RIETI Discussion Paper Series 24-E-025

Cascades of Tax Policy through Production Networks: Evidence from Japan

KOIZUMI, Hideto
RIETI

RIETI
Research Institute of Economy, Trade & Industry, IAA
The Research Institute of Economy, Trade and Industry
https://www.rieti.go.jp/en/

RIETI Discussion Paper Series 24-E-025

February 2024

Cascades of Tax Policy through Production Networks: Evidence from Japan*

Hideto Koizumi

Research Institute of Economy, Trade and Industry

Hitotsubashi University

Abstract

The effectiveness of tax policies targeting firms has been evaluated conventionally based on the effects on the firms that are directly affected by the tax policies. However, the indirect effects through the supply chains of the directly affected firms can also be of first-order importance. This paper estimates the indirect effects on firm performance of tax incentives for investment through production networks, exploiting the quasi-experimental event of an investment stimulus policy targeting small and medium enterprises and unique proprietary data of supply chains in Japan. After confirming the direct effects, I find evidence suggesting that the indirect effects on direct suppliers are even larger than the direct effects, while no discernible effects are found on downstream firms. The absence of downstream effects appears to stem from the fact that treated firms crowd out the market share of untreated large firms, leading to an insufficient change in market prices. In total, while the tax policy successfully stimulates the targeted small firms, its spillover effects are primarily confined to the upstream customers which tend to be large.

Keywords: Spillover Effects of Tax Policies, Bonus Depreciation, Production Networks

JEL classification: H25

*This study is conducted as a part of research at the Research Institute of Economy, Trade and Industry (RIETI).
. I would like to thank participants of the RIETI DP Seminar, where the draft of this paper was presented, for their helpful comments.
The author is grateful to Teikoku Data Bank (TDB) and TDB Center for Advanced Empirical Research on Enterprise and Economy, Graduate School of Economics, Hitotsubashi University (TDB-CAREE) for the provision of data sets. The author also thanks Masatoshi Abe, Yasushi Hara, and Yoshiki Hiramine for extensive research assistance. This project is funded by JSPS KAKENHI Grant Number 21K13312.

1 Introduction

The effectiveness of tax policies targeting firms has been evaluated conventionally by the effects on firms directly affected by the tax policies. However, the indirect effects through the supply chains of the affected firms can be of first-order importance. In particular, this paper focuses on an investment stimulative tax policy. An increase in output, induced by heightened capital investment due to the tax policy, must be followed by a greater demand for the actual ingredients of output to boost production (e.g., screws required for assembling a car), leading to potential upstream propagation. Simultaneously, an exogenous increase in the supply of the directly affected firms' output could drive down the market price of their products, potentially benefiting downstream firms as well. This paper estimates these indirect effects, exploiting a quasi-experimental event of an investment stimulus and unique proprietary data of supply chains in Japan.

The policy I study, "bonus" depreciation for small and medium enterprises (SMEs), allows SMEs to deduct an additional percentage of capital expenditures in the first year of an asset's tax life. Although this tax policy was not specifically directed at particular industries, there is a variation in the degree of policy benefits emerging from the discounting factor and the fact that longer-lived assets experience a larger reduction in the present value cost of investment. This is because bonus depreciation accelerates deductions from further in the future. Therefore, those firms in the industries that tend to use such longer-lived assets are affected by the policy more than other firms. Using this quasi-experimental variation in the treatment intensity and the firm-level input-output linkage data, I construct the supply chains of those firms with greater policy exposures and investigate the spillover effects of the tax policy.

Exploiting the natural experimental setting and unique data set, I found statistically significant and economically meaningful direct *and* indirect effects of tax incentives for investment. The extensive micro-data on inter-firm transactions allow me to trace and quantify the extent of cascades along supply chains. Specifically, I observed that SMEs among the direct suppliers experience a sales increase that is half the magnitude of the treatment effects. Additionally, there is weak evidence suggesting that the estimate for average direct suppliers, typically large in size, is even greater than the estimate for the treated firms.[1] The fact that the estimate for the spillover effects on the direct suppliers is larger than the direct effects may spoil the justification for the policy restriction to SMEs.

Even if the policymakers had aimed at stimulating the entire economy somehow through stimulating SMEs, the findings of this paper demonstrate the asymmetry of the spillovers through the production networks. In contrast to the literature indicating significant negative effects of a natural disaster on downstream firms, I do not find statistically significant effects on the downstream firms. Although alternative channels, such as mark-ups, market elasticities, and increased sales to final consumers, could potentially play a role, my findings reveal that treated firms increased the number of direct customers following the policy change. This increase might have occurred through the displacement of market share previously held by large firms in the same industries and markets, which were ineligible for the policy.

These findings are developed in two steps. The first stage is the estimation of direct effects. Zwick and Mahon (2017) (hitherto, ZM) estimated the direct effects of bonus depreciation in the U.S. by comparing the performance of firms in industries that, on average, invest in long-lived assets with those in industries that invest in short-lived assets. Their exposure measure is at the industry level. For this study, I apply to Japan their established exposure measure. Despite a difference in class life definition between the U.S. and Japan, a manual check of all listed class lives of equipment between the two countries reveals a high positive correlation, roughly at 0.7. To mitigate potential measurement errors arising from the discrepancies, and to construct the supply chain of the treated firms, I follow Garrett et al. (2020) in constructing a binary treatment variable. This approach

[1] Note that I cannot formally test that the difference between spillover effects and direct effects is statistically significant, due to the lack of a theory for the heterogeneous analysis of the doubly-robust difference-in-differences estimators that I will explain in the later sections.

ensures that, as long as Japanese and U.S. firms tend to purchase similar equipment, the policy measure I employ can capture the direct effects of the policy.

However, since I hypothesize the presence of spillovers through production networks, simply using all the firms in industries that tend to invest in short-lived assets as the control group would contaminate the estimate. Then, to overcome this issue, I follow Carvalho et al. (2021). Based on a theoretical prediction/assumption that shock propagation decays as subjects are farther away from the shock source, they estimated the spillovers of a negative shock caused by a natural disaster through production networks by comparing the directly unaffected firms that are close to the directly affected firms to those directly unaffected firms that are far distant away in the directly affected firms' supply chains using a difference-in-differences (DID) framework. Similarly, I compare the sales values of the treated firms to the sales values of those untreated firms that are far distant away in the treated firms' supply chains using a DID framework.[2]

The assumption behind this econometric design is the parallel trend of outcomes between the treated and control group firms in the absence of the policy, meaning that my measure of policy exposure is not correlated with other shocks which coincide with the implementation of bonus depreciation and affect sales. I address this identification threat in the following ways. First, I graphically demonstrate that changes in sales are uncorrelated with the policy exposure prior to the initial implementation. Second, I show that my results are robust after controlling for (medium-level) subsector-by-year-by-prefecture fixed effects, implying that (1) the threat coming from differential trends across subsectors is limited, (2) prefecture-level policies or shocks do not confound my estimates, and (3) local subsector trend is also not a confouding factor. With the empirical strategy, my baseline estimation shows that treated firms' sales grew by approximately 4 to 5% after the policy implemented in 1998 for 1993-2003 sample depending on specifications.

[2]Since the vast majority of firms are indirect transaction partners to each other, few firms are connected to a certain set of firms in general. Then, I resort to their identification strategy that uses firms that are distant from the treated firms in the firms' production network.

Given the first-stage result, I examine whether the presence of direct and indirect input-output linkages to the treated firms had an impact on firms' performance in the years after the tax policy. In particular, I compare the post-policy sales of firms at different distances—in the supply chain network sense—from the affected firms to a control group of firms that are relatively more distant. This is also similar to the approach used in Carvalho et al. (2021). The identification threat to this specification is that the direct or indirect connection to the affected firms correlates with other shocks which coincide with the implementation of bonus depreciation through which sales are affected. This threat is particularly concerning since large firms tend to trade with many more firms than small firms do, so these firms tend to be the direct transaction partners of the affected firms.

To alleviate this issue, I resort to conditional DID methods. These methods generally rely on two assumptions. One is the so-called conditional parallel trend assumption which states that in the absence of treatment, the average outcome of the treated and control groups would have evolved in parallel conditional upon covariates that predict the treatment status. The other is an overlap assumption which states that at least there is a small probability that some fraction of the population is treated and that for each value of the covariates, there is at least a small likelihood that the unit lies in the control group.[3]

In particular, I follow the doubly-robust difference-in-differences (DRDID) estimators proposed by Sant'Anna and Zhao (2020) and use the multi-period version of DRDID suggested by Callaway and Sant'Anna (2021), both of which are extended versions of the semiparametric DID with inverse probability weighting (IPW) approach suggested by Abadie (2005). In contrast to the semiparametric DID approach that requires the model for propensity score to be correctly specified, the DRDID approach is valid if either the propensity score model or the outcome estimation model is correct. For the propensity score model, I use covariates relevant to firm sizes together with physical distance to treated firms since Bernard et al. (2019) empirically demonstrate that physical distance

[3]This assumption is to circumvent the so-called "irregular identification" problem as discussed in Khan and Tamer (2010).

plays a major role in forming a transaction link between firms.

I support the conditional parallel trend assumption in the standard graphical demonstration of the pre-trend. Using DRDID estimates, I graphically show no differential trend in sales among firms at different distances in supply chains prior to the policy shock, except the indirect suppliers. As for the indirect suppliers, the effects seem to be within the extrapolated linear pre-trend in the spirits of Dobkin et al. (2018) up to some post periods, and therefore I will refrain from making a causal inference over the estimate for the indirect suppliers. Furthermore, following Garrett et al. (2020), I try different definitions of the treatment status. I confirm that the results remain unchanged. Although the assumption underlying the research design is fundamentally unverifiable, my empirical strategies and robustness checks significantly alleviate the identification threat.

These results contribute to two strands of literature. One is a growing literature that studies the impacts of investment stimulative tax policies. The previous literature extensively examines the direct effects of such tax policies (e.g., among many papers, closely related ones are Hall and Jorgenson (1967); Cummins et al. (1994); House and Shapiro (2008); Edgerton (2010); Zwick and Mahon (2017); Ohrn (2018); Ohrn (2019); Fan and Liu (2020); Guceri and Liu (2019); Garrett et al. (2020); Curtis et al. (2021); Tuzel and Zhang (2021)) in various settings. This study is the first contribution to the literature by demonstrating the significance of indirect effects through production networks.

The other is a growing literature of production networks. Among related papers,[4] Carvalho et al. (2021) study the cascading effects of the Great Earthquake in Japan through production networks using a similar data set. I use a similar identification strategy that they propose. Liu (2019) develops a rich model of industrial policies and production net-

[4]See, e.g., Acemoglu et al. (2016) for industry-level analyses; Barrot and Sauvagnat (2016) for input-specificity analyses; Boehm et al. (2019) for firm-level cross-country transmission; Demir et al. (2022) for the propagation of financial shocks by liquidity-constrained importers; Carvalho et al. (2018) for the propagation of demand shocks on innovation activities; Ozdagli and Weber (2017) for monetary policy shocks; Auer et al. (2019) for inflation comovements across countries; Takafumi et al. (2022) study how the economy can quickly manage the supply chain disruptions with a cut in the chain by finding new trade partners using the Great East Japan Earthquake of 2011.

works and extensively tests it with empirics. Estimating parameters of his model, he computes the predicted aggregate impacts of industrial policies in China and South Korea, considering the propagating effects of the historical sectoral policies through industry-level input-output linkages. This paper differs from his study in that (a) he studies sectoral policies targeting a particular sector unlike tax incentives for investment and (b) my study focuses on reduced-form estimates of the spillover effects through granular firm-level supply chains rather than a rich model prediction in Liu (2019). Thus, my study complements these studies in the production network literature. My paper uses granular firm-level network data and provides distinct results that unlike negative shocks such as a natural disaster, positive shocks including investment stimulus may not cascade over downstream firms if the shocks affect only a portion of firms in a market.

2 Policy Background

The Japanese government implemented 30% accelerated depreciation in June 1998 for small and medium enterprises (SMEs) ("*Chushokigyo Toshi Sokushin Zeisei*") as part of "*Sogo Keizai Seisaku*", to stimulate the economy by encouraging SMEs to buy more machinery and equipment. This was a response to Asian Currency Crisis in 1997. This policy was implemented together with increased taxable base that effectively decreased the corporate tax rates from 37.5 to 34.5% for non-SMEs and from 28 to 25% for SMEs in 1998. The effective corporate tax rates further decreased to 30% for non-SMEs and 22% for SMEs in 1999 without changing the taxable base. While the detailed definition of SMEs under these corporate-targeted tax policies can be found in the Appendix, the main difference between SMEs and large corporations is whether a firm has more than 100 million yen common stock (or initial capital for most private firms).[5] Those SMEs with common stock less than 30 million yen can choose between 30% bonus depreciation

[5]Common stock is part of shareholders' equity in balance sheets.

and 7% tax credit of eligible investment expenditures. The bonus depreciation and tax credit policy has been implemented every year since 1998. The bonus rate, tax credit rate, and corporate tax rates were not changed during the period of my study sample.

