Identifying Treatment Effects on Productivity: Theory with An Application to Firms' AI Adoption*

Zhiyuan Chen† Moyu Liao‡ Karl Schurter§

November 6, 2022

Abstract

This paper provides a semi-structural econometric framework for analyzing the treatment effects on productivity. We embed the potential outcome framework into a dynamic firm model with firm-level productivity that can be affected by some discrete treatment. The treatment can either be externally assigned or chosen by the firm. We characterize conditions for non-parametrically identifying production functions. Then we provide conditions for identifying treatment effects objects for absorbing treatment and generalize them to non-absorbing treatment. Our study has general implications for empirical studies on evaluating treatment effects on productivity. We apply our method to study the effect of AI adoption on productivity growth. Our results robustly show that AI adoption has no significant effect on the firm-level productivity in China.

Keywords: productivity; potential outcomes; dynamic treatment effects; non-parametric identification

*First Draft: April 2022. We thank comments from Ulrich Doraszelski, Sun Jae Jun, Marc Henry, Mark Roberts, Yan Shen, and participants at Penn State econometrics seminar, Asian Meeting of Econometric Society, CCER Summer Institute, IAER econometrics workshop, PSU-Cornell IO-econometrics conference. We thank Gabriele Rovigatti for sharing his Stata code for Rovigatti and Mollisi (2018) with us.

†School of Business, Renmin University. Email: chenzhiyuan@rmbs.ruc.edu.cn.

‡School of Business, Nanjing University. Email: liaomoyu@nju.edu.cn.

§Department of Economics, Pennsylvania State University. Email: kes380@psu.edu.

1 Introduction

Productivity is one of the most important yet mysterious indicator in economics. In the past three decades, advancements in approaches for estimating producer-level productivity have triggered a substantial literature on detecting productivity drivers from various perspectives.[1] Notably, productivity is not directly observed and has to be inferred from data. To recover productivity, existing production function estimation approaches rely on structural assumptions on firm behavior and Markovian productivity process (e.g., Olley and Pakes (1996); Levinsohn and Petrin (2003); Ackerberg et al. (2015); Gandhi et al. (2020)). However, the empirical regressions of estimated productivity on hypothetical productivity drivers are usually in conflict with these underlying assumptions. This can overturn the identification of production functions, and more importantly, the desired effects of internal and external drivers on productivity. While this internal inconsistency has drawn some attention from empirical researchers, there is a lack of formal analysis on this critical issue.[2]

In this paper, we provide an econometric framework for identifying and estimating treatment effects on productivity. We start with a dynamic firm model which includes a treatment influencing firms' productivity and/or production functions. The treatment that drives the change of productivity can either be chosen by the firm (e.g., R&D investment, exporting, importing, etc.) or externally given (e.g., trade liberalization, environmental regulation, tax cuts, etc.). We allow that the treatment can influence both the productivity evolution and the production function. To formalize the discussion of causal inferences, we embed a model of potential outcomes of productivity into the structural firm model. In the model, a firm has two potential productivities, but only the 'realized' productivity corresponding to its treatment status is effective in the production process. Econometrician has the classical issue of not observing the other 'unrealized' potential productivity. Compared to the classical dynamic firm investment model, we allow firms to choose the treatment status based on both of their potential productivities.

Our econometric framework has several advantages over the entirely structural method. First, the potential outcome framework enables researchers to use the language of program evaluation to describe treatment effect; Second, we do not need to completely spec-

[1] This strand of literature covers a wide range of fields including trade and development (e.g., Pavcnik (2003), De Loecker (2007), Amiti and Konings (2007), De Loecker (2013), Yu (2015), Brandt et al. (2017)), industrial organization (e.g., Doraszelski and Jaumandreu (2013), Braguinsky et al. (2015)), political economics (e.g., He et al. (2020), Chen et al. (2021), Chen et al. (2021)), public economics (Liu and Mao, 2019), etc.

[2] See Section 6.1 in the IO handbook chapter by De Loecker and Syverson (2021) for a short but insightful discussion on this problem.

ify the firm treatment decision behavior by virtue of the potential outcome framework; Third, the structural firm model allows us to conduct various counterfactual analyses that are not available if only using the potential outcome model.

