

Lima, Gonçalo; Moeller, Jakob; Schmandt, Marco; Westheide, Christian

Working Paper

The Dynamics and Spillovers of Management Interventions: A Comment on Bianchi and Giorcelli (2022)

I4R Discussion Paper Series, No. 66

Provided in Cooperation with:
The Institute for Replication (I4R)

Suggested Citation: Lima, Gonçalo; Moeller, Jakob; Schmandt, Marco; Westheide, Christian (2023) : The Dynamics and Spillovers of Management Interventions: A Comment on Bianchi and Giorcelli (2022), I4R Discussion Paper Series, No. 66, Institute for Replication (I4R), s.l.

This Version is available at:
<https://hdl.handle.net/10419/276960>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



No. 66

I4R DISCUSSION PAPER SERIES

The Dynamics and Spillovers of Management Interventions: A Comment on Bianchi and Giorcelli (2022)

Gonçalo Lima

Jakob Moeller

Marco Schmandt

Christian Westheide

September 2023

I4R DISCUSSION PAPER SERIES

I4R DP No. 66

The Dynamics and Spillovers of Management Interventions: A Comment on Bianchi and Giorcelli (2022)

**Gonçalo Lima¹, Jakob Moeller², Marco Schmandt³,
Christian Westheide⁴**

¹European University Institute, Fiesole/Italy

²Vienna University of Economics and Business, Vienna/Austria

³Technical University of Berlin/Germany

⁴University of Vienna/Austria

SEPTEMBER 2023

Any opinions in this paper are those of the author(s) and not those of the Institute for Replication (I4R). Research published in this series may include views on policy, but I4R takes no institutional policy positions.

I4R Discussion Papers are research papers of the Institute for Replication which are widely circulated to promote replications and meta-scientific work in the social sciences. Provided in cooperation with EconStor, a service of the [ZBW – Leibniz Information Centre for Economics](https://www.zbw.eu/), and [RWI – Leibniz Institute for Economic Research](https://www.rwi-essen.de/), I4R Discussion Papers are among others listed in RePEc (see IDEAS, EconPapers). Complete list of all I4R DPs - downloadable for free at the I4R website.

I4R Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

Editors

Abel Brodeur
University of Ottawa

Anna Dreber
Stockholm School of Economics

Jörg Ankel-Peters
RWI – Leibniz Institute for Economic Research

The Dynamics and Spillovers of Management Interventions: A Comment on Bianchi and Giorcelli (2022) *

Gonçalo Lima, Jakob Moeller, Marco Schmandt, Christian Westheide

September 14, 2023

Abstract

Bianchi and Giorcelli (2022) study the long-term and spillover effects of a management intervention program on firm performance in the US, between 1940 and 1945. The authors find that the Training Within Industry (TWI) program led to positive effects which lasted for at least 10 years. Firm sales of treated firms increased by 5.3% in the first year after implementation, peaking at 21.7% after 8 years, before reducing to 16% gains after a decade. The authors claim that the program generated long-lasting changes in managerial practices. Finally, the program also led to positive spillover effects on the supply chain of treated firms.

First, we reproduce the paper's main findings. Second, we test the robustness of the results to (1) changing the main specification sample and (2) testing other difference-in-differences estimators, using the same data, provided by the authors. We find that the results are robust to these changes. All point estimates in the study remain statistically significant and of similar magnitude.

While the paper's finding reproduce and replicate, challenges in reproducing results we encountered lead us to recommend improvements to journals' code policies.

*Authors: Lima: EUI. E-mail: goncalo.lima@eui.eu, Moeller: WU (Vienna University of Economics and Business). E-mail: jakob.moeller@wu.ac.at, Schmandt: Technical University of Berlin. E-mail: m.schmandt@tu-berlin.de. Westheide: University of Vienna. E-mail: christian.westheide@univie.ac.at. We thank Nicola Bianchi and Michela Girocelli, as well as Abel Brodeur and the Institute for Replication for helpful comments and their cooperation.

1 Introduction

[Bianchi and Giorelli \(2022\)](#), henceforth BG, study the long-term effects of the Training Within Industry (TWI) program on firm performance. The TWI was implemented in the United States (US) between 1940 and 1945. It was a voluntary US government-sponsored program that offered free management training to US firms involved in war production. Training was separated into three different modules: Job-Instructions (J-I) aimed at establishing standard procedures for operations; Job-Relations (J-R) aimed at better managing and motivating workers; and Job-Methods (J-M) aimed at improving production processes. Of the 11,575 applicant firms, 48% did not receive any training. Most trained firms received either one or two of the modules.

BG measure the dynamic effects of TWI on sales, productivity and return on assets (ROA) of treated firms, using an event study design. The authors compare applicant firms that underwent training with those that did not, considering a timeframe of 5 years prior to and 10 years following program implementation. The key identifying assumption is that absent treatment the outcome variables for treated firms would have followed trends parallel to those of untreated firms in post-treatment periods. Furthermore, the authors investigate the specific impacts of each training module, as well as spillover effects on the suppliers of treated firms.

The panel dataset used in the study is constructed by the authors, matching digitized detailed data on firm participation in the program with information on balance sheets and income statements between 1935 and 1955. The dataset used for our analyses is the same as the one provided by the authors in their replication package.

BG describe their main results as: “the effects of the TWI training on firm performance were positive and lasted for at least 10 years after the program implementation (...) [t]he effects on productivity were large and persistent, spanning from a 6% increase within 1 year of the training to a 27% increase after 10 years.”

Point estimates in each period after training were significant at a 1% level.

In this paper, we investigate whether the results presented in BG are reproducible and replicable, and further test the sensitivity of the main results to the following changes: (1) use annual sales and ROA in levels rather than their logarithmic transformations; (2) vary the fixed effects employed in the main specification; (3) employ modern difference-in-difference estimators that are robust to heterogeneity in treatment effects in the presence of staggered adoption of treatment.

Note that a previous reproduction of Figure 3 and Table 4 in the paper has been conducted by Joanne Haddad at the Institute for Replication¹. We also successfully reproduced BG's Figure 3 and Table 4 using their provided code, in Stata. Furthermore, we attempted to reproduce the full set of results using the code and data supplied in the replication package. We were able to reproduce most results after solving some issues related to the versions of the software packages used. We were ultimately able to reproduce the results reported in the main body of the paper. However, some results contained in the appendix could not be reproduced due to missing data, and some discrepancies between the reported results and those obtained in the reproduction remained. We also successfully reproduced the main results of the paper when reimplementing the analysis in R.

Our sensitivity analyses show that, overall, the results are robust to changes to the main specification. First, we change the outcome variables from logs to levels. With this specification, we test whether the finding of no differential pre-trends is invariant to a log transformation². Because the distributions of pre-treatment outcomes in the treatment and control groups are highly similar, we also reject differential pre-trends when running the analysis in levels instead of logs. Moreover, the dynamic long-term effects in levels continue to be large and significant. Second, starting with an identical specification to the one in the paper, we change only the type of fixed effects included in the regression. In particular, we test for the

¹DOI: <https://doi.org/10.48152/ssrp-ap7p-3e29>.

²See, e.g., [Roth and Sant'Anna \(2023\)](#) for a discussion of the strict conditions for the validity of parallel trends to be insensitive to monotonic transformations.

sensitivity of the estimates by iteratively changing the fixed effects included. We find no relevant differences for any of the main outcomes. Third, we find the results to be robust to the use of different difference-in-difference estimators, namely [Sun and Abraham \(2021\)](#) and [Callaway and Sant'Anna \(2021\)](#).

In addition to these tests for the robustness the main results, we conduct two supplementary analyses. First, we test whether the distributions of pre-treatment covariates and outcomes appear consistent with an assignment that is as good as random. Based on a bootstrap that randomly assigns treatment status 5000 times, we find that the differences in the variables observed in the real sample is neither too large nor unexpectedly small compared to the distribution of differences obtained with a random assignment of treatment. Second, we adapt an analysis in the appendix that considers firm exit as an outcome. While this analysis in BG suffers from a survivorship bias favoring treated firms, our modified analysis still shows a positive and significant, if smaller, effect of the training program on firm survival.