While the record from the period between 1998 and 2003 is not available, the report by the Ministry of Finance from the earliest possible period (FY 2011) shows that among all the applications for the bonus depreciation and tax credits of this policy for SMEs, approximately 95% (198.2 billion yen) of the total amounts was through bonus depreciation and the 5% (11.5 billion yen) was tax credits.[6] Note that all the policies above are implemented together and are endogenous of the macroeconomic shock which impedes the identification of direct and indirect effects through a simple pre-post time-series estimation approach. Thus, one needs an identification method that is plausibly exogenous to the shock.

2.1 Treatment Intensity Measure Constructed by ZM

To address this issue, I follow ZM who estimate the direct effects of bonus depreciation in the U.S. In the absence of bonus depreciation, the Modified Accelerated Cost Recovery System (MACRS) in the U.S. lays out tax rules for the depreciation of newly purchased assets. The present value of depreciation deductions associated with $1 of investment can be expressed as

$$z^0 = \sum_{t=0}^{T} \frac{1}{(1+r)^t} D_t,$$

where T is the class-life of the asset, D_t is the fraction of the dollar that is depreciated in year t, and r corresponds to the rate used to discount future cash flows. MACRS rules determine T and D_t in each period for each type of investment. Longer-lived assets are depreciated more slowly over longer lives and have smaller z^0s than shorter-lived assets.

[6] Retrieved on January 25, 2024 at https://warp.da.ndl.go.jp/info:ndljp/pid/10311345/www.mof.go.jp/tax_policy/reference/stm_report/houkoku01.pdf. The name of the report is "sozeitokubetsusochi no tekiyou jittaichosa no kekka ni kansuru hokokusho."

Then, tax deductions provided by longer-lived assets are generally less than shorter-lived assets in present value terms.

Bonus depreciation allows firms to immediately write off v percent of eligible investments. The remaining $1v$ percent are depreciated following the MACRS rules. Then, the policy reduces the present value cost of investment by $v(1z^0)$. Since this effect is larger for the asset with smaller z^0—i.e., assets with longer class-lives—, and thus z^0 captures a measure of bonus depreciation treatment intensity.

ZM compute an industry-level measure of z^0 in the following procedure. They first obtain z^0 for each asset class defined by MACRS using 7 percent as the discount rate. Next, they use administrative tax return data on sample firms to calculate the share of each eligible asset class purchased at each 4-digit NAICS industry level. Finally, ZM weight the asset-class-level z^0s by the industry shares to create z_j^0 for each industry j, which measures the present value of depreciation deductions for the average asset industry j purchases. As noted in Garrett et al. (2020), there is a considerable variation in z_j^0's even within a specific sector. Later, I conduct a similar exercise after applying this measure to Japanese industries.

2.2 Application of ZM Treatment Measure to Japanese Industries

Using the US administrative data, ZM computes the industry-level (continuous) exposure measure of U.S. bonus depreciation in 2002. Assuming that US and Japanese firms in the same industry tend to buy similar equipment and machinery, I apply their measure to Japanese industries.[7] I conducted a thorough manual comparison of of all the asset items for depreciation between the U.S. in 2002 and Japan in 1998 and finds a high positive correlation at 0.7. With this, as long as Japanese and U.S. firms have a tendency to purchase similar equipment if they are in the same industry, the policy measure I use can

[7] Komori (2003) compares Japanese and US depreciation systems. His paper implies that while there are differences in the length of depreciation periods for many goods, the length tends to be similar (e.g., computer is 5 years in US and 6 years in Japan.)

capture the direct effects of the policy in Japan.[8]

I use concordance tables provided by the United Nations to match Japanese industries categorized by 4-digit JSIC with U.S. industries categorized by 4-digit NAICS.[9] Using this concordance, I assign the treatment intensity measures to each of JSIC industries that have corresponding industries to 4-digit NAICS industries listed in the ZM industry-level treatment intensity file.[10]

Partly to reduce the measurement error coming from a discrepancy, I follow Garrett et al. (2020) to construct a binary treatment variable. They categorize industries into "treated" industries if they are in the bottom third of the z_j^0 distribution. According to them, they discretize the treatment variable for two reasons. First, it removes the effects of outliers in the z_j^0 distribution—e.g., the power generation industry has a z_j^0 that is much lower than other industries. Second, z_j^0 values rely on an assumption about the discount rate. Their discretized treatment measure eliminates this assumption. This paper also discretizes its treatment measure not just for the two reasons but also for mitigating measurement errors caused by differences between U.S. and Japanese industries with respect to purchasing patterns of eligible equipment. I use the 33rd percentile as the cutoff for the balance across the sample sizes of different network distances including the treatment and control group. I conduct robustness checks with 25th and 40th percentiles of the z_j^0 distribution as cutoffs. I find that the results remain unchanged.

Note that I do not observe which SMEs chose bonus depreciation over tax credits. If there is somehow a correlation between the depreciation schedule and the tendency to choose tax credits over bonus depreciation, then this may cause an upward or downward bias in my estimates. However, given the disproportionate usage of bonus depreciation over tax credits as mentioned above, the bias is expected to be negligible if there is any.

[8]Not exactly the same, but Japan used similar depreciation rules to MARCS in 1998. See Komori (2003).

[9]The crosswalks are available at https://unstats.un.org/unsd/classifications/Econ. There is no direct crosswalk between JSIC and NAICS, and therefore, I first use a crosswalk between ISIC and NAICS, and then I use a crosswalk between ISIC and JSIC.

[10]When there are multiple JSIC industries corresponding to a single 4-digit NAICS industry, I take a simple average of the intensity measures.

3 Data

I rely on a proprietary dataset compiled by the private credit reporting agency Teikoku Data Bank Ltd. (henceforth, TDB) to construct a firm-level production network of supplier-customer linkages. Generally speaking, firms give information to TDB in the course of obtaining credit reports on potential suppliers and customers. This information contains a set of firm-level characteristics, together with the identities of the firms' suppliers and customers. The TDB data uses its only industry codes that closely match the Japan Standard Industrial Classification (JSIS). TDB collects firm data on employment, sales, capital stock, and the location of the firm's headquarters. Firms in their data set report the date on which its fiscal year ends as well. On the other hand, TDB started collecting data on investment only after 2000 with a restricted sample.

The TDB sample is neither a census nor a representative survey since the entry of any particular firm occurs at the request of TDB's clients. Consequently, TDB does not annually update the data on every firm. Thus, I restrict my sample to the subset of firms which report sales figures and firm-level covariates for all 11 years between 1993 and 2003. I start with 1993 fiscal year (FY) since the TDB industry code started using in 1993 a new industry code that match JSIS more closely than before at disaggregate levels. Given the starting year, and given that expanding a period greatly reduces sample size (especially, small firms), I set the ending year at FY 2003.[11] This leaves us with a balanced panel data of 424,367 firms across all the prefectures in Japan.

To examine biases in the sample, I compare the 1997 TDB dataset with 1999 Economic Census. I chose FY 1997 for TDB dataset since the main policy of interests started in June 1998, and 1999 Economics Census is the census conducted closest to this time period. Figure 1 displays the comparison of industrial composition at the JSIS major classification level. As it shows, there is no major difference between the two. Furthermore, Figure 1 shows the comparison of geographic distribution at the prefecture level using the same

[11] I tried different year ranges, and the results remain unchanged.

Table 1: Firm Size Distribution

	Number of Employees									
	$0-4$	$5-9$	$10-19$	$20-29$	$30-49$	$50-99$	$100-299$	$300-999$	$1000-1999$	$2000+$
TDB	0.19	0.23	0.23	0.11	0.10	0.08	0.05	0.01	0.003	0.002
Census	0.50	0.21	0.14	0.05	0.04	0.03	0.02	0.01	0.001	0.001

Notes: This table reports the fraction of firms with the number of employees in each of the respective bins. "TDB" refers to the 1997 TDB dataset. "Census" refers to the 1999 Economic Census for Business Frame.

two datasets. We can see no major difference between the two.

On the other hand, partly because I use a relatively long period (11 years) and require the dataset to be balanced, the final dataset contains very small firms much fewer than the 1999 census. Table 1 shows the proportion of firms in each size bin based on the number of employees. We can see that the proportion of firms with less than 5 employees is disproportionately small in the TDB sample compared to the census. Thus, the results of this paper should be interpreted with caution. Given that ZM find that smaller firms face greater liquidity constraints and thus benefit more from the bonus depreciation policy, at least the direct effects may be underestimated in this paper.

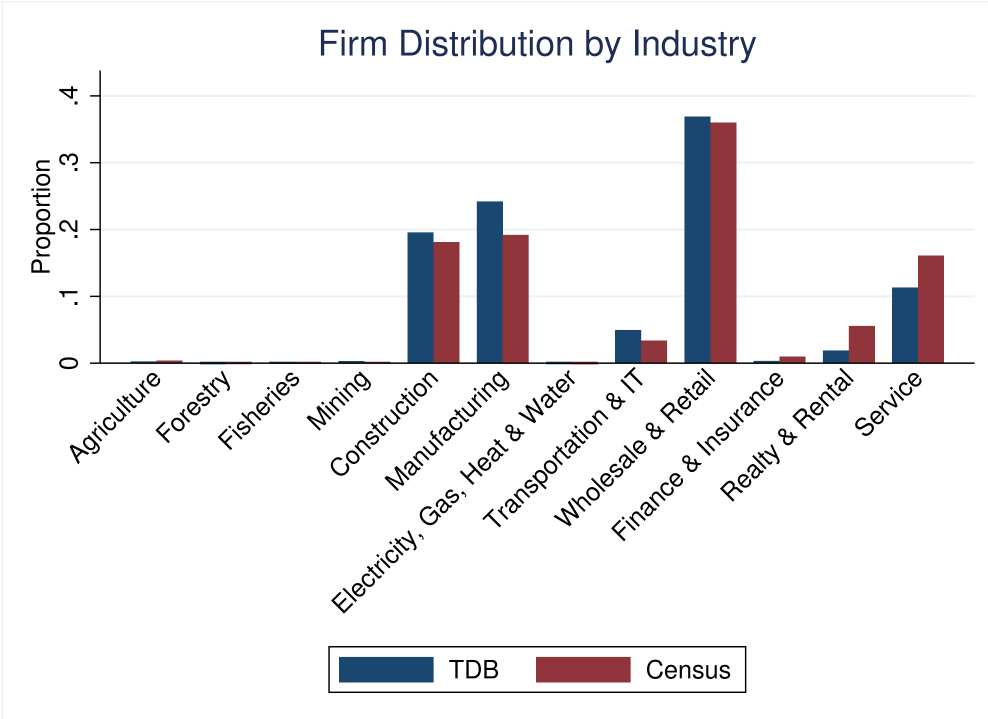
In addition to the basic statistics, each firm in the TDB dataset also provides a list of its transaction partners, allowing me to construct the production network of supplier-customer linkages for the firms in my sample. Given the occurrence of the tax policy in June 1998, I construct this network using the transaction data collected as of January 1998.

The TDB-based supplier-customer linkages have two limitations. First, the data only captures a binary measure of inter-firm supplier-customer relations. While the data contains information on whether one firm is another firm's supplier or customer, I do not observe a yen measure on their transaction volume.

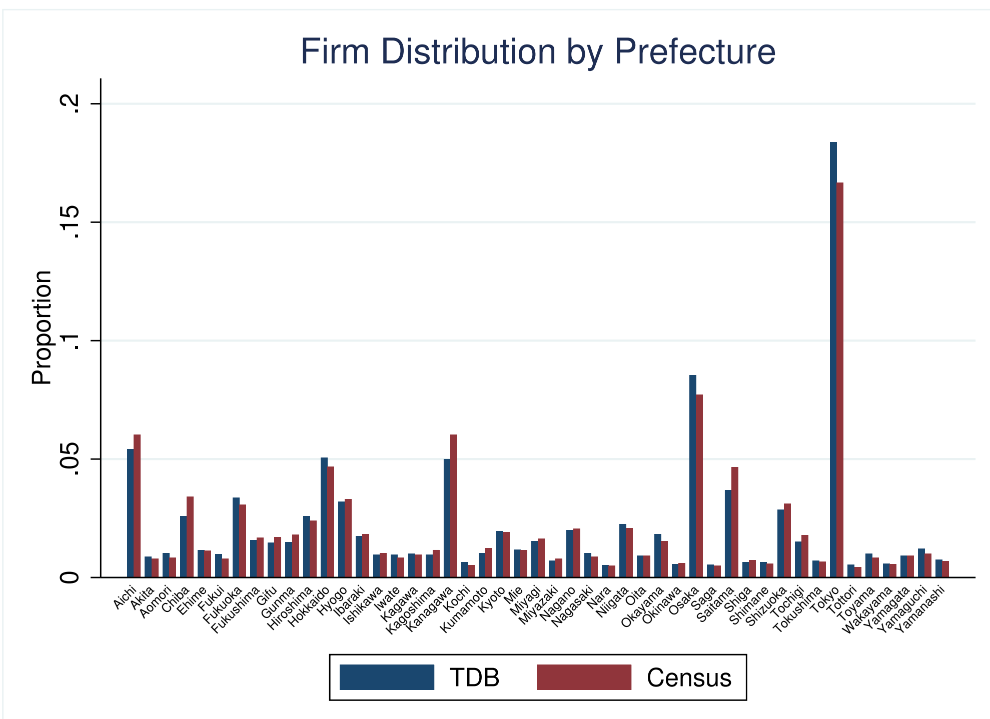
Second, the forms used by TDB limit the number of suppliers and customers that firms can report to 9 each. Nonetheless, given that each firm in the dataset may also be reported by other firms as a transaction partner, I overcome this limitation by augmenting the customer and supplier relations with those reported by other firms. That is, I construct a firm's transaction network by supplementing the list of suppliers (customers) reported

Figure 1: Industrial Composition and Geographical Distribution

(a) Comparison of Industrial Composition between TDB and Census



(b) Comparison of Prefecture Composition between TDB and Census



by the firm itself with the reports of other firms that state the firm as their customer (supplier). This procedure leads to the list of suppliers and customers of firms that have more than 9 transaction partners per category, including gigantic firms that transact with several thousand firms. With this procedure, I restrict my sample to the subset of firms that have at least one transaction partner within the TDB database.