We consider a simple setting to illustrate the basic ideas. The baseline model features a binary treatment indicator and a general non-parametric Markovian process for the vector of potential productivity. The treatment variable is assumed to be mean-independent of the productivity process shocks. We are interested in the treatment effect on productivity for the treated units, which is the difference between the realized productivity and counterfactual productivity. As the treatment effect on productivity is likely to change overtime, and the dynamic treatment effect can be decomposed into two parts: (1) The switching effect, which appears only at the changing point of treatment status; (2) The trend effect, which is due to the change in the productivity process after the treatment status switches.

Using the proposed generic framework, we study the non-parametric identification of the treatment effects on productivity. The first step is to recover the realized productivity [3] correctly. We show that if the potential productivity shocks are mean independent of the treatment variable, the production functions in persistent treatment regimes[4] can be non-parametrically identified up to a constant difference that depends on the treatment status. To achieve such identification result, we need two groups of units: One group is treated for at least two periods, the other is untreated for at least two periods. This means that we do not require a group of never-treated units to identify the production functions.[5] We also propose several stronger restrictions on the potential productivity evolution process that are less restrictive for using the data.

In light of our identification strategy, we illustrate how the existing approaches may fail to identify production function parameters in the context of endogenous productivity process. We examine two popular methods: the first is the ex-post regression approach which estimates productivity by assuming an exogenous productivity process[6] and then regressing the estimated productivity on an interested treatment variable (e.g., Pavcnik (2003), Amiti and Konings (2007), Yu (2015), He et al. (2020)); the second is a structural method which adds the interested treatment variable into the productivity evolution and relevant control functions for productivity (e.g., De Loecker (2013), Doraszelski and Jau-

[3]As we mentioned above, the realized productivity is the potential productivity that corresponds to the firm's treatment status. We call it the realized productivity because it participates in the production process.

[4]If the treatment status for a firm is constant for at least two periods, then we say the firm is in a persistent treatment regime.

[5]For example, a four-period panel of firms in which all firms are untreated in the first two periods and treated in the last two periods is sufficient.

[6]That is, the treatment variable is omitted in the productivity evolution process

mandreu (2013), Chen et al. (2021)). We show that both methods suffer from the problem of potential model mis-specification on the productivity evolution, which causes biases in evaluating treatment effects productivity.

Given the recovered effective productivity, we proceed to the identification of treatment effects. We first show that the conventional parallel trend assumption to identified ATTs is likely to fail if the treatment choice is not randomly assigned. We thus propose a 'conditional parallel trend assumption, which is imposed on the potential productivity evolution process and has economic meaning. Under the conditional parallel trend assumption, the instantaneous ATT is identified without further requirement. For the identification of dynamic ATTs, we tailor the 'control cohort' method (Sun and Abraham, 2021) in the dynamic treatment effect literature to our productivity context. In particular, the strutural firm model guides on which firms can be used as control cohort when we study the dynamic ATTs. Given the conditional parallel trend assumption, for each treated firm, we can identify the dynamic ATTs using two methods. The first method is to match a controlled firm based on their pre-treated realized productivity. The second approach is to simulate the counterfactual untreated potential productivity for a factual treated firm. The matching method requires a weaker model assumption but higher data quality. Conversely, the simulation-based method requires assumption on the residual of the productivity process, but it can be done for more flexible data environments.

We provide a Monte Carlo study by considering an extended productivity process that incorporates treatment variables. We illustrate that accounting for the regime-switching is essential for correctly evaluating the treatment effects. Our identification strategy performs well in detecting the full dynamic treatment effects.

To illustrate our method, we apply our method to study the adoption of AI technology on Chinese firms' productivity. We use the dataset of listed firms in the Chinese stock markets from 2005 to 2019. To identify the adoption of AI technology, we use text analysis of the listed firms' annual report to detect the reporting of using AT technology. Suitable for our methodology, our sample covers a relative long period before any AI technology was available, which allows us to simulate the counterfactual productivity path using the rich pre-trend information. Our empirical results show that the AI adoption has no significant causal effect on the productivity. Our results are robust to a series of robustness checks including different production function specifications and different simulation strategy.