Our findings altogether do not cast doubt on any of BG's important findings qualitatively, nor are they quantitatively much different from those reported in the paper.

In addition to the successful reproduction and replication of BG's work, our analyses have yielded another valuable insight. While BG diligently adhered to data and code policies, the reproduction of their results was far from straightforward. We found ourselves depending on older versions of a user-written Stata package, luckily still publicly available. This experience underscores the critical importance of comprehensive code policies that ensure result reproduction without dependence on the backward compatibility of future software versions. Specifically, we urge journals to mandate authors to provide details about the software version used in their study and to include any software modules not part of the base distribution of the used software or programming language. We also contend that appendices significantly contribute to the robustness evaluation of published articles. Therefore, they should be subject to the same data and code policies

2 Reproducibility

2.1 Computational reproducibility based on replication package

Using the authors' Stata code included in the replication package, we attempted to replicate all results contained in the main body of the paper and the appendix. We were ultimately able to reproduce identical results to those in the paper for most of the analyses. In the following subsections, we report on issues encountered when trying to reproduce the results, and on the analyses that yielded different results or that were not replicable with the data and code provided.

2.1.1 Stata versions and packages The code contained in the replication package does not include information on the Stata version used by the authors, nor does it contain install commands for third-party packages used in the code. We fix the latter issue by installing packages, initially from the Boston College Statistical Software Components (SSC) archive, whose missingness raises errors when trying to execute the code.³

With respect to the former issue, we initially attempt to run all the code using Stata 15.1 (released November 2017), and most subsequent issues do not depend on the Stata version employed. However, we encountered slight discrepancies in the results obtained for Table A10, which reports p-values using the bootstrap-based algorithm proposed by [Westfall and Young \(1993\)](#). The code provided does set a random seed, so that we infer that the bootstrap results reported in the paper were produced using a different random number generator. Based on information contained in the [Stata manuals](#), we conclude that the results reported in the paper were likely produced using a Stata version earlier than version 14. When we set the Stata version to 13, we are able to exactly reproduce the results of Table A10. While the effects of using different Stata versions are ultimately inconsequential in the case of BG, this general issue points to the importance of setting the Stata

³Other errors were raised due to a small number of hard-coded file paths and slight naming inconsistencies, which we fixed.

version in replication code.

While using a newer Stata version rarely “breaks” code written for an earlier version, this lack of backward-compatibility is likely more frequent for user-written packages. Moreover, SSC does not provide access to historical versions of such packages, such that it is impossible to recreate the functionality of earlier versions unless the author makes the code available in other ways, e.g., by hosting it on GitHub. Many of BG’s results rely on the highly popular [reghdfe](#) package, which enables linear regressions with high-dimensional fixed effects. The package remains under active development, which means that some of the functionality can be expected to change. We initially run the replication code after installing version 5 of [reghdfe](#), which is currently available via SSC. While we obtained identical point estimates to those reported in BG, for many of the analyses [reghdfe](#) did not provide estimates of the standard errors following the message “Warning: variance matrix is nonsymmetric or highly singular”. To figure out whether the version used was the cause of this issue, we installed the latest version 6 of [reghdfe](#) from [GitHub](#), which, via the option “version”, allowed us to explore differences between the older version 3, the current SSC version 5, and the developer version 6. We were able to reproduce BG’s results when using version 3, whereas the inability to compute standard errors for many analyses was the same for version 5 and 6. We take no stance on whether the standard errors computed using version 3 are accurate despite the concerns expressed by the warning message.

Differences between versions of [reghdfe](#) also occur with respect to the p-values reported in Table 2, panels A and B, and Table A2, panels A and B. These differences result from the multicollinearity of sector dummies contained in the regressions with the county-sector fixed effects. In versions 5 and 6, the sector dummies are excluded due to multicollinearity, whereas this is not the case in version 3, which was employed in producing the results reported in the paper. The sector dummies should not have been included in these regressions in the first place, though the correct p-values, while generally smaller than those reported in the paper, do not

lead to economically important differences in the interpretation of Tables 2 and A2.

Note that the availability of older versions of user-written Stata commands is rather uncommon, such that our successful search for the reasons behind the inability to reproduce results may be impossible in other cases. Thus, more generally, the issue we encountered points to the importance of including files containing the user-written commands employed in replication packages.⁴

2.1.2 Missing code or data The replication code failed to produce a subset of the results because of missing variables or missing code. The original authors helpfully provided us with the required modified data files and code. For Table A7, the code producing the baseline results contained in Panel A is missing. The code supplied to us by the authors successfully reproduced the table. The replication code could not compute the results report in Table A10c because the data does not include a variable named “balanced” that the code requires. Upon being provided with a dataset that, additionally to the data contained in the replication package, includes this variable, we were able to reproduce Table A10c.⁵ Figures A5 B and E could not be reproduced because the data does not include the indicator variable “county” but only the triple interaction “county_sector_time”, which does not allow the reconstruction of the variable “county”. The data provided to us by the authors enabled us to reproduce the figure.

2.1.3 Transcription errors The appendix tables contain several transcription errors and instances of incorrectly rounding up coefficients and p-values. In Table A2, panel D, the p-value for TWI in 1943 should be rounded to 0 (or indicate < 0.01) instead of 0.01. As Tables A8 and A15 contain multiple errors, we reproduce the original and corrected tables in Tables 1 and 2. While the differences do not substantively affect the interpretation of results, these inaccuracies emphasize the

⁴See also the [Stata Coding Guide](#) by Julian Reif and the post [How to make user-written Stata commands really reproducible](#) on the World Bank’s Development Impact blog.

⁵The table contained in the appendix contains one inconsequential transcription error, flipping the p-values for the dependent variables “Log repairs” and “Log injuries” for the predictor “J-R x Post”.

benefits of a fully automated production of tables.

2.1.4 Different results Table A9 reports results for clustering the standard errors at the subdistrict level. While we obtain results for Panels A and B that are identical to those reported in BG, those for Panel C differ. The differences do not lead to qualitatively different conclusions from those reported in BG. However, they cannot be explained by rounding errors or different versions of software packages employed, such that we cannot reconcile the differences in results. We report the original and the reproduced results in Table 3.

2.2 Reproducibility in R

In addition to using the Stata code provided by the authors, we also reproduce the main results of the paper in R. In particular, we reproduce panels A, B, and C of Figure 3, showing the difference-in-differences coefficient estimates, and Table 3, showing the coefficients of the interactions between the training variables and periods 1, 5, and 10. We are able to reproduce the first three panels of Figure 3. We are also able to reproduce the majority of the results in Table 3, with a few exceptions. The coefficients and standard errors we estimate are not always equal to those reported by BG. In almost all cases, however, they are of the same sign and significance levels. We report our estimates in Table 4.

3 Replication

In this section, we apply several sensitivity checks to test whether results are robust to changes in the empirical specifications. First, we test whether the empirical strategy is robust to a monotonic transformation of the outcome variables. We then check for sensitivity of the results to different fixed effects which were not considered in the robustness checks of the original article. Finally, we check whether the results are robust to changes in the estimators used.

For our analysis, unless indicated otherwise, we rely on the same specifica-

tion and difference-in-differences strategy comparing treated to untreated applicant firms, before and after the participation in the training program, specified by Equation 1 in the original article.

All robustness checks were decided upon reading the paper and inspecting the replication package provided by the authors.

3.1 Change in Outcome Variables

We first investigate whether changing the outcome variables changes the validity of the event study approach followed by BG. We rely on the original specification with the exception that we use annual sales and ROA as the dependent variables instead of their logarithmic transformations. As in the main specification, the sample is restricted to firms that just received one or none of the training modules and did not exit the market throughout the period of analysis.