With this data structure, I construct a measure of network distance to the set of directed affected industry firms for all firms in my sample. Exploiting the 1998 production network data, I first label the immediate customers and suppliers of treated firms as, respectively, “downstream distance 1” and “upstream distance 1” firms. Similarly, I then designate a firm as “downstream distance 2” if it was listed in 1998 as a customer of at least one downstream distance 1 firm and was not a distance 1 firm itself. With a similar recursive procedure, I identify the set of firms at various upstream and downstream distances from treated industry firms right before the intervention of policy intervention. To retain a sufficient number of observations for the control group, I use those firms which are at distance 3 or more away from the treated industry firms. One may think that those large firms in the treated industries can be a control group since they were not qualified for the policy. However, these large firms tend to transact with many more firms than SMEs and tend to be distance 1 or 2 away firms in the dataset. Thus, given that I expect spillover effects through production networks, I cannot use them as a control group.[12]

Figure 2 displays the location of the headquarters of the firms in my sample. It shows that the sample firms are not concentrated in a particular region, alleviating the concern on a geographic bias in the supply chain.

Table 2 summarizes baseline characteristics of firms in fiscal year 1997. One can tell that those direct suppliers and customers of the treated firms are relatively larger than the rest of the groups and have noticeably more transaction partners. This is because those

[12]While large firms are known to change transaction partners frequently, SMEs change less frequently. Since the treated firms are SMEs, 82% of the direct customers remain as the direct customers of the treated firms in the post-treatment period, while 80% remain as the direct suppliers.

Figure 2: Location of Firms

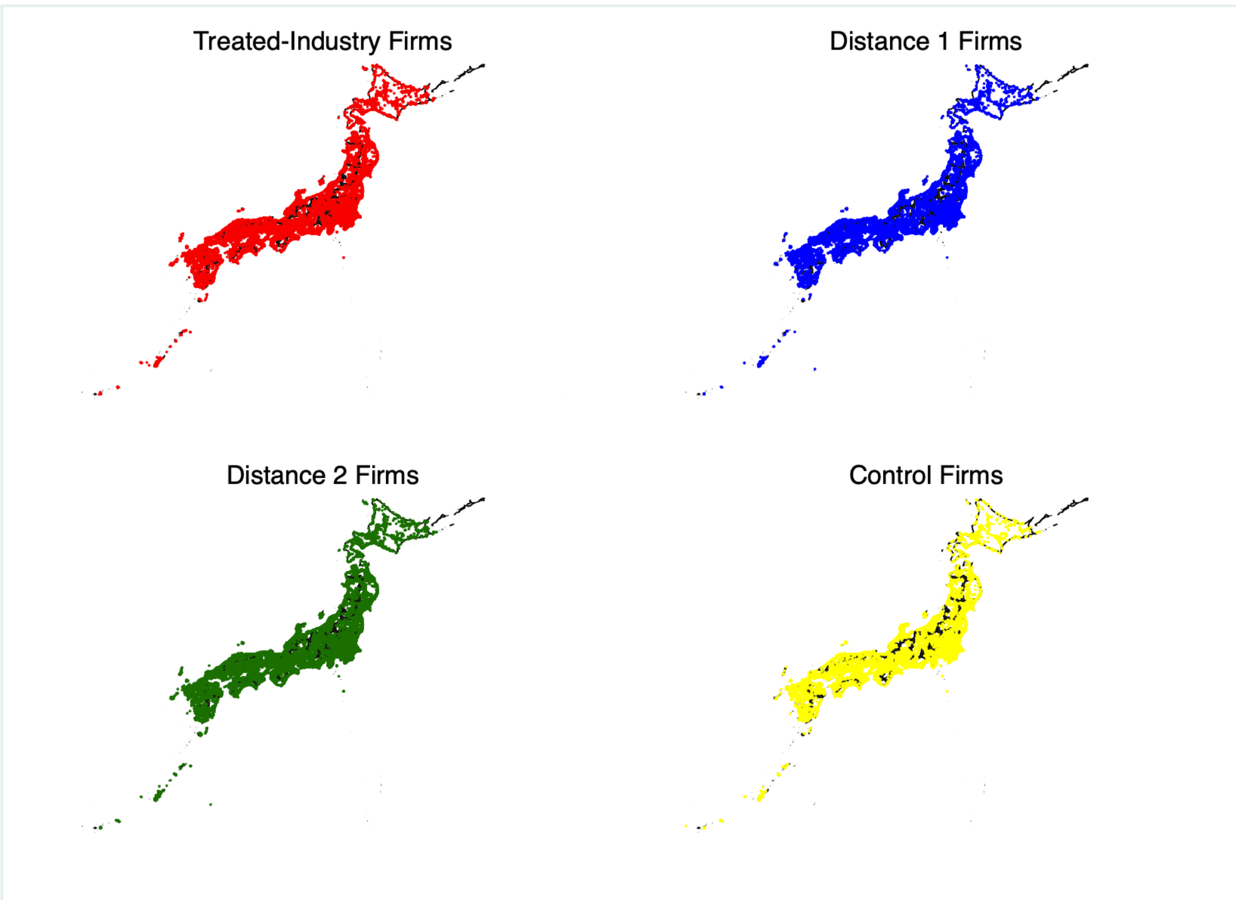


Table 2: Summary Statistics by Distance

	Treated	Upstream 1	Downstream 1	Upstream 2	Downstream 2	Control
Log of Sales	6.00	6.84	6.88	6.23	6.09	5.32
	(1.28)	(1.77)	(1.77)	(1.40)	(1.34)	(1.12)
Log of Employment	2.79	3.15	3.34	2.93	2.70	2.24
	(1.07)	(1.46)	(1.49)	(1.20)	(1.14)	(0.94)
Log of Capital Stock	9.40	9.97	10.14	9.69	9.52	9.16
	(0.88)	(1.57)	(1.63)	(1.13)	(1.05)	(0.90)
Age	27.91	29.75	29.72	25.99	25.89	22.40
	(12.99)	(14.48)	(14.46)	(12.86)	(12.75)	(11.85)
N of Suppliers	2.75	12.22	11.30	3.79	2.89	1.61
	(6.51)	(75.41)	(63.58)	(8.81)	(2.79)	(1.09)
N of Customers	3.40	16.13	14.49	3.70	4.37	1.75
	(6.45)	(71.13)	(70.27)	(7.33)	(10.84)	(1.70)
Customers' log sales	16.16	16.77	15.78	16.39	17.18	14.50
	(2.52)	(2.42)	(2.25)	(2.41)	(2.11)	(1.66)
Suppliers' log sales	16.30	15.80	16.76	18.31	16.83	14.75
	(2.69)	(2.41)	(2.73)	(2.33)	(2.84)	(2.12)
Observations	75684	42590	56158	126995	172618	65860

large firms tend to transact with many more firms than SMEs. These large firms may have differential trends compared to SME control group firms, and therefore I overcome the selection issue by an alternative approach that allows for covariate-specific trends, which will be discussed in the next section. My main covariates for the selection model include the size-related variables and the physical distance from the treated firms with which these firms transact since physical distance is known to be one of the primary factors of firm linkages.

With this TDB dataset, I apply the ZM treatment intensity measures to TDB-defined industries (hitherto TDB industries) that closely follow 4-digit JSIC.[13] As done by Garrett et al. (2020), I compute the coefficient of variation within each sector. Figure B1 demonstrates a considerable variation in the z_j^0's even when the ZM measures are applied to Japanese firms by showing the coefficient of variation within each sector normalized by that of the manufacturing sector.

[13]When there are multiple 4-digit JSIC industries corresponding to a single TDB industry, I take a simple average of the ZM intensity measures. If there are multiple TDB industries to a single JSIC industry, I assign the same ZM intensity measure for this JSIC industry to all the corresponding TDB industries.

4 Econometric Framework

There are two steps: estimation of the direct effects and that of the indirect effects.

4.1 Direct Effects

First, I estimate the direct effects of the tax policy on firms' performance.

$$Y_{ipNt} = \alpha_i + \delta_{pt} + \beta D_{N,t} + \varepsilon_{ipNt}, \tag{1}$$

where Y_{ipNt} is firm i's outcome variable such as sales value in natural logarithmic scale at year t which lies in industry N and is located at prefecture p, and $D_{N,t}$ is the difference-in-differences term that is equal to 1 for the treated firms at period between 1998 and 2003 and 0 in the other periods and 0 for the firms that are 3 or more distance away in the treated firms' transaction networks throughout all the periods. All the regressions in this paper are clustered at the TDB industry level that corresponds to 4-digit NAICS levels, to address the concerns raised by Bertrand et al. (2004) about errors being correlated within policy units (industries). As noted above, we cannot use large firms in the treated industries as a control group since these large firms tend to transact with many more firms than SMEs and tend to be distance 1 or 2 away firms in the dataset. Thus, given that I expect spillover effects through production networks, I cannot use them as a control group.

The identification threat to the estimation of β is the lack of a parallel trend on the outcome variable prior to the policy intervention in 1998. Considering the sensitivity issue raised by Roth and Sant'Anna (2022), I chose the natural logarithmic scale for the outcome variables since there is a large and skewed heterogeneity in firms' scales. Thus, scale-free percentage change measures across time periods are more suitable to justify the underlying assumption.

As supports for the parallel trend assumption, I provide pre-trend analysis results. I estimate the following pre-trend analysis specification;

$$Y_{ipNt} = \kappa_i + \kappa_{pt} + \sum_{t \neq 1993} \beta_t D_N \lambda_t + \hat{\varepsilon}_{ipNt}, \tag{2}$$

where $D_N = 1$ if industry N is the treatment industry, and 0 otherwise, $\lambda_t = 1$ event if event time $= t$, and I include the first year 1993 as reference year and thus omitted from the graph. I cluster standard errors at N industry level since the treatment category is at N industry level.

4.2 Spillover Effects through Production Networks

Next, I turn to the second stage estimation on spillover effects. Recall that Upstream 1 and Downstream 1 firms tend to be much larger than the control group firms. This is because large firms tend to directly transact with many firms (e.g., Toyota). Since the remarkable size difference could imply a differential trend in the counterfactual, rather than relying upon the standard parallel trend assumption, I resort to the Doubly-robust difference-in-differences (DRDID) estimator from Sant'Anna and Zhao (2020). The estimator is an extended version of the semiparametric DID with inverse probability weighting (IPW) approach suggested by Abadie (2005). DRDID estimators are consistent if either (but not necessarily both) a propensity score or outcome regression working models are correctly specified.

DRDID estimators require two standard assumptions in the conditional DID estimators: the so-called conditional parallel trend assumption and overlap assumption. The former assumption essentially states that the parallel trend assumption holds conditional on covariates that drive a differential trend. In other words, the assumption allows for covariate-specific time trends. The latter assumption states that some fraction of the population is treated and that for every value of the covariates, there is a positive probability that the unit is not treated.

Since my sample is a panel data with multiple time periods, I use Callaway and

Sant'Anna (2021) version of DRDID that extends the two-period DRDID estimator from Sant'Anna and Zhao (2020) to the one with multiple time periods. The idea is simple; for the conventional event study plot, they suggest to conduct two-period DRDID for two subsequent periods. In my case, for an event study plot, they suggest to implement DRDID on 1993 vs. 1994, 1994 vs. 1995, and so forth, for the pre-treatment period, and then 1997 vs. 1998, 1997 vs. 1999, and so forth, for the post-treatment period. For the estimate, I follow their aggregation scheme, which is essentially averaging out the pre- and post-treatment estimates. I conduct the DRDID for every network distance. For instance when I conduct DRDID on Upstream 1 firms, I restrict the sample to Upstream 1 firms and the control group first, and then I implement DRDID. The details of the specification from Sant'Anna and Zhao (2020) and Callaway and Sant'Anna (2021) are explained in the Appendix D.[14]

To take out the direct effects on the connected firms that may vary across different network distances, I control for the (smallest-level) industry-by-year-by-prefecture fixed effects. Note that since some industries perfectly predict network distance in my sample, the inclusion of these fixed effects as covarites in the selection equation will violate the overlap assumption. Then, I use an outcome residualized by these fixed effects.[15]

To compute the propensity score, I use the following covariates: the outcome of the previous period, the natural logarithm of capital stock, that of employment, that of the shortest physical distance to any of the treated firms being connected, and the average physical distance to the treated firms. The first three covariates are measures of firm sizes, and the other two are used since geographic distance is known to affect a firm

[14] For the actual implementation, I follow Rios-Avila et al. (2023) and use *csdid* Stata package written by the authors including Callaway and Sant'Anna.

