "Free Primary Schooling in the DRC? Where we are on the road to reform," `https://blogs.worldbank.org/en/education/free-primary-schooling-drc-where-we-are-road-reform` August 2021. Accessed: 2026-03-08.

Lucas, Adrienne M and Isaac M Mbiti, "Does Free Primary Education Narrow Dender Differences in Schooling? Evidence from Kenya," *Journal of African Economies*, 2012, *21* (5), 691–722.

Njeru, Wilfred K., Moses N. Muiruri, George M. Njeru, and Ruth Thinguri, "An Investigation on Impact of Free Primary Education on Quality of Education in Kenya Primary Schools," *Journal of Education and Practice*, 2014, *5* (20), 26–31.

Roth, Jonathan, Pedro H.C. Sant'Anna, Alyssa Bilinski, and John Poe, "What's trending in difference-in-differences? A synthesis of the recent econometrics literature," *Journal of Econometrics*, 2023, *235* (2), 2218–2244.

Strumpf, Erin C. Koski Alissaand, Jay S. Kaufman, John Frank, Jody Heymann, and Arijit Nandi, "The Impact of Eliminating Primary School Tuition Fees on Child Marriage in Sub-Saharan Africa: A Quasi-Experimental Evaluation of Policy Changes in 8 Countries," *PLOS One*, 2018, *13* (5), e0197928.

Sun, Liyang and Sarah Abraham, "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*, 2020.

Table 1: Countries Included in the Analysis

Country	FPE Year
Benin	2006
Burkina Faso	2007
Burundi	2005
Cameroon	2000
Democratic Republic of Congo	2019
Ethiopia	1995
Ghana	1996
Kenya	2003
Lesotho	2006
Madagascar	2003
Malawi	1994
Mozambique	2005
Namibia	2013
Rwanda	2003
Tanzania	2001
Togo	2008
Uganda	1997
Zambia	2002

FPE indicates the year free primary education was introduced as national policy. Data on the introduction of free primary education comes from Koski et al. (2018), Latif and Adelman (2021), and Filmer (2023).

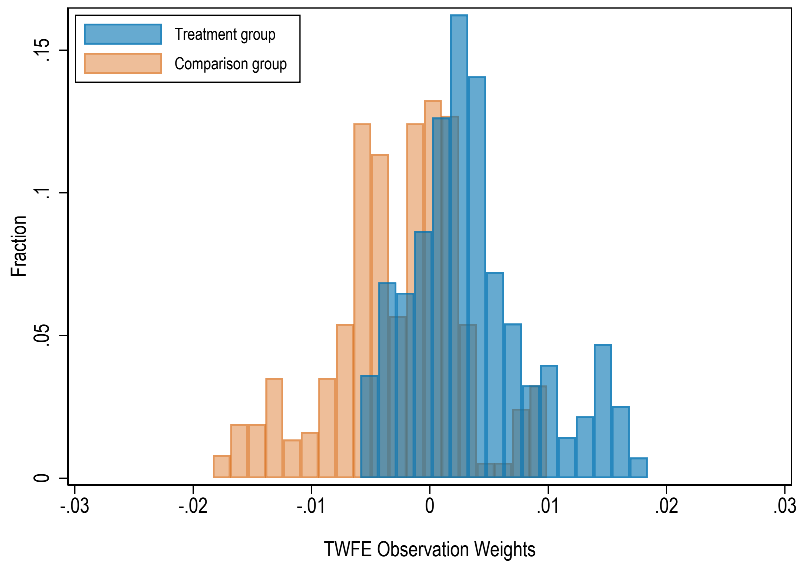
Table 2: Difference-in-Differences Estimates of Impacts of Free Primary Education

	Dep. Var.: Primary School...	
	Enrollment	Completion
	(1)	(2)
Free primary education	19.85	7.06
	(7.06)	(4.41)
	[0.01]	[0.13]
Country fixed effects	Yes	Yes
Year fixed effects	Yes	Yes