Figure 1 replicates Panels A, C, D and F in Figure 3 of the original article, but with outcome variables expressed in levels. Likewise, Table 6 reports the coefficients for each period of analysis. We find no evidence of differential pre-trends, as most coefficients prior to treatment periods are statistically indistinguishable from zero. We find that untreated firms averaged about 217 million dollars in sales in the period prior to treatment. In the third year after treatment, the treatment effect on annual sales is 25 million dollars (Figure 1, Panel A; Table 6, Column 3). At the same time, the effect on ROA is an increase by about 0.2 percentage points by the third year after treatment, compared to a pre-treatment control mean of 3% ROA (Figure 1, Panel B; Table 6, Column 4). Furthermore, we replicate the first differences for both outcomes in levels, in Panels C and D of Figure 1.

We conclude that the existence of no differential pre-trends is robust to an analysis with outcomes in levels and, more importantly, that the finding of large and statistically strong treatment effects persists when choosing an analysis in levels rather than logs.

3.2 Change in Fixed Effects

Figure 2 depicts dynamic treatment effects for each of the specifications tested in this section and each outcome. Table 7 reports the dynamic treatment effects of the TWI program on log annual sales associated with each specification.

First, we investigate whether the results are sensitive to a specification with firm, time and year fixed effects. Time indicators measure distance in time of a given observation from the first year in which it was treated, expressed in years. By contrast, year indicators indicate the calendar year of each observation. By including year fixed effects, the specification takes into account macroeconomic trends which are not captured by relative time indicators, since the training program is implemented in a staggered way across a 5-year period. We find no substantial differences relative to the specification in the paper, showing that the results in the original article are not being driven by such calendar year effects. Column 2 in Table 7 reports the main coefficients associated with this specification, for the outcome of log annual sales. Results are qualitatively similar to the main specification in the original article (presented in Column 1, Table 7). If anything, insignificant differential pre-trends are closer to zero, whereas treatment effects are slightly larger and more precisely estimated, especially in the first years after treatment.

Second, we investigate whether results change by alternating fixed effects by application window and application day. Figure 2 shows that the results are robust to the inclusion of each of these fixed effects, at a time. Table 7, in particular, shows that, at least for the case of log annual sales, these specifications uncover dynamic treatment effects that are both qualitatively and quantitatively identical to the preferred specification in the original study (Columns 3 and 4)⁶.

3.3 Alternative DID estimators

We investigate the possibility of heterogeneous treatment effects, by testing the robustness of the results to two other difference-in-difference (DID) estimators,

⁶For parsimony, we do not report coefficient tables for log ROA and log TPF.

namely, [Callaway and Sant’Anna \(2021\)](#) and [Sun and Abraham \(2021\)](#). For both estimators, we include firm and calendar year fixed effects. As in the main specification, we restrict the sample to firms that are in the panel throughout the whole period and received at most one of the treatment modules. In neither specification we include covariates.

In recovering the dynamic average treatment on the treated (ATT), we run both estimators comparing treated firms, in a given year, only with those that are never treated. With [Callaway and Sant’Anna \(2021\)](#) we estimate different unique ATTs for each cohort of firms treated at the same point in time. Specifically, for each cohort of firms treated in a given year t , we compute dynamic treatment effects, i.e. the difference for never treated firms in periods $t, t-1, t-2, \dots$, and $t+1, t+2, \dots$. We then aggregate the cohort-period ATTs into a unique ATT by period, replicating the event study design of BG. For inference, we use wild bootstrap to compute standard errors⁷. In turn, [Sun and Abraham \(2021\)](#) show that the leads and lags in event studies using OLS two-way fixed effects (TWFE) type of estimators—as BG’s design—may be contaminated with information from other leads and lags. In the case of treatment effect profile heterogeneity, estimates from OLS will fail to identify the ATT for a given time period. As in [Callaway and Sant’Anna \(2021\)](#), [Sun and Abraham \(2021\)](#) computes a conditional ATT for each cohort-period cell, using, however, an alternative aggregation method for the event study coefficients⁸.

Figure 3 depicts the dynamic treatment effects for each estimator, and each of the three outcomes. The blue circles show, once more, the effects of the main specification in BG, using OLS. The hollow triangles and squares depict the estimates using the alternative DID estimators. Vertical bars display the 95% confidence intervals. Despite minor changes in the magnitude of the effects, the results remain both statistically and economically significant. Especially in the first periods after treatment, the main specification in the paper returns slightly lower coefficients. Still, there is no evidence of differential pre-trends, and the qualitative interpreta-

⁷We use the Stata command `csdid`, to run [Callaway and Sant’Anna \(2021\)](#) estimator.

⁸We use the Stata command `eventstudyinteract`, to run [Sun and Abraham \(2021\)](#) estimator.

tion of the results remains largely unchanged. Therefore, we conclude that the main results of the paper are robust to potentially heterogeneous treatment effects.

3.4 Probability of firm exit and survivorship bias

Table A4 in BG reports tests of whether firms participating in the program differ in their probability of exit. The table shows large reductions in the probability of exit for treated firms. However, the data considered in the regressions reported in the table span the whole period from the beginning of the program until ten years after its end. Since most firms are treated well after the launch of the program, these firms must have necessarily survived for some time, which introduces survivorship bias. We reestimate the exit regressions while limiting the sample to firms that survived at least until the last year of the program. Table 8 shows the original results in Panel A, and the survivorship bias-corrected ones in Panel B. The effects on exit are quantitatively smaller, though still statistically and economically significant.

3.5 Random assignment of treatment

BG assert that there is no systematic difference between treated and non-treated firms prior to the intervention. This assertion is supported by their findings in Panel A of Table 1, which demonstrates that these two groups of firms show hardly any significant differences in their pre-treatment characteristics according to 1939 financial statements, with return on assets (ROA) as the sole exception. The p-values that BG report in Table 1 are calculated based on regressions clustered at the level of subdistricts and application windows. It is debatable whether this clustering, which leads to larger standard errors and thereby makes it less probable to observe significant differences in means, is necessary in this instance. Instead, we conduct tests for mean differences in the selected variables between treated and non-treated firms using t-tests without clustering to test for the sensitivity of the absence of significant differences. The results of our tests are reported in Table 9. Like BG, we consistently fail to reject the null hypothesis of equal means, with the

exception of differences in ROA. Thus, the qualitative results from Table 1 remain valid, even in the absence of clustering, where standard errors would typically be larger.

It is, unfortunately, not possible to test whether including those pre-treatment variables (e.g. ROA) in the main specifications alters the main results (or parallel trends), since it is not possible to merge the provided datasets that include pre-treatment variables (as of 1939) and the dataset containing post-treatment covariates and outcomes.

Considering the fact that there is a statistical difference between treated and control firms for only one in 15 variables, program participation appears to be as good as random. To confirm that the observed small differences between treated and control groups are indeed unremarkable in the hypothetical case of a random allocation, we turn to the bootstrap.

We use 5000 bootstrap iterations, assigning treatment to a randomly selected subset of firms in the sample, keeping the number of treated firms constant at the actual number of treated firms. We then calculate the difference in means for the variables in Panel A of Table 1 in BG between the randomly ‘treated’ and ‘non-treated’ firms. By doing so, we obtain a distribution of the differences in the variables under the assumption of treatment assignment occurring at random. We then compare the true differences between treated and non-treated firms to that distribution. Figure 4 shows the true differences (in red) and the distributions of differences if treatment assignment were random. For all variables, the observed difference between groups is within a plausible range compared to the differences under random assignment. We find no evidence for systematic differences between treated and non-treated firms that are larger, or indeed smaller, than what would be expected if treatment assignment were random.