[15] While partialling out fixed effects only from the left-hand side of a reduced-form equation is a common practice in empirical literature, as pointed out by Gormley and Matsa (2014), not partialling out the fixed effects from the right-hand side could result in an omitted variable bias in the estimates caused by a potential correlation between the treatment and fixed effects that are not partialled out. In my case, a potential omitted variable comes from a correlation bewtween the network distance and industry-specific shocks that are uncorrelated with firm-level size and physical distance. As in the direct effects, I mitigate this concern by graphically showing the conditional parallel-trend in the pre-treatment periods.

linkage. Furthermore, I include the polynomials of these covariates upto the third order since overfitting costs little but the fourth order turns out to be not only computationally difficult but also fails to converge. Note that the results are robust when I only include polynomials up to the second order.

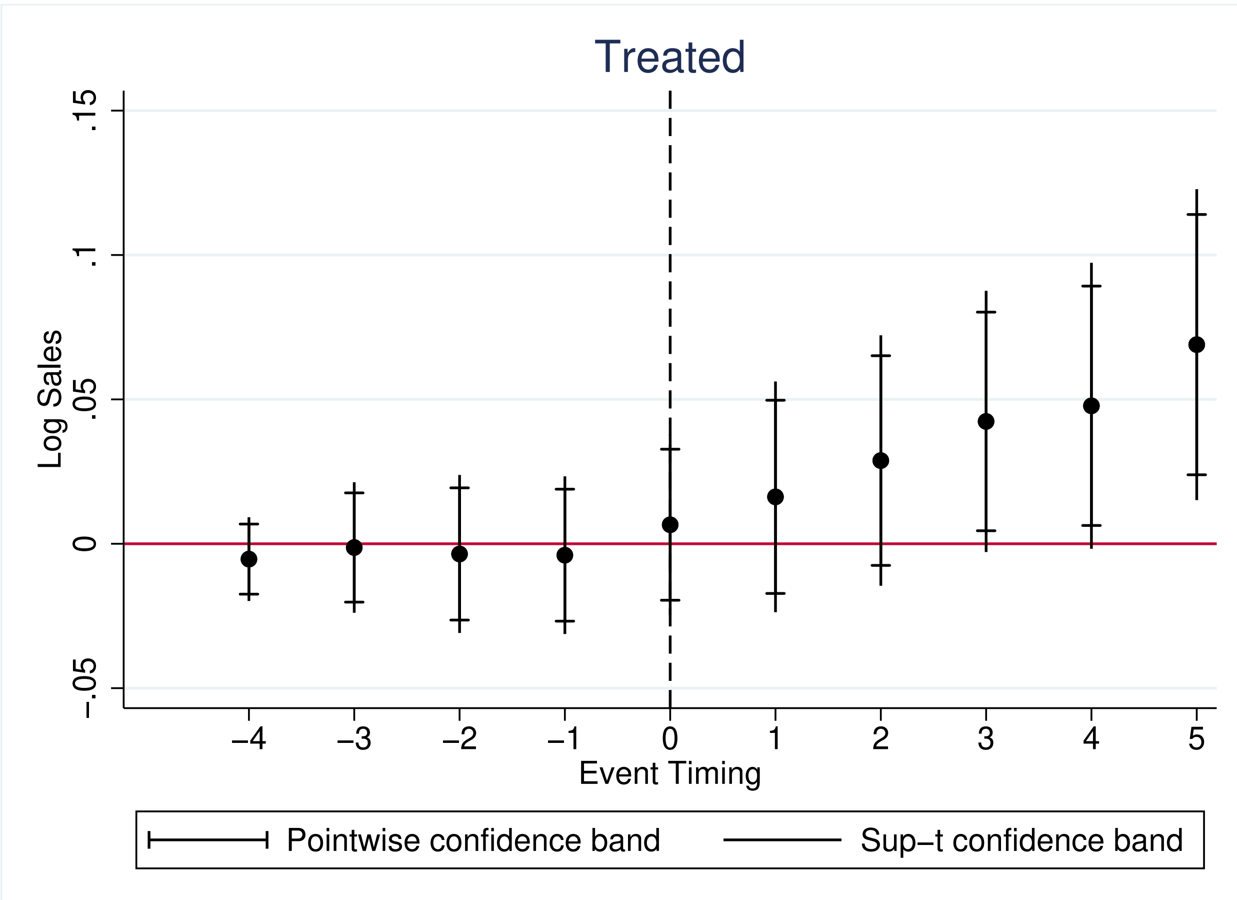
When I plot the dynamics of the effects, I follow the best practices of event study plots suggested by Freyaldenhoven et al. (2021) and use the (simultaneous) uniform confidence bands for the standard DID and similar uniform bands proposed by Callaway and Sant'Anna (2021) for DRDID estimators to graphically show the parallel trend in a conservative way. The motivation for the uniform confidence intervals (CIs) is to circumvent the multiple-hypothesis testing implicitly conducted with multiple time periods. For DRDID estimators, I follow Callaway and Sant'Anna (2021) and only display the uniform confidence bands.

5 Results

In Figure 3, x-axis indicates time period where 0 is equal to FY1998 that corresponds to the year the policy took effect, -1 means FY 1997, and so forth. It displays the coefficients and 95% confidence interval from the regression of equation (2) and appears to support the parallel trend. I exclude the reference year 1993 from the display in the graph.

Table 3 demonstrates the direct effects estimated by equation (1). The first column corresponds to the baseline model of equation (1), the second column includes two-digit JSIC industry-by-prefecture-by-year fixed effects, the third column restricts the control group sample to SMEs whose capital stock level is smaller than 100 million yen roughly equal to one million dollars, the fourth column restricts the treatment group to firms whose aggregated net-tax profits in the post period are negative, and the fifth column uses the number of employees as the outcome variable. All the outcome variables are in a logarithmic scale. One can see the statistically significant and economically meaningful direct

Figure 3: Pre-trend Test of Direct Effects



Notes: x-axis indicates time period where 0 is equal to FY1998 that corresponds to the year the policy took effect, -1 means FY 1997, and so forth. Coefficients with 95% confidence intervals and uniform sup-t confidence intervals from equation (2). The reference year 1993 is excluded to display from the graph.

Table 3: First-stage Direct Effects of Investment Stimuli

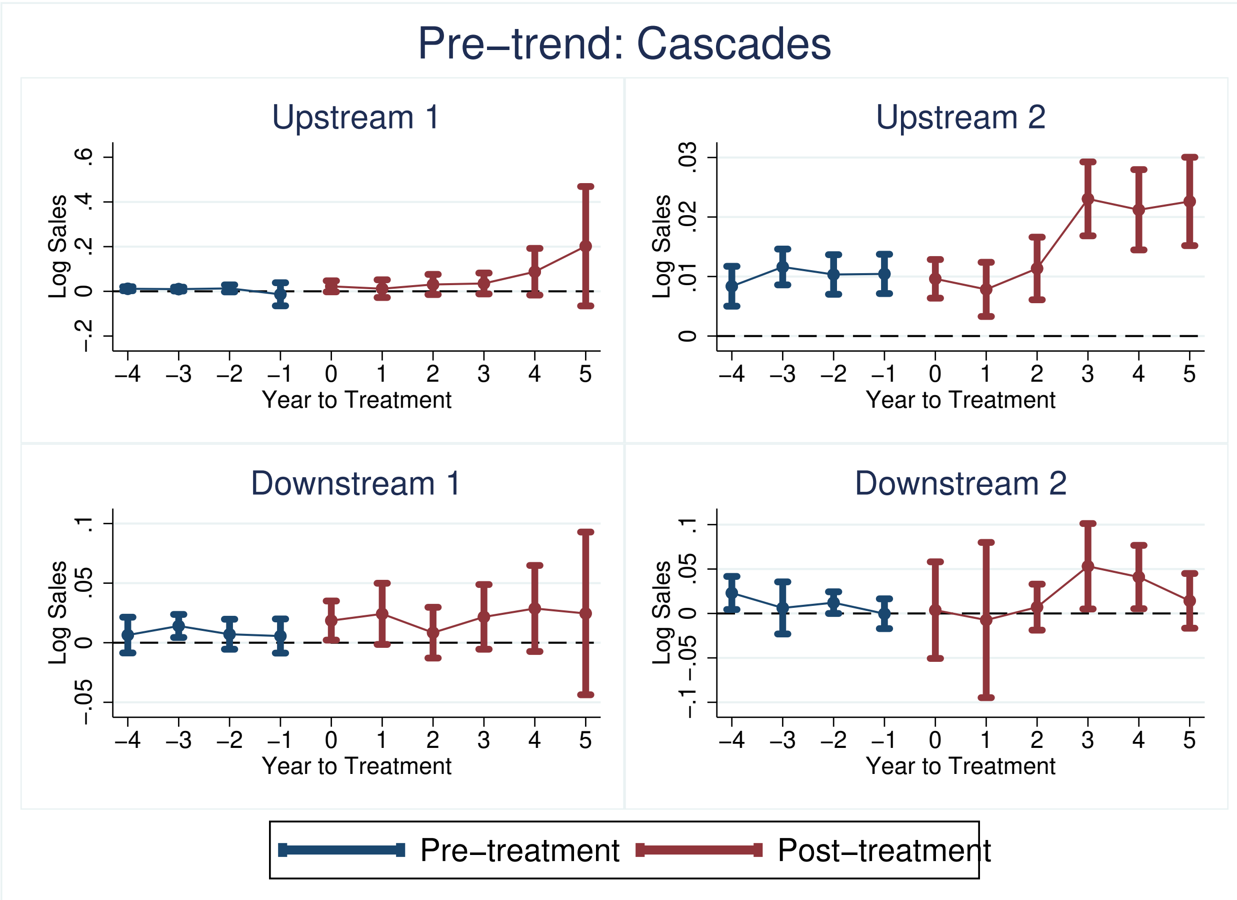
	(1) Sales	(2) Sales	(3) Sales	(4) Sales	(5) Employment
DD	0.038	0.045	0.054	-0.081	-0.015
	(0.014)	(0.014)	(0.014)	(0.019)	(0.014)
N	1556984	1553716	1542595	1020291	1553716
2-digit JSIC FE x Year FE	No	Yes	Yes	Yes	Yes
SME Control	No	No	Yes	No	No
Unprofitable Firm	No	No	No	Yes	No

Notes: Standard errors in parentheses are clustered at the industry level. The first column corresponds to baseline estimates, the second column includes two-digit JSIC industry-by-prefecture-by-year fixed effects, the third column restricts the control group to small and medium firms whose capital stock level is smaller than 100 million yen roughly equal to one million dollars, the fourth column restricts the treatment group to firms whose aggregated net-tax profits in the post period are negative, and the fifth column uses the number of employees as the outcome variable. All the outcome variables are in a logarithmic scale.

effects at 4% increase in sales after the policy under the baseline model. Furthermore, the results are robust to the inclusion of 2-digit industry-by-year-by-prefecture fixed effects, mitigating the concern of industry-trend effects. Since only SMEs were eligible for the bonus depreciation, I conduct a robustness check by restricting the control group sample to SMEs. The results are similar, while I use the full control group sample as my baseline to compare the estimates with the spillover effects. The fourth column is another robustness check that uses unprofitable firms for the treatment group, which will be discussed in more details later. The last column is the same as the second column except that the outcome variable is the number of employees. The estimate is insignificant and shows no effect on the employment of the affected firms, an interesting contrast to the positive effects found in the previous literature. While I will discuss on this no employment effect in more details later, the lack of effects also implies that the revenue increase is driven by an increase in the other input in production, which is capital.

Figure 4 displays the coefficients and analytical 95% confidence interval from DRDID for each distance group (Upstream 1, Upstream 2, Downstream 1, Downstream 2). For each distance group, I restrict sample to that distance group and the control group, and then I estimate DRDID with this restricted sample. The graph appears to roughly sup-

Figure 4: Pre-trend Test for Second-stage Spillover Effects



Notes: x-axis indicates time period where 0 is equal to FY1998 that corresponds to the year the policy took effect, -1 means FY 1997, and so forth. Coefficients with 90% uniform confidence intervals from the DRDID estimation. The reference year 1993 is excluded to display from the graph.

port the parallel trend for each group except Upstream 2. Although the increasing trend accelerates toward the end of post-treatment periods like Upstream 1, the presence of pre-trend makes the coefficients invalid for Upstream 2. Therefore, I will not make a conclusive remark on the spillover effects over the indirect suppliers.