This paper contributes to the literature in following ways. First, this study is closely related to a large body of literature on evaluating certain macro policies or firm-level actions on producer-level productivity (e.g., Pavcnik (2003); Amiti and Konings (2007); De Loecker (2007); Doraszelski and Jaumandreu (2013); Braguinsky et al. (2015); Brandt et al. (2017); Chen et al. (2021)). Although the underlying structural assumptions for estimating productivity is recognized, there is no econometric framework for analyzing the identification properties and estimation strategies for evaluating treatment effects on productivity. By embedding potential outcomes of productivity into the structural firm model, we provide an econometric framework and propose identification strategies for detecting the treatment effects on productivity. Second, our study is also related to the study on identifying the dynamic treatment effects (see Heckman and Navarro (2007); Abbring and Heckman (2007); Vikström et al. (2018); Sun and Abraham (2021)). In these studies, the realized outcome is assumed to be directly observed. However, in the scenario of productivity estimation, the outcome variable, which is the productivity, is estimated based on structural assumptions. Our study contributes an application of inferring dynamic treatment effects using structurally estimated productivity. More broadly, the analysis in this study can be generalized to a series of policy evaluation studies which uses structurally estimated outcomes. The current study suggests that to correctly evaluate the treatment effects, the possible objective policy impacts should be considered in the structural estimation of the outcomes. Although our study is focused on the treatment effects on productivity, our semi-structural approach has general implications for conducting policy evaluations by combining the potential outcome framework and structural modelling.

The rest of this paper is organized as follows. Section 2 describes the econometric framework for analyzing the treatment effects on productivity. We discuss the identification of production function in Section 3, and treatment effects on productivity in Section 4. Section D is the Monte Carlo simulation. Section 6 concludes the paper.

2 The Econometric Framework

2.1 A Firm Model with Treatment and Potential Productivity

A firm produces with a Hicks-neutral production technology. Both production technology and the productivity's evolution are affected by some treatment D_{it}. The treatment indicator $D_{it} \in \{0, 1\}$, with $D_{it} = 1$ indicating the firm receives the treatment. The treatment can be imposed externally (e.g., trade liberalization, environmental regulations, etc.) or chosen by the firm (e.g, R&D investment, importing and exporting, etc.). In period t, firm i has the following production function

$$Q_{it} = e^{\omega_{it}+\eta_{it}} F(K_{it}, L_{it}, M_{it}, D_{it}; \boldsymbol{\beta}), \tag{1}$$

where Q_{it} is the output quantity, ω_{it} is the realized productivity, η_{it} is some ex-post shock of productivity that is not known when firm make current period input choices, K_{it} is the capital, L_{it} is the labor, M_{it} is the material, D_{it} is the treatment, and $\boldsymbol{\beta}$ is the parameter vector. The dimension of $\boldsymbol{\beta}$ can be infinite when the production function is non-parametric. Moreover, $\boldsymbol{\beta}$ can also include a set of time dummies to account for a secular trend in the production function (e.g., Doraszelski and Jaumandreu (2013)). Note that we allow the treatment D_{it} as an input factor,[7] which captures possible impacts on managerial efficiencies (Chen et al., 2021).

There are two potential productivity outcomes ω_{it}^0 and ω_{it}^1. The binary treatment D_{it} determines the realized productivity through the following equation

$$\omega_{it} = \omega_{it}^1 D_{it} + \omega_{it}^0 (1 - D_{it}). \tag{2}$$

The firm knows its potential productivities when making decisions, but the econometrician does not. To facilitate our exposition, we define an indicator for treatment changes.