4 Conclusion

We reproduced and replicated [Bianchi and Giorcelli \(2022\)](#). The authors' replication package allowed the reproduction of the results in the main body of the text, and most of those contained in the appendix. Furthermore, we successfully reproduced the main results in an implementation of our own in R. The results hold in several sensitivity analyses, and we also confirm that the treatment assignment appears as good as random. We commend the authors for the robustness of their important results, and for a replication package that left little to be desired with respect to compliance with the code and data policy.

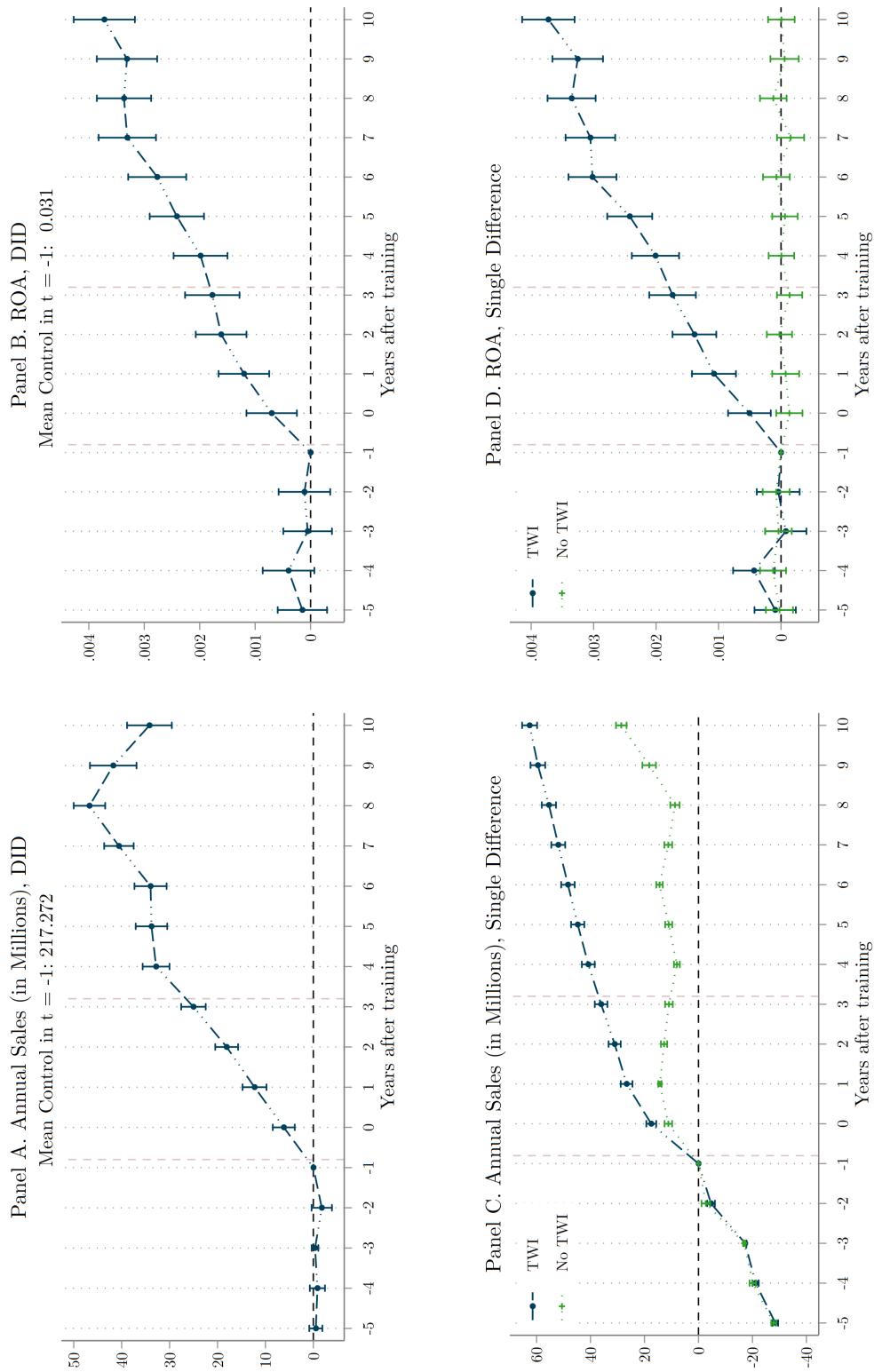
That said, we gained a second more general insight pertaining to journals' code and data policies and their implementation. Although the core functionality of commonly used software packages is typically backward-compatible, exceptions do exist. Consequently, it is crucial for replication packages to specify the software versions used. Further complications arise with packages outside the base distributions; these are more likely to break backward compatibility and previous versions might not always be accessible. We therefore recommend that journals mandate authors to not only specify the versions of these packages used but also include these packages in their replication packages.

References

- Bianchi, N. and M. Giorcelli (2022). The dynamics and spillovers of management interventions: Evidence from the training within industry program. *Journal of Political Economy* 130(6), 1630–1675.
- Callaway, B. and P. H. Sant’Anna (2021, dec). Difference-in-Differences with multiple time periods. *Journal of Econometrics* 225(2), 200–230.
- Roth, J. and P. H. C. Sant’Anna (2023). When Is Parallel Trends Sensitive to Functional Form? *Econometrica* 91(2), 737–747.
- Sun, L. and S. Abraham (2021, dec). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics* 225(2), 175–199.
- Westfall, P. H. and S. S. Young (1993). *Resampling-based multiple testing: Examples and methods for p-value adjustment*, Volume 279. John Wiley & Sons.

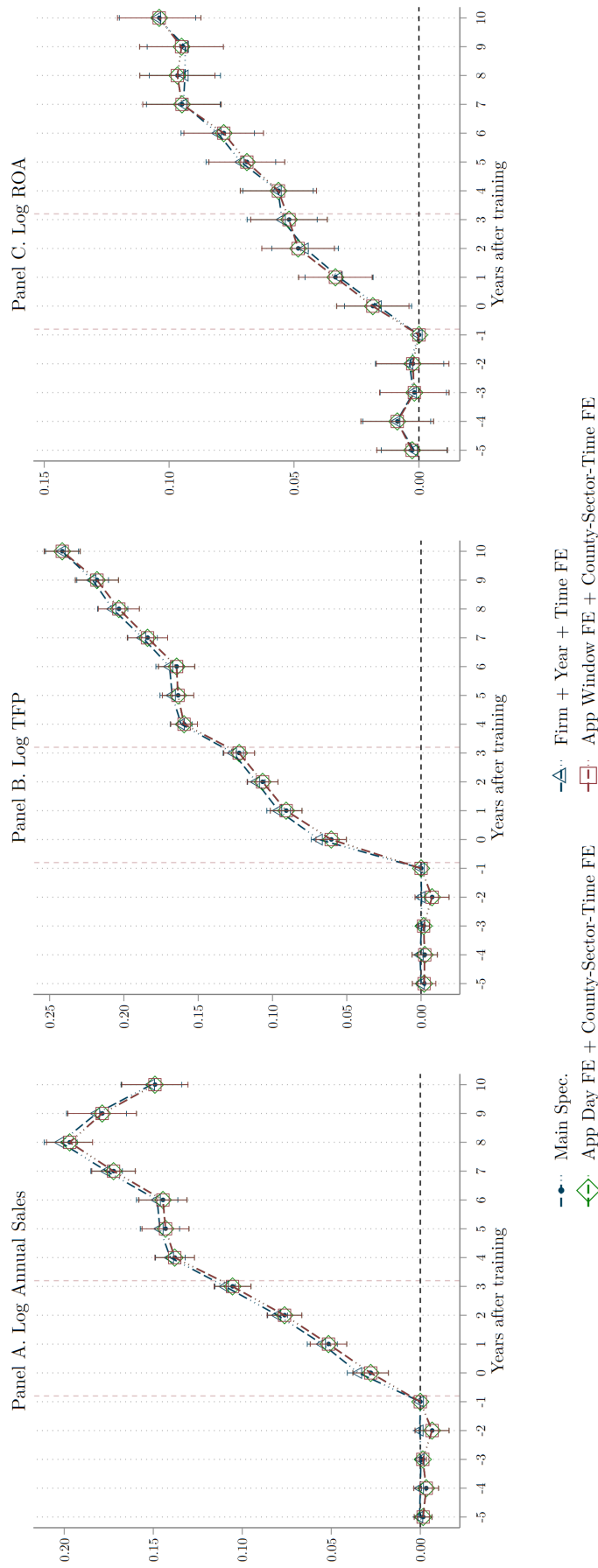
5 Figures

Figure 1: Yearly Effects of TWI Training - Outcomes in Levels



The figure depicts coefficients from an event study, which varies the outcomes from the specification in Equation 1 in [Bianchi and Giorelli \(2022\)](#). Vertical bars represent 95% confidence intervals.