Table 4 shows the regression results for DRDID estimation. The first column corresponds to the results of DRDID, and the second column shows the DRDID results with SME sample. The first row shows that there is spillover effects on the direct suppliers that is even larger than the direct effects with the full sample. Note that while the spillover effect on the direct suppliers are statistically significant at the five percent significance level when restricted to the SME sample, the effect is significant only at the 10 percent level (p-

Table 4: Second-stage Spillover Effects on Sales

	(1) Sales	(2) Sales
Upstream 1 × Post	0.065	0.028
	(0.037)	(0.009)
Upstream 2 × Post	0.016	0.017
	(0.002)	(0.003)
Downstream 1 × Post	0.021	0.008
	(0.014)	(0.017)
Downstream 2 × Post	0.019	0.009
	(0.015)	(0.016)
SME only	No	Yes

Notes: Standard errors in parentheses are clustered at the (finest-digit) industry level. The first column corresponds to baseline DRDID estimates, and the second column is the same estimation except that the sample is restricted to SMEs whose capital stock level is smaller than 100 million yen roughly equal to one million dollars.

value = 0.079 for the two-sided test and 0.039 for the one-sided test) with the full sample. Thus, it is weak evidence. On the other hand, we do not see statistically significant effects on the downstream firms. This is a stark contrast to the existing scholarship which finds downstream spillovers.

5.1 Additional Robustness Checks

5.1.1 Different Cutoffs of Treatment Category

As mentioned above, Garrett et al. (2020), I use the 30th percentile of the z_j^0 distribution as the cutoff for the discretized treatment measure for my main results. I conduct robustness checks with the 25th and 40th percentiles of the z_j^0 distribution and confirmed no change in results, as shown in Appedix C.

5.1.2 Unprofitable Firm Sample

Since bonus depreciation only affects firms with positive taxable income, I check robustness by estimating the direct effects using those firms which have negative net-tax profits in the post-treatment period on average. As the fourth column in Table 3 shows, these firms do not show positive effects of bonus depreciation.[16]

6 Discussions and Limitations

6.1 Heterogeneity by Network Distance

As demonstrated above, there is a considerable heterogenetiy in the spillover effects by the network distance of connected firms. With the sample restricted to SMEs, we found statistically significant effects on the direct suppliers of the treated firms that are comparable to the direct effects, and when we did not restrict the sample, we found weakly statistically significant effects that are even larger than the direct effects. This means that some large firms enjoy larger spillover effects than SMEs, implying the potential influence of markups.[17] However, I humbly note that the presence of such distortions is unfortunately hard to pin down among firms other than the aforementioned sales size differences, due to the limitation of our datasets.

Furthermore, the larger estimate needs to be interpreted with caution due to the nature of the binary measure for the treatment. While the direct policy effects are taken out fully from the indirect effects estimation by controlling for the industry-by-year fixed effects, the binary measure for the treatment in the estimation of direct effects could cause a downward bias since the control group also experiences some direct effects as well. Although the difference between the treatment and control group can detect the presence of direct effects, it would not capture a precise magnitude.

In contrast to the upstream spillovers, we do not see statistically significant effects on downstream firms, whether they are large or small firms. This is a stark contrast to the

[16]One may think that (positive) profit levels in the pre-treatment period can be used for another robustness check of the direct effects. I do not conduct this robustness check since those small firms which evade corporate taxes by reporting a small negative taxable income can change their behavior in response to the tax policies in 1998.

[17]See, e.g., Baqaee and Farhi (2020) for models that show that the presence of distortions such as markups can alter the productivity shocks' propagation patterns.

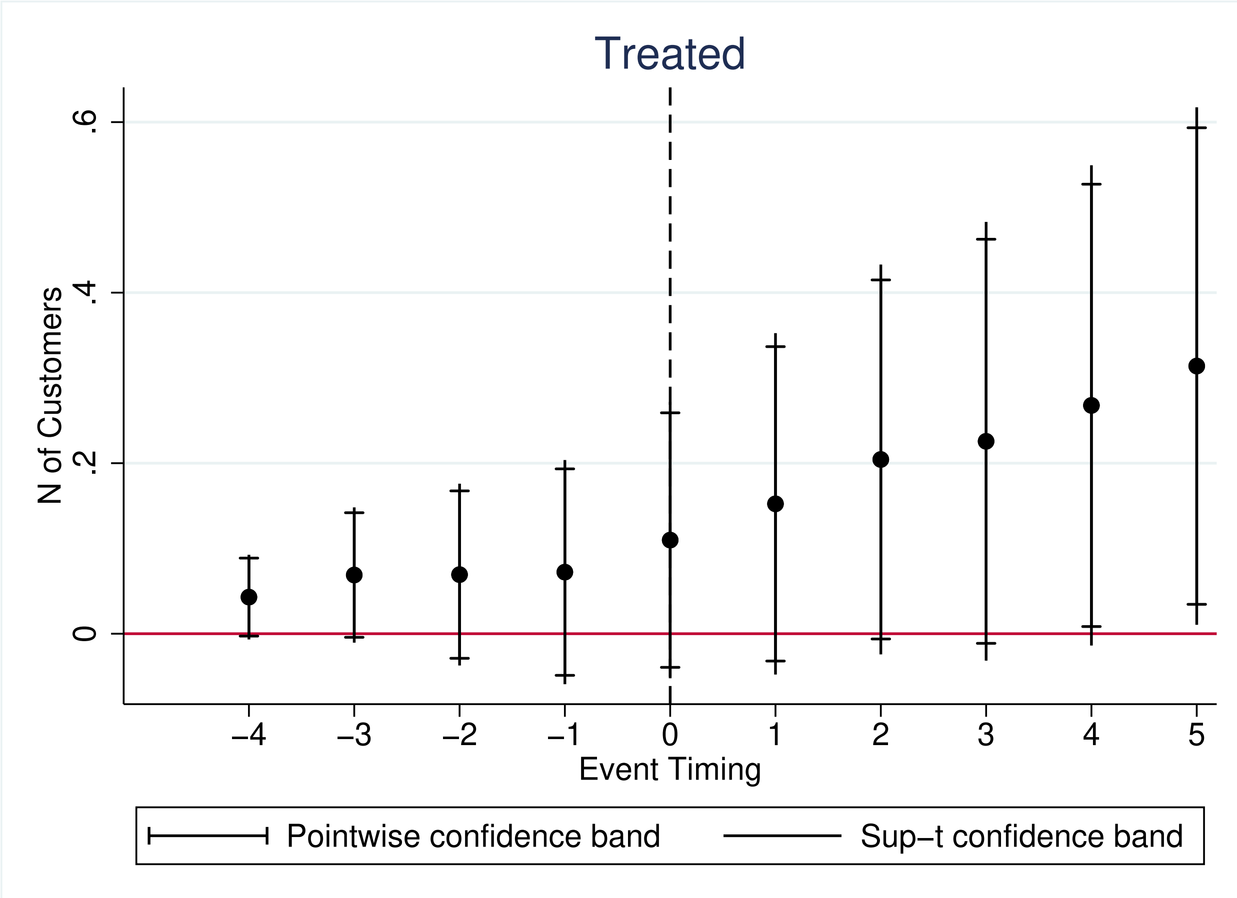
existing research on spillover effects in production networks. Carvalho et al. (2021) find that the negative shock of the 2011 Great East Japan Earthquake that hit a certain region of Japan resulted in negative spillover effects on upstream and downstream firms almost at equal magnitudes with slightly stronger effects on downstream firms, and the effects fade away at the farther network distance.

Given that the treated firms increased sales, either the prices or quantity of their products sold must have increased. The treated firms could have possibly increased the price of their products since, as theoretically analyzed by Grassi et al. (2017), the positive productivity shock might have increased their markups. The price data is not available to us, so we cannot unfortunately investigate this channel.

On the other hand, if the treated firms increased the supply of their products, which is expected to drive down the price of their products, the customer firms must have increased their production as well. And yet, we do not see a statistically significant sales increase in the customer firms. Note that the treated firms include industries such as utilities, pipeline transport, railroad, accommodation, and food manufacturing. These industries tend to have two commonalities. One is that they sell to final consumers. Thus, the treated firms increased sales perhaps by selling more to the final consumers. The other commonality is that these industries tend to sell highly specialized goods. This might result in relatively inelastic demand in their markets in which they do not need to lower the prices of their goods much to increase their sales. I cannot test this route since the data does not have information on the price and volume of each transaction.

Furthermore, the treated firms could have searched for new customer firms to avoid lowering the prices. If the treated firms produce and sell more to the same customer firms, then they need to lower the prices of their products. Firms tend to avoid lowering prices since they do not need to provide a reason when they lower prices but they do need to explain hard when they increase prices. To sell more without decreasing the price of their products, they can search for new customers as long as their products are sufficiently

Figure 5: Graphical Results on the Direct Effects on the Number of Direct Customers



Notes: x-axis indicates time period where 0 is equal to FY1998 that corresponds to the year the policy took effect, -1 means FY 1997, and so forth. Coefficients with 95% confidence intervals and uniform sup-t confidence intervals from equation (2). The reference year 1993 is excluded to display from the graph.

unique. Since the industries of the treated firms are heterogeneous, I expect them to produce heterogeneous goods in general. Figure 5 demonstrates such firm behavior. The number of direct customers of the treated firms started taking off after the policy year compared to the control group firms, providing suggestive evidence for the conjecture. [18]

[18] On the other hand, there is no good theoretical justification to conduct both DID and DRDID on the large firms of the treated industries. Given the large size difference, the standard parallel trend assumption may not hold. Meanwhile, the conditional parallel trend assumption conditional on firm size needs the overlap condition to hold. However, the fact that these large firms are supposed to have more than 100 million common stock would violate this condition between these large firms and the control group if we use firm size as covariates. Therefore, it is difficult for me to provide causal arguments over a potential sales decrease in these large firms due to the share crowding through DID or DRDID.

6.2 Concurrent Tax Policies

6.2.1 Sales Tax Increase in 1997

There were some other tax policies implemented around 1998. First, for fiscal consolidation, Japan increased the sales tax rate from 3% to 5% in April 1997. This must have affected the Japanese economy, raising some identification concern if this differentially affected the treated firms. The dynamic effects in Figure 3 demonstrate that there was no differential effect in 1997, mitigating this concern.

6.2.2 First-year Simplified Method

The government announced on January 9, 1998 that the government would abolish in April 1998 the First-year Simplified Method (*Shonendo Kanbenhou*) that allowed the firms to depreciate half of the depreciation amount of the first year for investment goods (except structures such as buildings) purchased in whichever month of the fiscal year. For example, suppose a firm whose fiscal month is April buys a computer in March 1993, and suppose that the firm is allowed to depreciate 20% of the purchased amount for the first year. Then, the simplified method allowed the firm to immediately depreciate half of 20%, which is 10%, for the 1993 fiscal year tax return, and depreciation the other half throughout the rest of the year. Since April 1998, all firms must distribute the depreciation amount on a monthly prorated basis and divide the depreciation amount by the proportion of months left for depreciation. For instance, if a firm whose fiscal month is April buys a computer in March 1993, the firm is allowed to depreciate only 1 / 12 of 20% of the purchased amount. The abolishment announcement (and the anticipation effects from policy discussions before the announcement) might have created a "last-minute" demand and incentivized those treated firms whose fiscal month ends in or before April to purchase eligible equipment relatively more than the control group firms whose fiscal month ends in or before April. Notice that this change is a one-time shock and does not confound the effects in the later years which I found.

6.2.3 Corporate Tax Decrease

As mentioned above, the effective corporate tax rates were reduced in both 1998 and 1999, in response to the Asian Currency Crisis. One may wonder if this change affected the results in this paper. Using the Domestic Production Activities Deduction (DPAD), Ohrn (2018) finds that a decrease in the effective corporate tax rate increases investment. The ZM bonus depreciation intensity measure is constructed based upon depreciation duration schedule differences not based upon corporate tax rates and other indicators, which mitigates the identification threat to some extent. Meanwhile, I cannot eliminate the possibility that those industries that invest in relatively long (or short) duration equipment benefits more from the corporate tax cut. However, the effects of a bonus depreciation from Zwick and Mahon (2017) introduced in the period concurrent with the effective corporate tax cut through DPAD is consistent with the effects introduced in the period without DPAD. Assuming that Japanese and U.S. industry structures are similar, their findings further mitigate the concern.

6.2.4 No Employment Effect and Labor Laws in Japan

While I found an increase in sales, I found no effect of investment stimuli on employment. This is a stark contrast to the findings of the previous scholarship in the U.S. The zero effect is surprising at a glance but is consistent with high fixed costs of hiring a new employee in Japan due to strict labor laws. Japanese employment contracts are remarkably different from those in other countries in that Japanese firms hire new employees not for specific jobs or tasks but for general purposes as "official employees" or so-called *seishain*. This unique feature of Japanese contracts make it much harder for firms to lay off their employees even when firms are in downturn and need to downsize specific departments or projects since the Japanese contracts make it possible for these employees to be transferred from one department to another within the firms.[19] This implies that Japanese firms pay large fixed costs to hire an additional employee. Therefore, even if labor and equipment in the firms' production functions are imperfect substitutes, the decreased price of equipment by the bonus depreciation must have been insufficient for firms to hire more employees.