Definition 1. *(Treatment switching indicator) We define a treatment regime change indicator $G_{it} \in \{-1, 0, 1\}$: (1) Positive regime change: $G_{it} = 1$ if $D_{it} - D_{it-1} = 1$; (2) Unchanged regime: $G_{it} = 0$ if $D_{it} - D_{it-1} = 0$; (3) Negative regime change: $G_{it} = -1$ if $D_{it} - D_{it-1} = -1$.*

Conventionally, the realized productivity is assumed to follow a first-order Markov process (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg et al., 2015). We generalize this tradition to assume a Markov process for $(\omega_{it}^1, \omega_{it}^0)$:

$$\begin{pmatrix} \omega_{it}^0 \\ \omega_{it}^1 \end{pmatrix} = \mathbb{1}(G_{it} = 0)\bar{\boldsymbol{h}} \begin{pmatrix} \omega_{it-1}^0 \\ \omega_{it-1}^1 \end{pmatrix} + \mathbb{1}(G_{it} = 1)\boldsymbol{h}_i^+ \begin{pmatrix} \omega_{it-1}^0 \\ \omega_{it-1}^1 \end{pmatrix} + \mathbb{1}(G_{it} = -1)\boldsymbol{h}_i^- \begin{pmatrix} \omega_{it-1}^0 \\ \omega_{it-1}^1 \end{pmatrix} + \begin{pmatrix} \epsilon_{it}^0 \\ \epsilon_{it}^1 \end{pmatrix}, \tag{3}$$

where $\boldsymbol{h}_i^+$ (resp. $\boldsymbol{h}_i^-$) is the transition function when the regime switch is positive (resp. negative) for firm i. We allow the evolution at the transition process to possibly depend on i but impose the same evolution process when the treatment variable is constant.[8] Furthermore, the Markovian productivity process (3) is diagonal whenever there is no treatment status change:

Assumption 2.1. *(Diagonal Markov Process) The function $\bar{\boldsymbol{h}}$ satisfies*

$$\bar{\boldsymbol{h}} \begin{pmatrix} \omega_{it-1}^0 \\ \omega_{it-1}^1 \end{pmatrix} = \begin{pmatrix} \bar{h}_0(\omega_{it-1}^0) \\ \bar{h}_1(\omega_{it-1}^1) \end{pmatrix},$$

[7] Another equivalent formulation of the production function is $Q_{it} = e^{\omega_{it}} F(K_{it}, L_{it}, M_{it}; \boldsymbol{\beta}(D_{it}))$, which treats the treatment more like a factor influencing the organization of production.

[8] For example, different firms may have different regime switch time within a year, which may lead to the difference in $\boldsymbol{h}_i^+$ and $\boldsymbol{h}_i^-$.

where $\mathbb{E}[\epsilon_{it}^d|\omega_{it-1}^0, \omega_{it-1}^1] = 0$ *for* $d = 0, 1$.

Assumption 2.1 says that, the evolution of potential outcome ω_{it}^0 does not depend on the ω_{it}^1 if there is no switching in the treatment status. The assumed productivity evolution rule generalizes the productivity process considered in the productivity estimation literature. To see this, consider that $G_{it} = 0$ for all i and t, then the productivity evolution can be captured by $\omega_{it}^d = \bar{h}_d(\omega_{it-1}^d) + \epsilon_{it}^d$, for $d \in \{0, 1\}$. Therefore, we can think of the conventional productivity process (e.g. Olley and Pakes (1996)) as the case of no treatment, i.e. $D_{it} = 0$. The generalized productivity evolution process (3) also has economic meaning closely related to a wide range of empirical studies. We now give several examples for the productivity process (3). In real empirical setting, the potential productivity process can be thought as a mixture of the these examples.

Example 1. *(Parallel Shifted Productivity) In many empirical contexts, a policy simply shifts the productivity upwards. This context can be realized by imposing: (1) Initial period shift, i.e.* $\omega_{i1}^1 = \omega_{i1}^0 + C$ *almost surely for some constant* C*; (2)* $\epsilon_{it}^1 = \epsilon_{it}^0$ *almost surely for all* t*; (3) The evolution functions satisfy* $\bar{\boldsymbol{h}} = \boldsymbol{h}_i^+ = \boldsymbol{h}_i^-$ *for all* i*, and* $\bar{\boldsymbol{h}}_1(\omega) = \bar{\boldsymbol{h}}_0(\omega - C) + C$*. These conditions lead to* $\omega_{it}^1 = \omega_{it}^0 + C$ *almost surely for all* i, t.