Figure 2: Sensitivity to Different Fixed Effects



The figure depicts coefficients from an event study, which varies the fixed effects included in the specification in Equation 1 in [Bianchi and Giorcelli \(2022\)](#). Vertical bars represent 95% confidence intervals.

Figure 3: Alternative DID Estimators

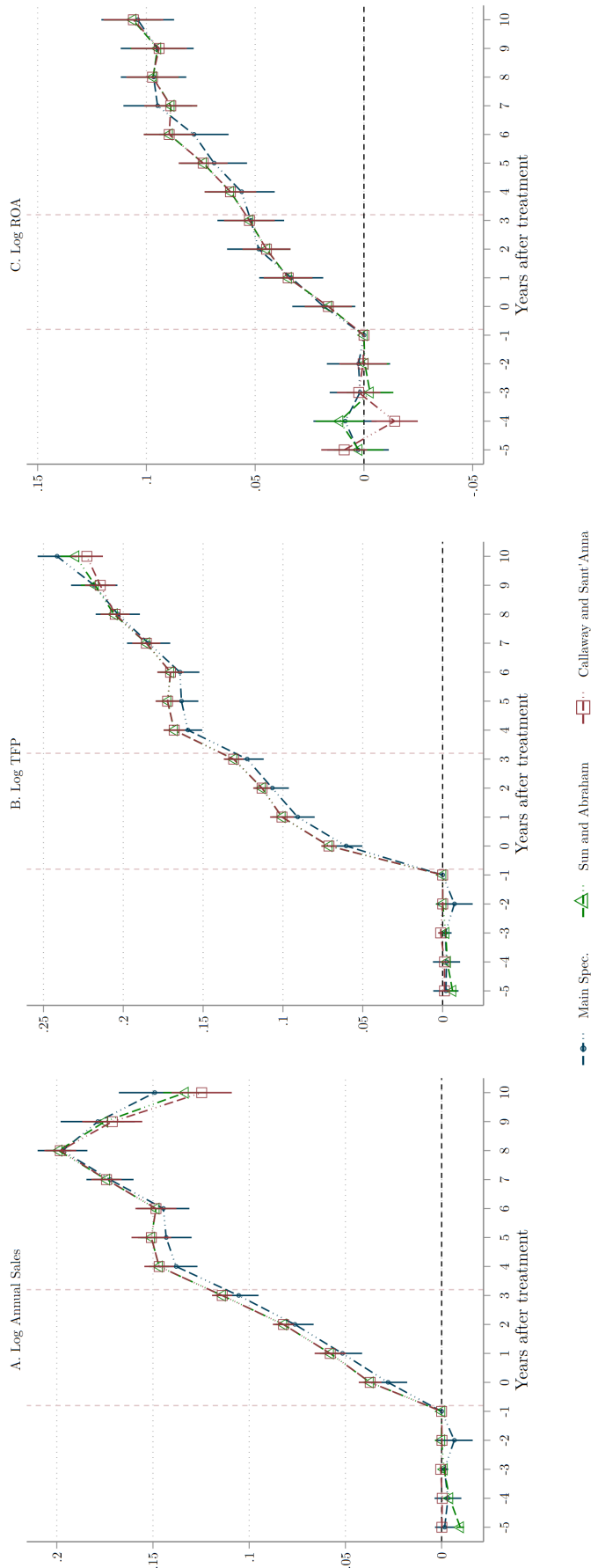
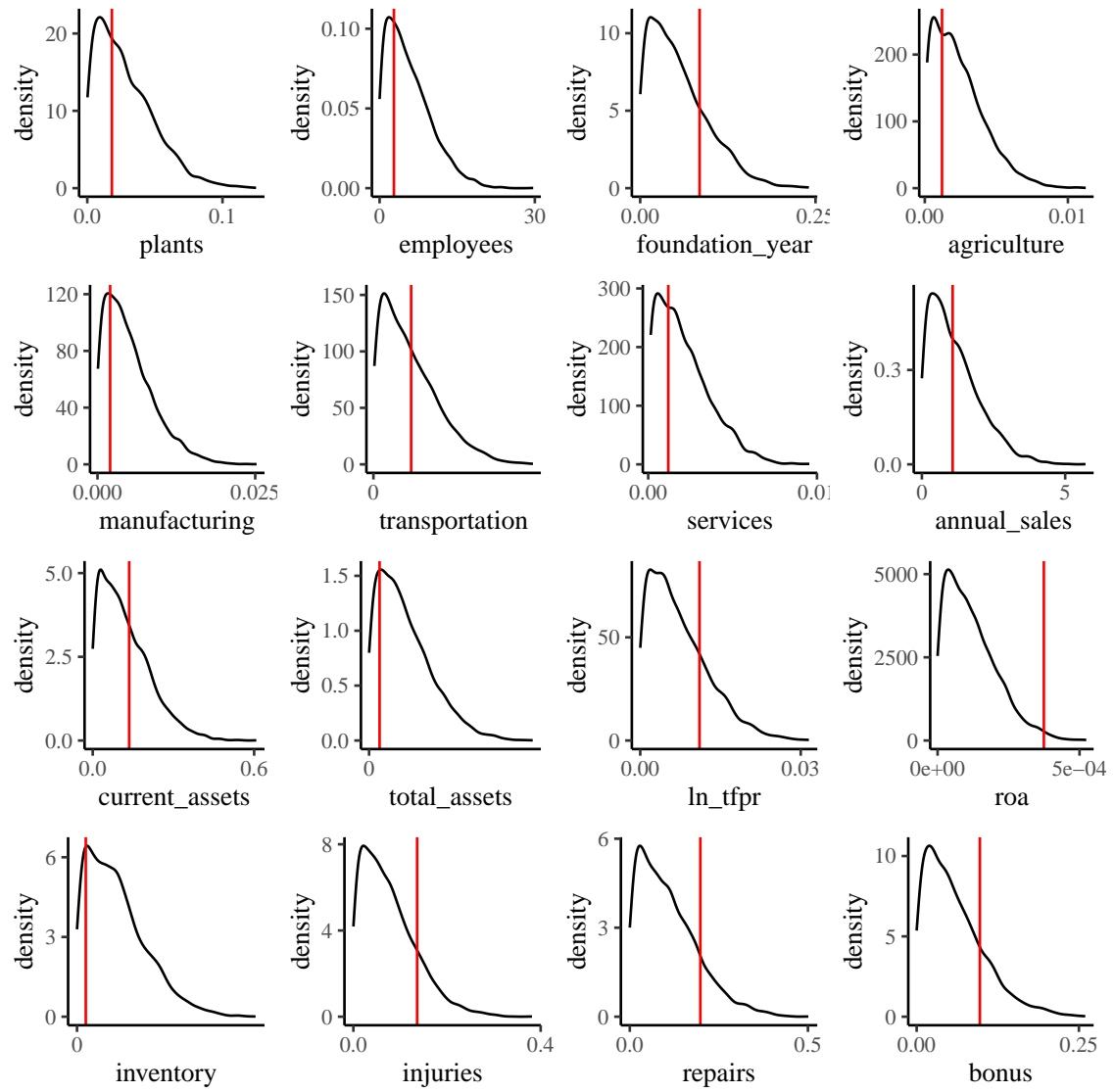