6.3 Technical Difficulty of Some Robustness Tests

6.3.1 Block Permutation Tests

As conducted (in slightly different forms) by Chetty et al. (2009), Ohrn (2018), and Zidar (2019), block permutation tests that randomly assign placebo implementation years and randomly pick placebo-treated industries appear to be desirable for this study as well. Ohrn (2018) conducted a series of 2,000 block permutation tests, which will take roughly 5,000 hours on the server of TDB. In addition to the time challenge, given that the TDB on-site data center to which contracted researchers have access conducts server maintenance on a regular basis, conducting block permutations of an even smaller number would be unfortunately technically challenging.

6.3.2 Placebo Production Network

Another placebo test that is desirable but technically challenging is to conduct the same analysis with a randomly generated production network, as done by Carvalho et al. (2021). Carvalho et al. (2021) start with the actual production network constructed using their data, draw a random production network uniformly at random while preserving the identity and the number of customers of all firms, and use the resulting network to redefine all firms' upstream and downstream network distances to the treated firms. This falsification exercise is unfortunately infeasible in my setting. The vast majority of firms

[19]See, e.g., https://shuchi.php.co.jp/the21/detail/8467.

in my sample lie within two network distances, and therefore many firms still lie in the plausibly affected network distances.

7 Conclusion

This paper estimates the indirect effects of tax incentives for investment through production networks, exploiting a quasi-experimental event of an investment stimulus and unique proprietary data of supply chains in Japan. I confirm the direct effects of such policies that are consistent with the existing literature. I further provide novel results that corporate tax policies cascade through production networks. These results are useful findings for policymakers who are evaluating the effectiveness of investment stimulative tax policies in the entire economy. Additionally, the results of this paper imply that the justification used for a policy to be restricted to SMEs may not be valid due to the spillovers through the production networks.

References

ABADIE, A. (2005): "Semiparametric Difference-in-Differences Estimators," *The Review of Economic Studies*, 72, 1–19.

ACEMOGLU, D., U. AKCIGIT, AND W. KERR (2016): "Networks and the macroeconomy: An empirical exploration," *Nber macroeconomics annual*, 30, 273–335.

AUER, R. A., A. A. LEVCHENKO, AND P. SAURÉ (2019): "International inflation spillovers through input linkages," *Review of Economics and Statistics*, 101, 507–521.

BAQAEE, D. R. AND E. FARHI (2020): "Productivity and misallocation in general equilibrium," *The Quarterly Journal of Economics*, 135, 105–163.

BARROT, J.-N. AND J. SAUVAGNAT (2016): "Input specificity and the propagation of idiosyncratic shocks in production networks," *The Quarterly Journal of Economics*, 131, 1543–1592.

BERNARD, A. B., A. MOXNES, AND Y. U. SAITO (2019): "Production networks, geography, and firm performance," *Journal of Political Economy*, 127, 639–688.

BERTRAND, M., E. DUFLO, AND S. MULLAINATHAN (2004): "How much should we trust differences-in-differences estimates?" *The Quarterly journal of economics*, 119, 249–275.

BOEHM, C. E., A. FLAAEN, AND N. PANDALAI-NAYAR (2019): "Input linkages and the transmission of shocks: Firm-level evidence from the 2011 Tōhoku earthquake," *Review of Economics and Statistics*, 101, 60–75.

CALLAWAY, B. AND P. H. SANT'ANNA (2021): "Difference-in-Differences with multiple time periods," *Journal of Econometrics*, 225, 200–230.

CARVALHO, V. M., M. DRACA, ET AL. (2018): "Cascading innovation," *Work. Pap., Univ. Cambridge, UK*.

CARVALHO, V. M., M. NIREI, Y. U. SAITO, AND A. TAHBAZ-SALEHI (2021): "Supply chain disruptions: Evidence from the great east japan earthquake," *The Quarterly Journal of Economics*, 136, 1255–1321.

CHETTY, R., A. LOONEY, AND K. KROFT (2009): "Salience and taxation: Theory and evidence," *American economic review*, 99, 1145–1177.

CUMMINS, J. G., K. A. HASSETT, R. G. HUBBARD, R. E. HALL, AND R. J. CABALLERO (1994): "A reconsideration of investment behavior using tax reforms as natural experiments," *Brookings papers on economic activity*, 1994, 1–74.

CURTIS, E. M., D. G. GARRETT, E. C. OHRN, K. A. ROBERTS, AND J. C. S. SERRATO (2021): "Capital investment and labor demand," Tech. rep., National Bureau of Economic Research.

DEMIR, B., B. JAVORCIK, T. K. MICHALSKI, AND E. ORS (2022): "Financial constraints and propagation of shocks in production networks," *Review of Economics and Statistics*, 1–46.

DOBKIN, C., A. FINKELSTEIN, R. KLUENDER, AND M. J. NOTOWIDIGDO (2018): "The economic consequences of hospital admissions," *American Economic Review*, 108, 308–352.

EDGERTON, J. (2010): "Investment incentives and corporate tax asymmetries," *Journal of Public Economics*, 94, 936–952.

FAN, Z. AND Y. LIU (2020): "Tax compliance and investment incentives: firm responses to accelerated depreciation in China," *Journal of Economic Behavior & Organization*, 176, 1–17.

FREYALDENHOVEN, S., C. HANSEN, J. P. PÉREZ, AND J. M. SHAPIRO (2021): "Visualization, identification, and estimation in the linear panel event-study design," Tech. rep., National Bureau of Economic Research.

GARRETT, D. G., E. OHRN, AND J. C. SUÁREZ SERRATO (2020): "Tax policy and local labor market behavior," *American Economic Review: Insights*, 2, 83–100.

GORMLEY, T. A. AND D. A. MATSA (2014): "Common errors: How to (and not to) control for unobserved heterogeneity," *The Review of Financial Studies*, 27, 617–661.

GRASSI, B. ET AL. (2017): "Io in io: Size, industrial organization, and the input-output network make a firm structurally important," *Work. Pap., Bocconi Univ., Milan, Italy*.

GUCERI, I. AND L. LIU (2019): "Effectiveness of Fiscal Incentives for R&D: Quasi-experimental Evidence," *American Economic Journal: Economic Policy*, 11, 266–291.

HALL, R. E. AND D. W. JORGENSON (1967): "Tax policy and investment behavior," *The American Economic Review*, 57, 391–414.

HECKMAN, J. J., H. ICHIMURA, AND P. E. TODD (1997): "Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme," *The review of economic studies*, 64, 605–654.

HOUSE, C. L. AND M. D. SHAPIRO (2008): "Temporary investment tax incentives: Theory with evidence from bonus depreciation," *American Economic Review*, 98, 737–68.

KHAN, S. AND E. TAMER (2010): "Irregular identification, support conditions, and inverse weight estimation," *Econometrica*, 78, 2021–2042.

KOMORI, R. (2003): "Recent Changes in Japanese Depreciation Systems And Japan-US Comparison," *Keizaigaku Ronsou*, 54, 1–24.

LIU, E. (2019): "Industrial Policies in Production Networks*," *The Quarterly Journal of Economics*, 134, 1883–1948.

OHRN, E. (2018): "The Effect of Corporate Taxation on Investment and Financial Policy: Evidence from the DPAD," *American Economic Journal: Economic Policy*, 10, 272–301.

——— (2019): "The effect of tax incentives on US manufacturing: Evidence from state accelerated depreciation policies," *Journal of Public Economics*, 180, 104084.

OZDAGLI, A. AND M. WEBER (2017): "Monetary policy through production networks: Evidence from the stock market," Tech. rep., National Bureau of Economic Research.

RIOS-AVILA, F., P. SANT'ANNA, AND B. CALLAWAY (2023): "CSDID: Stata module for the estimation of Difference-in-Difference models with multiple time periods," Statistical software components s458976, Boston College Department of Economics.

ROTH, J. AND P. H. C. SANT'ANNA (2022): "When Is Parallel Trends Sensitive to Functional Form?" .

SANT'ANNA, P. H. AND J. ZHAO (2020): "Doubly robust difference-in-differences estimators," *Journal of Econometrics*, 219, 101–122.

TAKAFUMI, K., S. TAKAFUMI, ET AL. (2022): "Supply Chain Dynamics and Resilience of the Economy during a Crisis," Discussion papers 22070, Research Institute of Economy, Trade and Industry (RIETI).

TUZEL, S. AND M. B. ZHANG (2021): "Economic Stimulus at the Expense of Routine-Task Jobs," *The Journal of Finance*, 76, 3347–3399.

ZIDAR, O. (2019): "Tax cuts for whom? Heterogeneous effects of income tax changes on growth and employment," *Journal of Political Economy*, 127, 1437–1472.

ZWICK, E. AND J. MAHON (2017): "Tax Policy and Heterogeneous Investment Behavior," *American Economic Review*, 107, 217–248.

Appendices

A Policy Background And Other Relevant Concurrent Tax Policies

In this section, I note the details of the tax policy of interest and other major tax policies and changes that were implemented around 1998.[20] First, for fiscal consolidation, Japan increased the sales tax rate from 3% to 5% in April 1997. The Asian Currency Crisis followed this, and Japan started experiencing a significant economic downturn in July 1997. Many banks went bankrupt toward the end of 1997, and the government allowed some of these banks to bail out. To stimulate the economy, the government announced on January 9, 1998 that the government would introduce special income tax allowances, decrease the corporate tax rate as mentioned in the main text while increasing the corporate tax base to maintain some level of fiscal health, and make several changes in the system of depreciation on investment goods.

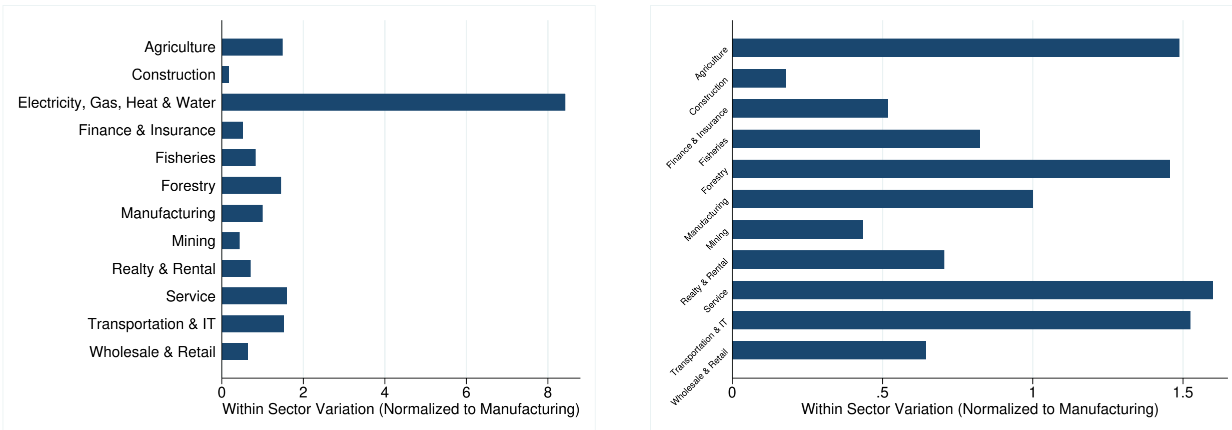
[20] The references come from the government reports: https://www.mof.go.jp/pri/publication/policy_history/series/h1-12/4_1_11.pdf, https://dl.ndl.go.jp/view/download/digidepo_3515892_po_553f.pdf?contentNo=6&alternativeNo=, and https://www.cao.go.jp/zei-cho/history/1996-2009/etc/1997/zeicho1.html.