Example 2. *(Divergence of Productivity when Treatment Diverges.) Consider a case where the binary treatment represents whether a firm invests in R&D. If a firm chooses to switch from not investing in R&D at* t *to investing in R&D at* $t + 1$*, then only* ω_{it}^0 *matters for the determination of* ω_{it+1}^1*. In this case, only the observed potential outcome before the regime switching matters for the productivity process. Similarly, if a firm decides to shut down the R&D center at period* $t + 1$*, then only* ω_{it}^1 *matters for determining the value of* ω_{it}^0.

This model can be captured by imposing: (1) $\boldsymbol{h}_i^+(\omega_{it}^0; \omega_{it}^1) = (h_{i0}^+(\omega_{it}^0), h_{i1}^+(\omega_{it}^0))'$*, where* h_{i0}^+ *and* h_{i1}^+ *are scalar functions; and (2)* $\boldsymbol{h}_i^-(\omega_{it}^0; \omega_{it}^1) = (h_{i0}^-(\omega_{it}^1), h_{i1}^-(\omega_{it}^1))'$*, where* h_{i0}^- *and* h_{i1}^- *are scalar functions. The heterogeneity in transition functions* $\boldsymbol{h}_i^+$ *and* $\boldsymbol{h}_i^-$ *across* i *can be induced by heterogeneity in the timing of decision. For example, some firms may choose to start R&D in the beginning of the year, while others may make the decision in the middle of the year.*

Example 3. *(Independent Productivity Evolution Process) In some cases, a firm needs to choose between two types of technologies. Each technology evolves without being influenced by the other technology. Firms can choose which technology to use. In this case, the regime switching is also diagonal and* $\bar{\boldsymbol{h}} = \boldsymbol{h}_i^+ = \boldsymbol{h}_i^-$*, and there is no heterogeneity in the evolution process during the transition period.*

Firms' Behavior and Timing of Firms' Decisions

We follow Ackerberg et al. (2015) and Gandhi et al. (2020) to distinguish the static inputs and the pre-determined inputs.

Assumption 2.2. *(Timing of Inputs) Capital K_{it} is determined at or before $t-1$, labor can be determined at or before $t-1$ or a static input chosen some time in period t. Intermediate input M_{it} is determined no sooner than other inputs after the realization of ω_{it}.*

The treatment variable can be either determined by the external environment or chosen by the firm. We distinguish these two cases and make the following assumption on its timing.

Assumption 2.3. *(Timing of Treatment) (1) When the treatment is externally imposed, D_{it} is determined at or before $t-1$; (2) When the treatment is a firm choice, D_{it} is chosen after the realization of $(\omega^0_{it-1}, \omega^1_{it-1})$ but before $(\omega^0_{it}, \omega^1_{it})$.*

We summarize the overall firm-decision timeline in the following graph:

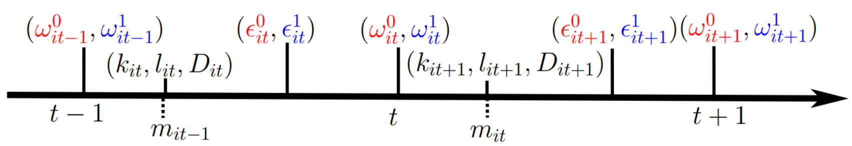


Figure 1: Timeline for firm's decision.

Let $S_{it} \equiv (K_{it}, L_{it}, D_{it}, \omega^1_{it}, \omega^0_{it}, \zeta_{it})$ be a vector of state variable, where ζ_{it} is some unobserved heterogeneity in the cost of taking the treatment. Define $\Pi(S_{it})$ as the per-period indirect profit function, the Bellman equation for the firm's dynamic programming problem is

$$V(S_{it}) = \max_{I_{it}, L_{it+1}, (D_{it+1})} \Big\{ E_\eta[\Pi(S_{it})] - C_I(I_{it}) - C_L(L_{it+1}) - C_D(D_{it+1}, \zeta_{it}) + \frac{1}{1+\rho}\mathbb{E}[V(S_{it+1})|S_{it}, I_{it}, L_{it+1}, D_{it}] \Big\} \tag{4}$$

where the I_{it} is the physical capital investment, $C_I(\cdot), C_L(\cdot), C_D(\cdot)$ are cost functions of investment, labor and adopting the treatment, respectively. The discounting rate is $1/(1+\rho)$. In problem (4), the notation (D_{it+1}) means that the treatment is not necessarily chosen by the firm. If D_{it+1} is endogenously chosen by the firm, $C_D(\cdot)$ entails the costs of selecting into the treatment.[9] If policy variable is externally assigned, D_{it} disappears from the firm's choice set. In this case, we impose that $C_D(\cdot) = 0$.[10] We summarize the timing assumption by formally stating firm i's information set $\mathcal{I}_{it}^F$:

Definition 2. *Firm i's time-t information set is given by*

$$\mathcal{I}_{it}^F = \{k_{it}, l_{it}, (\omega_{is}^0, \omega_{is}^1, D_{is-1}, k_{is-1}, l_{is-1}, m_{is-1}, \zeta_{is})_{s\leq t}\}.$$

In the case of externally assigned treatment, model (4) bears features similar to a large class of firm models considered in productivity estimation (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg et al., 2015; Gandhi et al., 2020). However, in the case of endogenously-chosen treatment, the structure of $C_D(\cdot)$ plays an important role in inferring the treatment effects on productivity.[11] Our firm model allows for the existence of unobserved heterogeneity ζ_{it}. This additional unobserved heterogeneity can bring additional difficulty of identifying the treatment effect on productivity. This is different from the endogenous productivity literature (Aw et al. (2011); De Loecker (2013); Doraszelski and Jaumandreu (2013); Peters et al. (2017)) who are interested in the productivity differences between treatment takers and non-treatment takers.

2.2 Treatment-Effect Objects

A switch of the regime, i.e. $G_{it} = 1$, influences the production process through three aspects. First, the level of productivity switches from ω_{it}^0 to ω_{it}^1. This change is instantaneous and may not be carried over time. Second, if treatment status persists, the productivity evolution process is changed from $\bar{h}_0$ to $\bar{h}_1$. This switch has a long-term effect that accumulates over time. Third, the production function can be different, i.e. the relative efficiency of inputs can be influenced by the policy.

In addition to the traditional individual treatment effect at time t: $\omega_{it}^1 - \omega_{it}^0$, we also consider other two types of treatment effects originating from the dynamic productivity process. We formally define these treatment effects as follows:

Definition 3. *The individual treatment effect for firm i at time t is $\omega_{it}^1 - \omega_{it}^0$. The trend effect is given by the function $\bar{h}_1(\cdot) - \bar{h}_0(\cdot)$.*

[9] For example, when D_{it+1} represents exporting choice, $C_D(\cdot)$ is the search and communications costs incurred when selling to foreign buyers. Also, when D_{it} is the R&D choice, $C_D(\cdot)$ is the costs installing research equipments and hiring research scientists.

[10] By assuming this, we exclude the complication of considering the problem of whether firms fully comply with the policy. See Section 3 in Abbring and Heckman (2007) for a general discussion on this issue.

[11] This is because $C_D(\cdot)$ is important for selection into treatment.

The trend effect comes from the structure of the potential productivity process. Unlike the traditional dynamic treatment effect literature where the objective outcome variable is usually observed, the productivity is unobserved, and the structural evolution process (3) is the key assumption that allows us to identify the production function parameters. Our goal is to discuss whether the treatment effects in Definition 3 are separately identified from each other and under what assumptions the treatment effects can be identified.

2.3 No-anticipation and Sequential Randomization Condition

We now briefly connect our method to the dynamic treatment effect literature (Abbring and Heckman, 2007). There are two key conditions in the dynamic treatment effect literature: No-anticipation condition (NA) and the Sequential randomization condition (SR). Since our framework combines both the potential outcome model and the structural equation model, we can use the structural model to verify whether NA and SR conditions hold or not. To simplify notation, we let $D_i^t = (D_{i1}, ..., D_{it})$, and $\omega_i^{dt} = (\omega_{i1}^d, ..., \omega_{it}^d)$ for $d = 0, 1$. We state the NA condition in our framework.