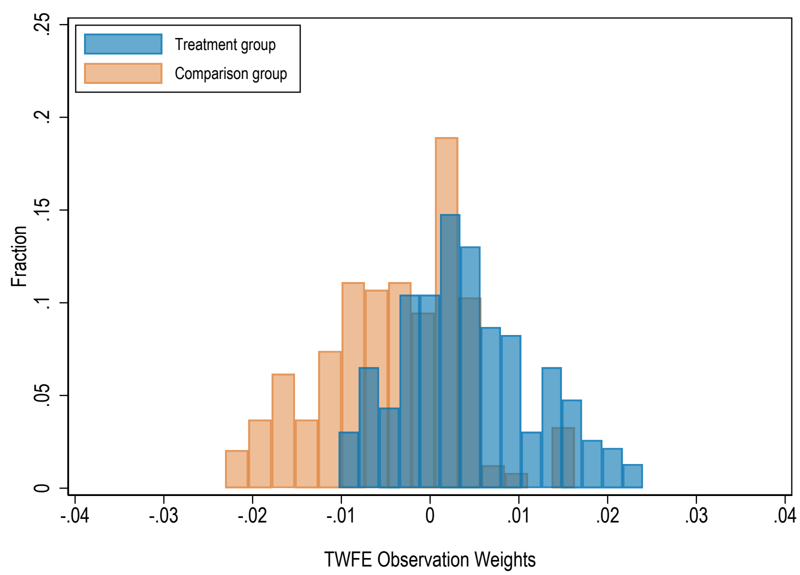
Dependent variable: gross enrollment ratio. Data on gross primary enrollment and primary school completion in 18 countries comes from the World Development Indicators, years 1981 through 2019. Standard errors (clustered at the country level) in parentheses; p-values in brackets.

Figure 1: TWFE Weights, by Treatment Status

Panel A: Dependent Variable: Gross Enrollment in Primary School



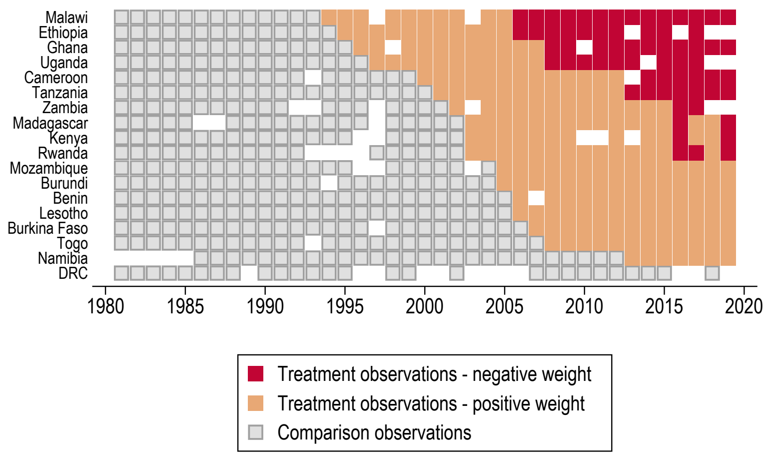
Panel B: Dependent Variable: Primary School Completion



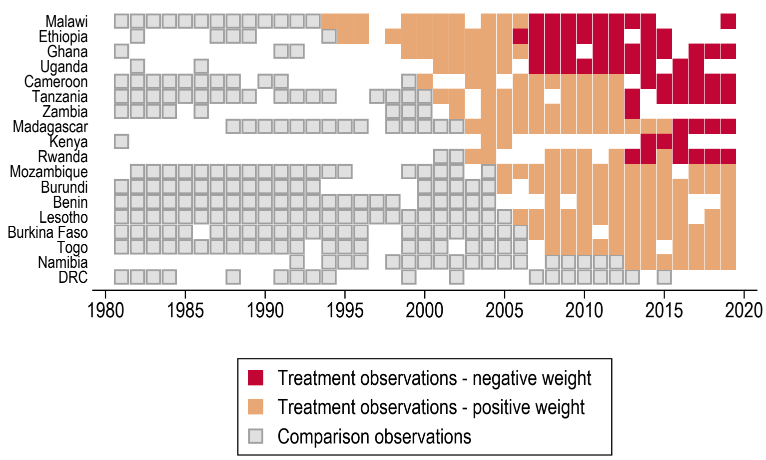
The figure presents histograms of the weights used to calculate the TWFE estimates of the impact of eliminated primary school fees on gross enrollment in primary school (Panel A) and primary school completion (Panel B). The weights are the residuals from a regression of treatment on country and year fixed effects, scaled by the sum of the squared residuals across all observations. See De Chaisemartin and d'Haultfoeuille (2020) for discussion. The weights are not identical in the two specifications because both panels are imbalanced, and the missing country-years differ across the two outcome variables.

Figure 2: Weights Used in TWFE, by Country and Year

Panel A: Dependent Variable: Gross Enrollment in Primary School



Panel B: Dependent Variable: Primary School Completion



The figure characterizes the weights used to calculate the TWFE estimates of the impact of eliminated primary school fees on gross enrollment in primary school (Panel A) and primary school completion (Panel B). The weights are the residuals from a regression of treatment on country and year fixed effects, scaled by the sum of the squared residuals across all observations. See De Chaisemartin and d'Haultfoeuille (2020) for discussion. The weights are not identical in the two specifications because both panels are imbalanced, andt the missing country-years differ across the two outcome variables (as shown in the figure).