B Treatment Intensity Variation within Sector

Figure B1: Treatment Intensity Variation in Each Sector

(a) Within Single-digit JSIC Variation in Duration

(b) Within Single-digit JSIC Variation in Duration without Utility Sector



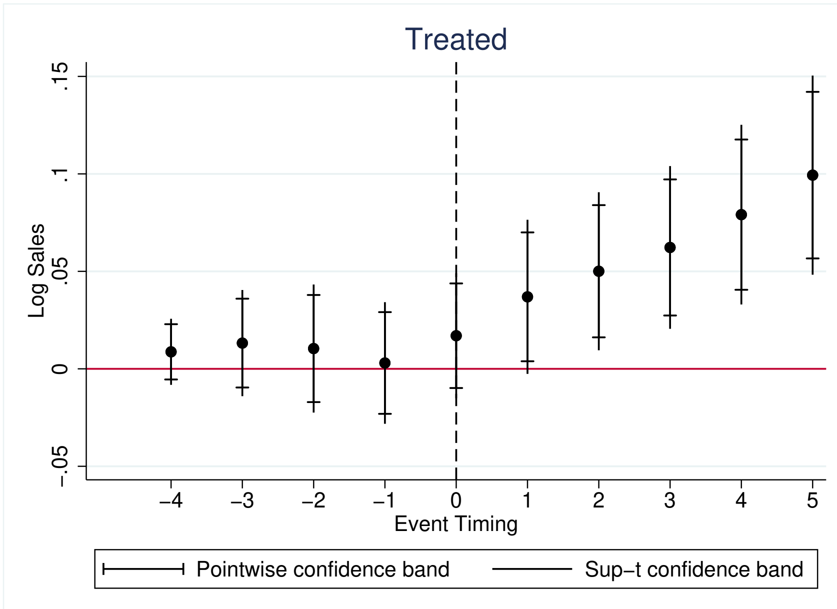
C Additional Results and Robustness Checks

C.1 Different Cutoffs for Treatment

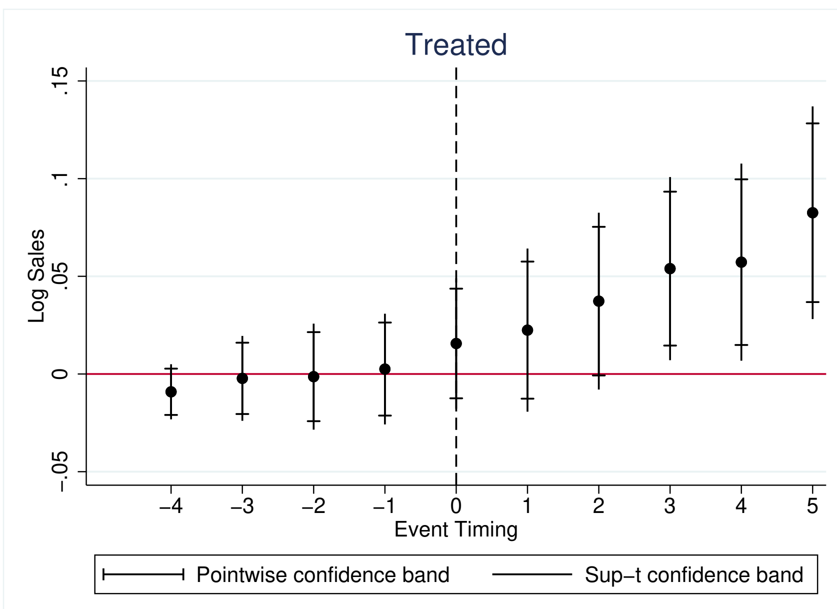
In the main text, I use the 33rd percentile of z_N measure adopted from ZM US industries to Japanese industries. I chose this cutoff for the balance of sample sizes across different network distances including the treated group. In this section, I show that the results in the main text are robust to the choice of this cutoff by demonstrating the same event study plots with the 25th and 40th percentiles as cutoffs for the treatment.

Figure C1: Pre-trend Test for Direct Effects with Different Cutoffs

(a) 25th percentile

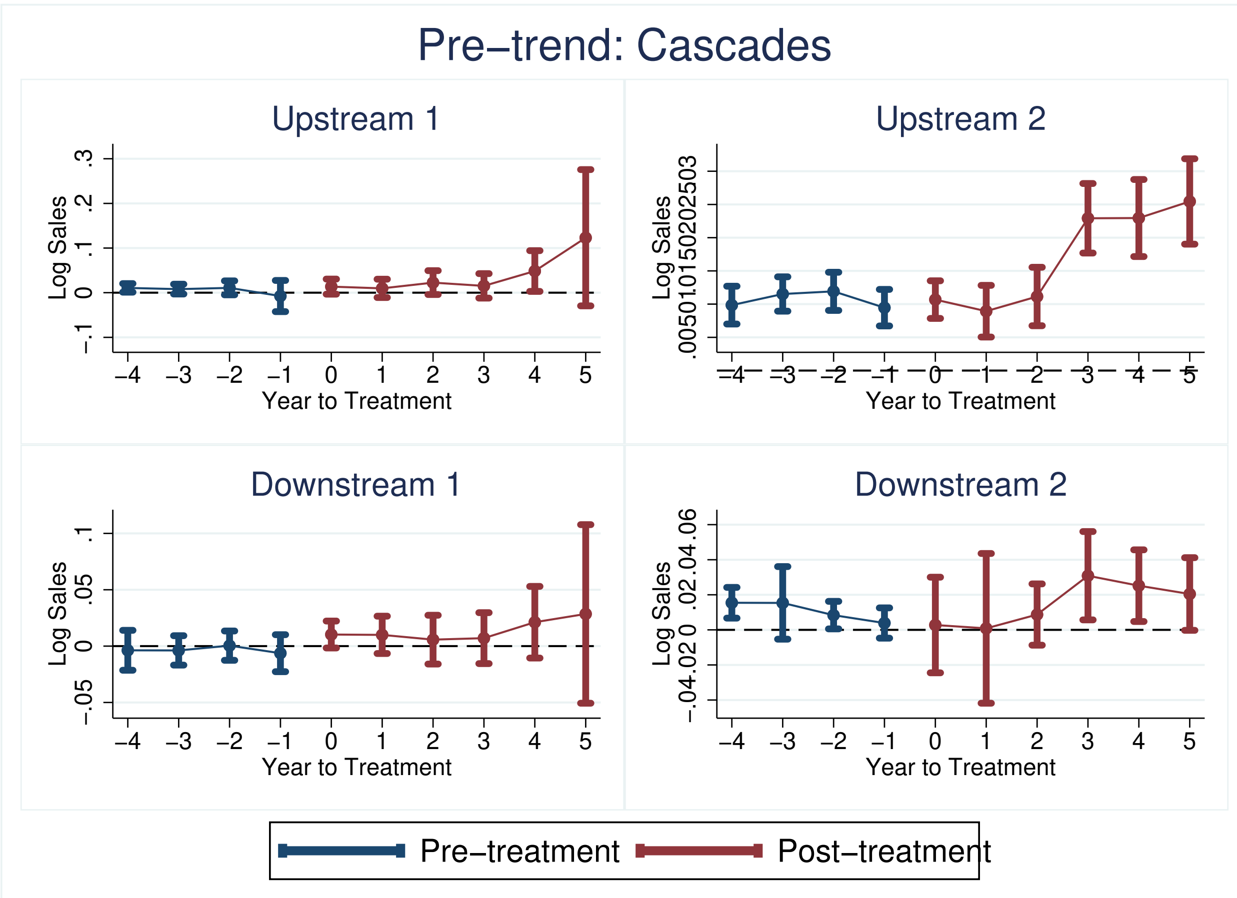


(b) 40th percentile



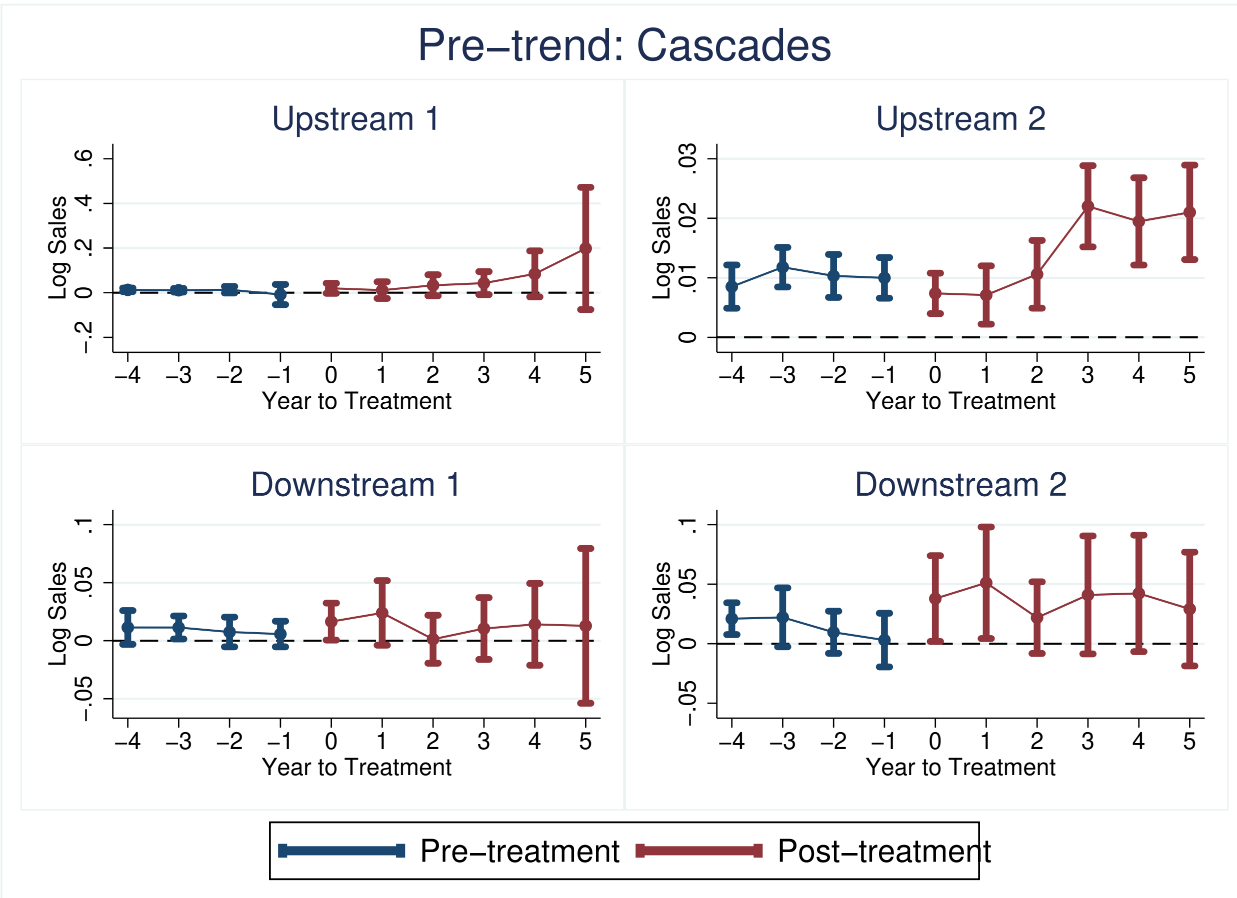
Notes: x-axis indicates time period where 0 is equal to FY1998 that corresponds to the year the policy took effect, -1 means FY 1997, and so forth. Coefficients with 95% point and sup-t uniform confidence intervals from equation (2). The reference year 1993 is excluded to display from the graph.

Figure C2: Pre-trend Test for Spillover Effects with 25th Percentile Cutoff



Notes: x-axis indicates time period where 0 is equal to FY1998 that corresponds to the year the policy took effect, -1 means FY 1997, and so forth. Coefficients with 90% uniform confidence intervals from the DRDID estimation. The reference year 1993 is excluded to display from the graph.

Figure C3: Pre-trend Test for Spillover Effects with 40th Percentile Cutoff



Notes: x-axis indicates time period where 0 is equal to FY1998 that corresponds to the year the policy took effect, -1 means FY 1997, and so forth. Coefficients with 90% uniform confidence intervals from the DRDID estimation. The reference year 1993 is excluded to display from the graph.

D Doubly-robust Difference-in-differences from Sant'Anna and Zhao (2020) and Callaway and Sant'Anna (2021)

In this section, I shall explain the DRDID estimators from Sant'Anna and Zhao (2020), which is an extended version of the semiparametric DID with inverse probability weighting (IPW) approach suggested by Abadie (2005). In contrast to the semiparametric DID approach that requires the model for propensity score is correctly specified, the DRDID approach is valid if either the propensity score model is correct or the outcome estimation model is correct.

D.0.1 Set-up

To formally introduce the estimator and its assumptions, I will introduce some new notations. Readers who are familiar with the estimator and the paper can safely skip this section. I first focus on a canonical two-period DID setup and introduce the definition and assumptions of DRDID estimators following Sant'Anna and Zhao (2020). Next, I will the case with multiple periods following Callaway and Sant'Anna (2021) and introduce their suggested way to apply DRDID to the multiple-period case. For both cases, I focus on a case in which researchers have access to panel data sets in this paper.

Suppose we have two periods: $t = 0$ as a pre-treatment period, while $t = 1$ as a post-treatment period. Let Y_{it} be the outcome of interest for unit i at time t. We assume that one has access to outcome data at $t = 0$ and $t = 1$. Let $D_{i,t}$ be a binary variable equal to one if unit i is treated in period t and equal to zero otherwise. Since we focus on a canonical DID set-up first, $D_{i0} = 0$ for every i, which allows us to write $D_i = D_{i1}$. Using the standard potential outcome notation, let $Y_{it}(0)$ be the outcome of unit i at time t if i receives no treatment by time t and let $Y_{it}(1)$ be the outcome for the same unit if it receives treatment. Then, the realized outcome for unit i at time t is $Y_{it} = D_i Y_{it}(1) + (1 - D_i)\, Y_{it}(0)$. A vector of pre-treatment covariates is denoted by X_{it} or in a two-period case, just X_i. We assume that the first element of X_i is a constant. The following is the standard assumption in the literature.