Figure 4: Bootstrapped differences between firms if treatment assignment were random



6 Tables

Table 1: Bianchi and Giorelli Table A8 and reproduction

		Bianchi and Giorelli, Table A8						Reproduction					
		J-I		J-R		J-M		J-I		J-R		J-M	
		Period 0	Period 3	Period 0	Period 3	Period 0	Period 3	Period 0	Period 3	Period 0	Period 3	Period 0	Period 3
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1)	Log machine repairs	-0.089*** (0.001)	-0.357*** (0.006)	0.053*** (0.001)	0.216*** (0.004)	0.053*** (0.001)	0.219*** (0.004)	-0.089*** (0.001)	-0.357*** (0.006)	0.053*** (0.000)	0.216*** (0.004)	0.053*** (0.000)	0.219*** (0.004)
(2)	Log workers' injuries	0.029*** (0.001)	0.118*** (0.004)	0.125*** (0.004)	0.029*** (0.001)	0.030*** (0.000)	0.121*** (0.005)	0.029*** (0.000)	0.118*** (0.004)	0.030*** (0.000)	0.125*** (0.004)	0.030*** (0.000)	0.121*** (0.005)
(3)	Register causes of breakdown	0.877*** (0.009)	0.877*** (0.020)	-0.019*** (0.006)	-0.018*** (0.007)	-0.020*** (0.007)	-0.016*** (0.006)	0.877*** (0.009)	0.877*** (0.020)	-0.019*** (0.006)	-0.018*** (0.007)	-0.020*** (0.007)	-0.016*** (0.006)
(4)	Job description for managers	0.001 (0.001)	0.001 (0.002)	0.876*** (0.006)	0.875*** (0.019)	-0.002 (0.002)	-0.002 (0.003)	0.000 (0.000)	0.000 (0.002)	0.876*** (0.006)	0.875*** (0.019)	-0.002 (0.002)	-0.002 (0.003)
(5)	Job description for workers	-0.019*** (0.007)	-0.020*** (0.008)	0.838*** (0.008)	0.826*** (0.021)	-0.015*** (0.005)	-0.014*** (0.006)	-0.019*** (0.007)	-0.020*** (0.008)	0.838*** (0.008)	0.826*** (0.021)	-0.015*** (0.005)	-0.014*** (0.006)
(6)	Training for workers	-0.025*** (0.007)	-0.030*** (0.009)	0.875*** (0.008)	0.875*** (0.017)	-0.015*** (0.005)	-0.013*** (0.006)	-0.025*** (0.007)	-0.030*** (0.009)	0.875*** (0.008)	0.875*** (0.017)	-0.015*** (0.005)	-0.013*** (0.006)
(7)	Bonus payment scheme	-0.008* (0.004)	-0.009 (0.006)	0.897*** (0.006)	0.902*** (0.015)	-0.007* (0.004)	-0.006 (0.003)	-0.008* (0.004)	-0.009 (0.006)	0.897*** (0.006)	0.902*** (0.015)	-0.007* (0.004)	-0.006 (0.004)
(8)	Suggestions from workers	0.001 (0.001)	-0.002 (0.003)	0.563*** (0.010)	0.552*** (0.026)	-0.002 (0.002)	-0.002 (0.004)	0.000 (0.000)	-0.002 (0.003)	0.563*** (0.010)	0.552*** (0.026)	-0.002 (0.002)	-0.003 (0.004)
(9)	Log unused input	0.039*** (0.001)	0.157*** (0.004)	0.039*** (0.001)	0.158*** (0.004)	-0.089*** (0.000)	-0.348*** (0.004)	0.039*** (0.000)	0.157*** (0.004)	0.039*** (0.000)	0.158*** (0.004)	-0.089*** (0.000)	-0.348*** (0.004)
(10)	Production planning	-0.027*** (0.008)	-0.032*** (0.009)	-0.018** (0.007)	-0.019** (0.008)	0.823*** (0.009)	0.827*** (0.023)	-0.027*** (0.008)	-0.032*** (0.009)	-0.018** (0.007)	-0.019** (0.008)	0.823*** (0.009)	0.827*** (0.023)
(11)	Marketing	-0.008* (0.004)	-0.007 (0.006)	-0.002 (0.002)	-0.001 (0.003)	0.840*** (0.008)	0.828*** (0.020)	-0.008* (0.004)	-0.007 (0.006)	-0.002 (0.002)	-0.001 (0.003)	0.840*** (0.008)	0.828*** (0.020)

This table contains the results shown in table A8 in the appendix to [Bianchi and Giorelli \(2022\)](#) in columns (1) to (6) and the results for the same table obtained from using the replication package in columns (7) to (12).

Table 2: Bianchi and Giorcelli Table A15 columns 5 and 6 and reproduction

	Bianchi and Giorcelli		Reproduction	
	Difference with applicants (1)	Difference with matched applicants (2)	Difference with applicants (3)	Difference with matched applicants (4)
	Panel A: Data from annual financial statements in 1939			
Plants	-3.47	-0.60	-3.47	-0.60
Employees	-420.06	-63.98	-420.06	-63.98
Foundation year	-8.01	-8.09	-8.01	-8.09
Agriculture	0.01	-0.01	0.00	-0.00
Manufacturing	-0.01	-0.01	0.00	-0.00
Transportation	0.01	0.01	0.00	0.00
Services	0.01	0.01	0.00	0.00
Sales	-101.78	-52.94	-101.78	-52.94
Current assets	-17.22	-12.39	-17.22	-12.39
Total assets	-60.81	-43.85	-60.81	-43.85
TFP	0.22	0.05	0.22	0.05
ROA	0.04	0.04	0.04	0.04
Inventory	-14.83	-10.75	-14.83	-10.75
Injuries	-9.52	-6.88	-9.52	-6.88
Repairs	-14.02	-10.09	-14.02	-10.10
Bonus payments	-6.02	-3.95	-6.02	-3.95
	Panel B: Workforce data from replacement lists in 1941			
Share African-Americans	-0.01	-0.01	-0.01	-0.01
Share women	-0.01	-0.01	-0.01	-0.01
Years of education	-1.51	-1.53	-1.51	-1.53
Age of workforce	-5.96	-5.90	-5.96	-5.90
	Panel C: WWII-related data from replacement lists and war contracts			
Share drafted employees (1942-1945)	-0.03	0.01	-0.03	0.01
Switched 3-digit SIC (1940-1945)	0	0.00	-0.02	-0.06
Switched 2-digit SIC (1940-1945)	-0.02	-0.05	-0.02	-0.04
Switched 1-digit SIC (1940-1945)	-0.02	-0.04	-0.02	-0.04
Number of contracts (1940-1945)	1.22	1.21	1.22	1.21
Value of contracts (1940-1945)	-6.85	-9.59	-6.85	-9.59

This table contains the results shown in columns 5 and 6 of table A15 in the appendix to [Bianchi and Giorcelli \(2022\)](#) and the results for the same columns obtained from using the replication package.