Assumption 1. *The data $\{Y_{it}, D_{it}, X_{it}\}_{i=1}^{n}$ are independent and identically distributed (iid).*

Next, note that the parameter of interest is the average treatment effect on the treated (ATT):

$$\tau = \mathbb{E}\left[Y_{i1}(1) - Y_{i1}(0) \mid D_i = 1\right],$$

which can be written as

$$\tau = \mathbb{E}\left[Y_1(1) \mid D = 1\right] - \mathbb{E}\left[Y_1(0) \mid D = 1\right] = \mathbb{E}\left[Y_1 \mid D = 1\right] - \mathbb{E}\left[Y_1(0) \mid D = 1\right] \quad (3)$$

where I drop subscript i to simplify notation and follow this convention through this paper. Now, I am ready to introduce the remaining standard assumptions in conditional DID methods, the conditional parallel trend assumption (PTA) and overlap assumption:

Assumption 2. *(conditional PTA)* $\mathbb{E}\left[Y_1(0) - Y_0(0) \mid D = 1, X\right] = \mathbb{E}\left[Y_1(0) - Y_0(0) \mid D = 0, X\right]$ *almost surely (a.s.).*

Assumption 3. *(overlap) For some $\varepsilon > 0$, $\mathbb{P}(D = 1) > \varepsilon$ and $\mathbb{P}(D = 1 \mid X) \leq 1 - \varepsilon$ a.s.*

Assumption 2 states that in the absence of the treatment, the average conditional outcome of the treated and the control groups would have evolved in a parallel manner. The important difference from the standard PTA is that Assumption 2 permits covariate-specific time trends, although it rules out unit-specific trends. On the other hand, Assumption 3 states that at least a small portion of the population is treated while for every value of X, there is at least a small likelihood that a unit is not treated.

Under Assumptions 1-3, there are two main estimation procedures to estimate the ATT. One is the outcome regression (OR) approach such as done in Heckman et al. (1997) which relies on researchers' ability to model the outcome evolution. Given the assumptions, one can estimate the ATT with

$$\widehat{\tau}^{\text{reg}} = \bar{Y}_{1,1} - \left[\bar{Y}_{1,0} + n_{\text{treat}}^{-1} \sum_{i|D_i=1} \left(\widehat{\mu}_{0,1}\left(X_i\right) - \widehat{\mu}_{0,0}\left(X_i\right)\right)\right], \tag{4}$$

where $\bar{Y}_{d,t} = \sum_{i|D_i=d,T_i=t} Y_{it}/n_{d,t}$ is the sample average outcome among units in treatment group d and time t, and $\widehat{\mu}_{d,t}(x)$ is an estimator of the true, unknown $m_{d,t}(x) \equiv \mathbb{E}\left[Y_t \mid D = d, X = x\right]$.

The other approach is the inverse probability weighted (IPW) approach suggested by Abadie (2005). This approach avoids directly modeling the outcome evolution and exploits the fact that under Assumptions 1-3, the ATT can be written as

$$\tau = \frac{1}{\mathbb{E}[D]}\mathbb{E}\left[\frac{D - p(X)}{1 - p(X)}\left(Y_1 - Y_0\right)\right]. \tag{5}$$

Abadie (2005) proposes the following IPW estimator:

$$\widehat{\tau}^{ipw,p} = \frac{1}{\mathbb{E}_n[D]}\mathbb{E}_n\left[\frac{D - \widehat{\pi}(X)}{1 - \widehat{\pi}(X)}\left(Y_1 - Y_0\right)\right], \tag{6}$$

where $\hat{\pi}(x)$ is an estimator of the true, unknown $p(X)$. and for a generic random variable Z such that $\mathbb{E}_n[Z] = n^{-1}\sum_{i=1}^{n} Z_i$.

The DRDID estimand combines these two approaches to form doubly robust moments/estimands for the ATT. Let $\pi(X)$ be an arbitrary model for the true, unknown propensity score. When panel data are available, let $\Delta Y = Y_1 - Y_0$ and define $\mu_{d,\Delta}^{p}(X) \equiv \mu_{d,1}^{p}(X) - \mu_{d,0}^{p}(X), \mu_{d,t}^{p}(x)$ being a model for the true, unknown outcome regression $m_{d,t}^{p}(x) \equiv \mathbb{E}\left[Y_t \mid D = d, X = x\right], d, t = 0, 1$. Given these notations, the DRDID estimand is defined as

$$\tau^{dr,p} = \mathbb{E}\left[\left(w_1^p(D) - w_0^p(D, X; \pi)\right)\left(\Delta Y - \mu_{0,\Delta}^{p}(X)\right)\right], \tag{7}$$

where, for a generic g,

$$w_1^p(D) = \frac{D}{\mathbb{E}[D]}, \quad \text{and} \quad w_0^p(D, X; g) = \frac{g(X)(1-D)}{1-g(X)} / \mathbb{E}\left[\frac{g(X)(1-D)}{1-g(X)}\right]. \tag{8}$$

The generic DRDID estimators are in the following form:

$$\widehat{\tau}^{dr,p} = \mathbb{E}_n\left[\left(\widehat{w}_1^p(D) - \widehat{w}_0^p(D, X; \widehat{\gamma})\right)\left(\Delta Y - \mu_{0,\Delta}^p\left(X; \widehat{\beta}_{0,0}^p, \widehat{\beta}_{0,1}^p\right)\right)\right], \tag{9}$$

where

$$\widehat{w}_1^p(D) = \frac{D}{\mathbb{E}_n[D]}, \quad \text{and} \quad \widehat{w}_0^p(D, X; \gamma) = \frac{\pi(X; \gamma)(1-D)}{1-\pi(X; \gamma)} / \mathbb{E}_n\left[\frac{\pi(X; \gamma)(1-D)}{1-\pi(X; \gamma))}\right], \tag{10}$$

such that $\widehat{\gamma}$ is an estimator for the pseudo-true γ^*, $\widehat{\beta}_{0,t}^p$ is an estimator for pseudo-true $\beta_{0,t}^{*,p}$, $t = 0, 1$, and for a generic β_0 and β_1, $\mu_{0,\Delta}^p(\cdot; \beta_0, \beta_1) = \mu_{0,1}^p(\cdot; \beta_1) - \mu_{0,0}^p(\cdot; \beta_0)$.

Sant'Anna and Zhao (2020) provide some guidance on the choice of first-step estimators to further improve the generic DRDID estimators. They propose the so-called "improved" DRDID estimator for the ATT proposed by Sant'Anna and Zhao (2020) which focuses on the case where a researcher is comfortable with linear regression working models for the outcome of interest, a logistic working model for the propensity score, and with covariates X included in all the nuisance models in a symmetric manner. Then, we consider the case in which

$$\pi(X, \gamma) = \Lambda\left(X'\gamma\right) \equiv \frac{\exp\left(X'\gamma\right)}{1+\exp\left(X'\gamma\right)}, \text{ and } \mu_{0,\Delta}^p\left(X; \beta_{0,1}^p, \beta_{0,1}^p\right) = \mu_{0,\Delta}^{lin,p}\left(X; \beta_{0,\Delta}^p\right) \equiv X'\beta_{0,\Delta}^p. \tag{11}$$

Their improved DRDID estimator is provided by the following three-step estimator

$$\widehat{\tau}_{imp}^{dr,p} = \mathbb{E}_n\left[\left(\widehat{w}_1^p(D) - \widehat{w}_0^p\left(D, X; \widehat{\gamma}^{ipt}\right)\right)\left(\Delta Y - \mu_{0,\Delta}^{\text{lin,p}}\left(X; \widehat{\beta}_{0,\Delta}^{wls,p}\right)\right)\right], \tag{12}$$

where the first two-steps consist of computing

$$\hat{\gamma}^{ipt} = \arg\max_{\gamma \in \Gamma} \mathbb{E}_n \left[DX'\gamma - (1-D)\exp\left(X'\gamma\right) \right]$$

and

$$\widehat{\beta}^{wls,p}_{0,\Delta} = \arg\min_{b \in \Theta} \mathbb{E}_n \left[\frac{\Lambda\left(X'\hat{\gamma}^{ipt}\right)}{1-\Lambda\left(X'\hat{\gamma}^{ipt}\right)} \left(\Delta Y - X'b\right)^2 \mid D = 0 \right],$$

where in the third and last step, one plugs the fitted values of the working models (11) into the sample analogue of $\tau^{dr,p}$.

Given this estimator, now we are ready to extend it to multiple periods. Consider the case with $\mathcal{T}$ periods and denote a particular time period by t where $t = 1, ..., \mathcal{T}$. The first assumption is a standard one.

Assumption 4. *(Irreversibility of Treatment). $D_1 = 0$ almost surely (a.s.). For $t = 2, \ldots, \mathcal{T}$,*

$$D_{t-1} = 1 \textit{ implies that } D_t = 1 \textit{ a.s.}.$$

Assumption 4 states that no i is treated at time $t = 1$, and that once i is treated, i will remain treated in the next period.

Let G be the time period when i first experiences the treatment. Under Assumption 4, for all units that eventually participate in the treatment, G defines to which group they belong. For the case whereby a unit is never treated in any time period, we arbitrarily set $G = \infty$. Denote by G_g a binary variable equal to one if a unit is first treated in period g (i.e., $G_{i,g} = \mathbf{1}\{G_i = g\}$)) and let C be a binary variable equal to one for units that do not receive the treatment in any time period (i.e., $C_i = \mathbf{1}\{G_i = \infty\} = 1 - D_{i,T}$)). Let $\mathcal{G} = \text{supp}(G) \backslash \{\bar{g}\} \subseteq \{2, 3, \ldots, \mathcal{T}\}$ be the support of G. Given these notations, the ATT of interest for a particular group g and time t is

$$ATT(g,t) = \mathbb{E}\left[Y_t(g) - Y_t(0) \mid G_g = 1\right]. \tag{13}$$

Now, with multiple time periods, we need the following standard no anticipation assumption.

Assumption 5. *(Limited Treatment Anticipation). There is a known $\delta \geq 0$ such that*

$$\mathbb{E}\left[Y_t(g) \mid X, G_g = 1\right] = \mathbb{E}\left[Y_t(0) \mid X, G_g = 1\right] \text{ a.s. for all } g \in \mathcal{G}, t \in \{1, \ldots, \mathcal{T}\} \text{ such that } t < g - \delta.$$

This assumption is satisfied in general when the treatment path is not a priori known and/or when units are not the ones who select treatment status. The next assumption the version of the conditional PTA with multiple time periods:

Assumption 6. *(Conditional Parallel Trends based on a "Never-Treated" Group). Let δ be as defined in Assumption 5. For fach $g \in \mathcal{G}$ and $t \in \{2, \ldots, \mathcal{T}\}$ such that $t \geq g - \delta$,*

$$\mathbb{E}\left[Y_t(0) - Y_{t-1}(0) \mid X, G_g = 1\right] = \mathbb{E}\left[Y_t(0) - Y_{t-1}(0) \mid X, C = 1\right] \text{ a.s..}$$

Lastly, the following is the muti-period version of the overlap assumption:

Assumption 7. *(Multi-period overlap). For each $t \in \{2, \ldots, \mathcal{T}\}, g \in \mathcal{G}$, there exist some $\varepsilon > 0$ such that $P\left(G_g = 1\right) > E$ and $p_{g,t}(X) < 1 - \varepsilon$ a.s..*

The DRDID estimand with multiple time periods is define as

$$ATT_{dr}^{nev}(g,t;\delta) = \mathbb{E}\left[\left(\frac{G_g}{\mathbb{E}\left[G_g\right]} - \frac{\frac{p_g(X)C}{1-p_g(X)}}{\mathbb{E}\left[\frac{p_g(X)C}{1-p_g(X)}\right]}\right)\left(Y_t - Y_{g-\delta-1} - m_{g,t,\delta}^{nev}(X)\right)\right], \tag{14}$$

where $m^{nev}_{g,t,\delta}(X) = \mathbb{E}\left[Y_t - Y_{g-\delta-1} \mid X, C = 1\right]$. Callaway and Sant'Anna (2021) demonstrate that we can use the time period $t = g - \delta - 1$—i.e., the most recent time period when untreated potential outcomes are observed for units in group g—as an appropriate reference time period under Assumption 5 and Assumption 6.

Furthermore, as in their Remakr 12, while the limited anticipation condition implies that $ATT(g;t) = 0$ for all $t < g - \delta$, it is common practice to also estimate these pre-treatment effects and use them to check the credibility of the underlying identifying assumptions. We can do this easily by replacing the "long differences" $(Y_t - Y_{g-\delta-1})$ with the "short differences" $(Y_t - Y_{t-1})$ for all $t < g - \delta$. I follow their suggestions and compare 1994 against 1993, 1995 against 1994, and so forth, and then use 1997 as the reference for all the post periods.

As stated in Remark 7 of Callaway and Sant'Anna (2021), one should include pre-treatment covariates that are potentially associated with the outcome evolution of $Y(0)$ during the post-treatment periods. I follow their suggestion to use a simple average to aggregate the ATT. For the propensity score model, then, I estimate the following polynomial logistic regression, using $t - 1$ for the pre-treatment periods and $g - \delta - 1$ for the post-treatment periods:

$$\text{logit}(p(X_i)) = a_0 + aX_i + bX_i^2 + cX_i^3 + e_i, \tag{15}$$

where X_i is a vector of covariates including the outcome of the previous period, the natural logarithm of capital stock, that of employment, that of the shortest physical distance to any of the treated firms being connected, and the average physical distance to the treated firms. The first three covariates are measures of firm sizes, and the other two are used since geographic distance is known to affect a firm linkage. Furthermore, I include the polynomials of these covariates upto the third order since overfitting costs little but the fourth order turns out to be not only computationally difficult but also fails to converge.