Table 3: Bianchi and Giorcelli Table A9 C and reproduction

	Log repairs (1)	Log maintenance (2)	Log injuries (3)	Log bonus (4)	Log strike (5)	Prob training (6)	Log inventory (7)	Log product lines (8)	Prob marketing (9)
A: Bianchi and Giorcelli, Table A9, Panel C									
J-I x Post	0.018 (0.046)	-0.023 (0.052)	0.012 (0.048)	0.035 (0.032)	0.006 (0.033)	-0.071*** (0.022)	-0.037 (0.024)	-0.026 (0.025)	0.002 (0.016)
J-R x Post	0.020 (0.045)	-0.073* (0.041)	-0.006 (0.042)	0.057** (0.028)	-0.139*** (0.038)	0.466*** (0.048)	-0.002 (0.020)	-0.000 (0.020)	-0.009 (0.011)
J-M x Post	0.015 (0.037)	0.012 (0.043)	0.007 (0.039)	0.077** (0.037)	-0.005 (0.024)	-0.025 (0.030)	-0.408*** (0.074)	0.025 (0.018)	0.368*** (0.045)
B: Reproduction									
J-I x Post	0.029 (0.027)	-0.021 (0.035)	0.037 (0.030)	0.032 (0.028)	0.029 (0.024)	-0.027 (0.024)	-0.036*** (0.012)	-0.004 (0.015)	0.019 (0.025)
J-R x Post	0.020 (0.030)	-0.048 (0.038)	-0.007 (0.027)	0.015 (0.024)	-0.113*** (0.022)	0.426*** (0.033)	0.007 (0.012)	-0.002 (0.014)	-0.027 (0.025)
J-M x Post	0.020 (0.025)	0.045 (0.040)	0.009 (0.024)	-0.007 (0.030)	0.021 (0.018)	-0.053** (0.026)	-0.269*** (0.049)	0.020* (0.012)	0.321*** (0.040)

This table contains the results shown in table A9 in the appendix to [Bianchi and Giorcelli \(2022\)](#) in panel A and the results for the same table obtained from using the replication package in panel B.

Table 4: Replication of Table 3 in R

	Log repairs	Log maintenance	Log injuries	Log bonus	Log strike	Prob training	Log inventory	Log product lines	Prob marketing
J-I x Period 1	-0.005 *** (0.000)	0.003 *** (0.000)	-0.017 * (0.01)	0.000 (0.000)	0.000 (0.001)	0.000 (0.000)	-0.01 (0.011)	0.001 (0.003)	0.000 (0.000)
J-I x Period 5	-0.033 *** (0.001)	0.014 *** (0.001)	-0.043 *** (0.011)	0.000 (0.000)	0.006 *** (0.002)	-0.051 *** (0.006)	0.007 (0.011)	-0.002 (0.004)	-0.028 *** (0.006)
J-I x Period 10	-0.057 *** (0.001)	0.025 *** (0.001)	-0.064 *** (0.012)	0.000 (0.000)	0.03 *** (0.004)	-0.062 *** (0.006)	0.008 (0.013)	-0.004 (0.006)	-0.064 *** (0.007)
J-R x Period 1	0.000 (0.000)	0.000 (0.000)	-0.006 (0.011)	0.005 *** (0.000)	0.000 (0.000)	0.000 (0.000)	0.007 (0.01)	0.000 (0.002)	0.000 (0.000)
J-R x Period 5	0.000 (0.000)	0.000 (0.000)	0.011 (0.011)	0.431 *** (0.000)	-0.268 *** (0.007)	0.628 *** (0.027)	0.008 (0.011)	-0.001 (0.004)	-0.023 *** (0.004)
J-R x Period 10	0.000 (0.000)	0.000 (0.000)	0.000 (0.011)	0.459 *** (0.001)	-0.268 *** (0.007)	0.927 *** (0.009)	0.033 *** (0.011)	0.000 (0.006)	-0.061 *** (0.007)
J-M x Period 1	0.000 (0.000)	0.000 (0.000)	-0.006 (0.01)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	-0.005 (0.005)	0.005 * (0.003)	0.000 (0.000)
J-M x Period 5	0.000 (0.000)	0.000 (0.000)	0.001 (0.01)	0.000 (0.000)	0.007 *** (0.002)	-0.038 *** (0.006)	-0.071 *** (0.006)	0.538 *** (0.009)	0.661 *** (0.026)
J-M x Period 10	0.000 (0.000)	0.000 (0.000)	0.006 (0.01)	0.000 (0.000)	0.013 *** (0.003)	-0.062 *** (0.006)	-0.1 *** (0.008)	0.606 *** (0.009)	0.93 *** (0.009)

Replication of Table 3 in Bianchi and Giorcelli using R. This table shows the coefficients of the interactions between the training variables and three selected period dummies (out of 15 period dummies included). The omitted period is the year before the TWI training (period 21). Standard errors in parentheses.

Table 5: Bianchi and Giorcelli Table 3 original

	Log repairs	Log maintenance	Log injuries	Log bonus	Log strike	Prob training	Log inventory	Log product lines	Prob marketing
J-I x Period 1	-.010*** (.000)	.005*** (.000)	-.020** (.010)	-.000 (.000)	.000 (.001)	-.001 (.001)	-.010 (.011)	.002 (.003)	-.001 (.001)
J-I x Period 5	-.033*** (.001)	.014*** (.001)	-.043*** (.010)	-.000 (.000)	.006*** (.002)	-.051*** (.006)	.007 (.010)	-.002 (.004)	-.028*** (.005)
J-I x Period 10	-.057*** (.001)	.025*** (.001)	-.064*** (.011)	.000 (.000)	.030*** (.004)	-.062*** (.006)	.008 (.012)	-.004 (.005)	-.064*** (.007)
J-R x Period 1	-.000 (.000)	.000 (.000)	.004 (.010)	.414*** (.000)	2 (.000)	.043*** (.010)	-.005 (.011)	.002 (.002)	.001* (.001)
J-R x Period 5	-.000 (.000)	.000 (.000)	.011 (.010)	.431*** (.000)	-.268*** (.006)	.628*** (.025)	.008 (.010)	-.001 (.004)	-.023*** (.004)
J-R x Period 10	-.000 (.000)	-.000 (.000)	-.000 (.009)	.459*** (.001)	-.268*** (.006)	.927*** (.008)	.033*** (.010)	-.000 (.005)	-.061*** (.006)
J-M x Period 1	.000 (.000)	.000 (.000)	-.001 (.010)	.000 (.000)	.000 (.000)	.001 (.001)	-.015*** (.005)	.015*** (.004)	.043*** (.010)
J-M x Period 5	-.000 (.000)	.000 (.000)	.001 (.009)	-.000 (.000)	.007*** (.002)	-.038*** (.005)	-.071*** (.006)	.538*** (.009)	.661*** (.024)
J-M x Period 10	-.000 (.000)	.000 (.000)	.006 (.009)	-.000 (.000)	.013*** (.002)	-.062*** (.006)	-.100*** (.007)	.606*** (.009)	.930*** (.008)
Observations	67,472	67,472	67,472	67,472	67,472	67,472	67,472	67,472	67,472
R ²	.142	.131	.141	.170	.165	.559	.406	.445	.648

Table 3 as found in Bianchi and Giorcelli.

Table 6: Yearly Effects of TWI Training - Outcomes in Levels

	(1)	(2)	(3)	(4)
	Log Annual Sales	Log ROA	Annual Sales (in Millions)	ROA
Period -5	-0.0016 (0.0025)	0.0027 (0.0072)	-0.52 (0.70)	0.00015 (0.00023)
Period -4	-0.0033 (0.0035)	0.0087 (0.0074)	-0.84 (0.81)	0.00040* (0.00024)
Period -3	-0.0014 (0.0010)	0.0018 (0.0071)	-0.35 (0.35)	0.000053 (0.00022)
Period -2	-0.0066 (0.0049)	0.0025 (0.0074)	-1.76 (1.08)	0.00011 (0.00024)
Period 0	0.028*** (0.0051)	0.018** (0.0074)	6.17*** (1.17)	0.00070*** (0.00023)
Period 1	0.052*** (0.0052)	0.033*** (0.0075)	12.3*** (1.27)	0.0012*** (0.00023)
Period 2	0.076*** (0.0050)	0.048*** (0.0074)	18.1*** (1.22)	0.0016*** (0.00023)
Period 3	0.11*** (0.0053)	0.052*** (0.0078)	25.0*** (1.31)	0.0018*** (0.00025)
Period 4	0.14*** (0.0057)	0.056*** (0.0078)	32.8*** (1.43)	0.0020*** (0.00025)
Period 5	0.14*** (0.0067)	0.069*** (0.0077)	33.8*** (1.68)	0.0024*** (0.00025)
Period 6	0.14*** (0.0069)	0.078*** (0.0081)	34.0*** (1.71)	0.0028*** (0.00027)
Period 7	0.17*** (0.0062)	0.095*** (0.0080)	40.6*** (1.57)	0.0033*** (0.00026)
Period 8	0.20*** (0.0066)	0.097*** (0.0077)	46.7*** (1.68)	0.0034*** (0.00025)
Period 9	0.18*** (0.0098)	0.095*** (0.0086)	41.8*** (2.48)	0.0033*** (0.00028)
Period 10	0.15*** (0.0095)	0.10*** (0.0085)	34.2*** (2.38)	0.0037*** (0.00028)
Observations	67,472	67,472	67,472	67,472
Sample	Same as Figure 3	Same as Figure 3	Same as Figure 3	Same as Figure 3
Mean Control Group in $t = -1$	19.126	-3.492	217.272	.031

The table reports event study coefficients for a specification identical to Equation 1, and reported in Figure 3 of [Bianchi and Giorcelli \(2022\)](#), using the outcomes annual sales (Column 3) and ROA (Column 4) in levels rather than logs, as in the original article. Columns 1 and 2 reproduce the results in the original article. Standard errors are reported in parentheses.

* $p < .10$, ** $p < .05$, *** $p < .01$.

Table 7: Sensitivity to Different Fixed Effects - Log Annual Sales

	(1)	(2)	(3)	(4)
Period -5	-0.0016 (0.0025)	0.00042 (0.00043)	-0.0016 (0.0025)	-0.0016 (0.0025)
Period -4	-0.0033 (0.0035)	0.000017 (0.00034)	-0.0033 (0.0035)	-0.0033 (0.0035)
Period -3	-0.0014 (0.0010)	-0.00036 (0.00027)	-0.0014 (0.0010)	-0.0014 (0.0010)
Period -2	-0.0066 (0.0049)	0.00025 (0.00021)	-0.0066 (0.0049)	-0.0066 (0.0049)
Period 0	0.028*** (0.0051)	0.035*** (0.0032)	0.028*** (0.0051)	0.028*** (0.0051)
Period 1	0.052*** (0.0052)	0.055*** (0.0043)	0.052*** (0.0052)	0.052*** (0.0052)
Period 2	0.076*** (0.0050)	0.080*** (0.0028)	0.076*** (0.0050)	0.076*** (0.0050)
Period 3	0.11*** (0.0053)	0.11*** (0.0029)	0.11*** (0.0053)	0.11*** (0.0053)
Period 4	0.14*** (0.0057)	0.14*** (0.0042)	0.14*** (0.0057)	0.14*** (0.0057)
Period 5	0.14*** (0.0067)	0.15*** (0.0056)	0.14*** (0.0067)	0.14*** (0.0067)
Period 6	0.14*** (0.0069)	0.15*** (0.0059)	0.14*** (0.0069)	0.14*** (0.0069)
Period 7	0.17*** (0.0062)	0.18*** (0.0045)	0.17*** (0.0062)	0.17*** (0.0062)
Period 8	0.20*** (0.0066)	0.20*** (0.0047)	0.20*** (0.0066)	0.20*** (0.0066)
Period 9	0.18*** (0.0098)	0.18*** (0.0086)	0.18*** (0.0098)	0.18*** (0.0098)
Period 10	0.15*** (0.0095)	0.15*** (0.0087)	0.15*** (0.0095)	0.15*** (0.0095)
Observations	67,472	67,472	67,472	67,472
Application Window FE	Yes	No	No	Yes
Application Day FE	Yes	No	Yes	No
County-Sector-Time FE	Yes	No	Yes	Yes
Time FE	No	Yes	No	No
Year FE	No	Yes	No	No
Firm FE	No	Yes	No	No

The table reports event study coefficients for a specification identical to Equation 1, and reported in Figure 3 of [Bianchi and Giorelli \(2022\)](#), for the outcome of log annual sales, alternating the fixed effects included. Column 1 reproduces the results in the original article. Standard errors are reported in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table 8: Bianchi and Giorcelli Table A4 and reproduction taking survivorship bias into account

	Exit	Exit	Exit	Exit	Exit	Exit
	(1)	(2)	(3)	(4)	(5)	(6)
A: Bianchi and Giorcelli, Table A4						
TWI	-0.257*** (0.013)	-0.249*** (0.012)	-0.254*** (0.013)			
J-I				-0.221*** (0.019)	-0.223*** (0.017)	-0.225*** (0.018)
J-R				-0.258*** (0.019)	-0.242*** (0.017)	-0.245*** (0.019)
J-M				-0.288*** (0.017)	-0.280*** (0.015)	-0.288*** (0.017)
B: Table A4 with sample adjusted for survivorship bias						
TWI	-0.199*** (0.013)	-0.193*** (0.012)	-0.198*** (0.013)			
J-I				-0.169*** (0.019)	-0.172*** (0.017)	-0.175*** (0.018)
J-R				-0.195*** (0.019)	-0.183*** (0.017)	-0.183*** (0.019)
J-M				-0.229*** (0.016)	-0.222*** (0.015)	-0.232*** (0.016)

This table contains the results shown in table A4 in the appendix to [Bianchi and Giorcelli \(2022\)](#) in panel A and the results for the same table obtained from adjusting the sample so as to correct for the survivorship bias in the treated sample in panel B.

Table 9: Replication of Table 1, Panel A

	Trained Firms Mean	Nontrained Firms Mean	Difference	t-Value	p-Value
Plants	6.09	6.07	-0.02	-0.51	0.61
Employees	1036.99	1039.77	2.77	0.39	0.70
Foundation year	1930.96	1931.05	0.08	1.22	0.22
Agriculture	0.03	0.03	-0.00	-0.41	0.69
Manufacturing	0.86	0.86	-0.00	-0.31	0.75
Transportation	0.09	0.09	0.00	0.82	0.41
Services	0.02	0.02	-0.00	-0.45	0.65
Sales	191.27	192.34	1.07	0.74	0.46
Current assets	18.47	18.33	-0.14	-0.91	0.36
Total assets	64.40	64.51	0.11	0.23	0.81
TFP	3.87	3.86	-0.01	-1.22	0.22
ROA	0.03	0.03	0.00	2.50	0.01 *
Inventory	15.36	15.38	0.02	0.18	0.86
Injuries	10.33	10.47	0.14	1.43	0.15
Repairs	15.28	15.48	0.20	1.46	0.14
Bonus payments	8.22	8.32	0.10	1.32	0.19

Replication of Bianchi and Giorcelli Table 1, Panel A. Differences in means between trained and nontrained firms.

Table 10: Bianchi and Giorcelli Table 1, Panel A, Columns 5, 6, and 7 original

	Trained Firms Mean	Nontrained Firms Mean	Difference p-Value
Plants	6.09	6.07	.588
Employees	1,036.99	1,039.77	.686
Foundation year	1930.96	1931.05	.229
Agriculture	.03	.03	.678
Manufacturing	.86	.86	.746
Transportation	.09	.09	.400
Services	.02	.02	.624
Sales	191.27	192.34	.444
Current assets	18.47	18.33	.341
Total assets	64.40	64.51	.808
TFP	3.87	3.86	.316
ROA	.03	.03	.010
Inventory	15.36	15.38	.854
Injuries	10.33	10.47	.141
Repairs	15.28	15.48	.134
Bonus payments	8.22	8.32	.176

Means for trained and untrained firms as found in Bianchi and Giorcelli Table 1, Panel A.