Assumption 2.4. *(NA) Let D_i^T and $\tilde{D}_i^T$ be two treatment sequence such that $D_i^t = \tilde{D}_i^t$ for any $t \leq T$. The no-anticipation condition holds if the potential $(\omega_{it}^0, \omega_{it}^1)$ generated under D_i^T coincides with the potential $(\tilde{\omega}_{it}^0, \tilde{\omega}_{it}^1)$ generated under $\tilde{D}_i^T$ for all $t \leq T$..*

The no-anticipation condition says that if two sequences of treatment coincides up to time t, then the potential productivity up to time t should also coincide. Given the Markovian evolution process (3), Assumption 2.4 holds as long as there is no anticipation in the productivity shocks: The shock sequence $(\epsilon_{is}^0, \epsilon_{is}^1)_{s\leq t}$ under D_i^t coincides with the shock sequence $(\tilde{\epsilon}_{is}^0, \tilde{\epsilon}_{is}^1)_{s\leq t}$ under $\tilde{D}_i^t$. We view Assumption 2.4 as a weak requirement since the shocks to productivity process are usually assumed to be unexpected by firms in the productivity estimation literature.

Another condition is the sequential randomization condition (Robins, 1997; Gill and Robins, 2001; Lok, 2008), which says that future potential outcomes are conditional independent of the current treatment status. Sequential randomization is crucial to the identification of treatment effects. We state the firm's SR condition in our framework.

Assumption 2.5. *(SR-F) $D_{it+1} \perp (\omega_{is}^1, \omega_{is}^0)_{s\geq t} | \mathcal{I}_{it}^F$.*

We call Assumption 2.5 the sequential randomization condition for firms since we condition on the firms' information set. This is slightly different from the traditional sequential randomization condition in Gill and Robins (2001), where they conditional on the econometrician's information set.

Our structural model implies that Assumption 2.5 holds when D_{it+1} is chosen by the firm according to (4). Indeed, from the firm's dynamic optimization problem, we know D_{it+1} is a function of $\mathcal{I}_{it}^F$, denoted by $D_{it+1} = g(\mathcal{I}_{it}^F)$. Then given the information set $\mathcal{I}_{it}^F$, D_{it+1} is a degenerative variable and thus Assumption 2.5 holds. When the treatment variable is externally imposed, and the assigner randomizes the treatment up to the firm's knowledge, i.e. $D_{it+1} = \tilde{g}(\mathcal{I}_{it}^F, \eta_{it})$ for some η_{it} independent of $(\omega_{is}^1, \omega_{is}^0)_{s\geq t}$, then SR-F also holds. We will come back to the traditional SR condition after we consider the econometrician's problem of recovering the firm-level productivity.

3 Recovering the Unobserved Productivity

Econometric analysis of the productivity relies on the Markov property of the evolution of productivity (3). However, the econometrician does not know the unobserved potential productivity.

Assumption 3.1. *The econometrician has access to the instrument set $\mathcal{Z}_{it} = \mathcal{I}_{it}^F / \{(\omega_{is}^1, \omega_{is}^0, \zeta_{is})_{s\leq t}\}$. Moreover, $E[\epsilon_{it}|\mathcal{Z}_{it}] = 0$, and $E[\eta_{it}|\mathcal{Z}_{it}, M_{it}] = 0$.*

Assumption 3.1 is standard in the classical production function estimation literature, and it is typically implied by the firms' timing assumption. The econometrician cannot observe the potential productivity and the hidden cost heterogeneity ζ_{is}. We will maintain Assumption 3.1 throughout the rest of this paper.

3.1 Recovering the Productivity in the Absence of Treatment

We first review the case where $D_{it} = 0$ for all i and t, i.e. there is no treatment at all. As a result, the realized productivity $\omega_{it} = \omega_{it}^0$ plays the role of influencing final output quantities. There are two strands of literature that use different moments to identify the production function parameters. For the gross output production function, we follow the GNR (Gandhi et al., 2020) method and use an additional material-to-revenue first order condition. For the value-added production function, we follow the ACF (Ackerberg et al., 2015) method and material proxy approach. In both cases, a conditional mean zero assumption on the productivity shocks are imposed. We use the lower and upper case letters represent logs and levels of the corresponding variables, respectively.

GNR Approach. The GNR first-order condition approach uses the following material-to-revenue share equation

$$\mathbb{E}\left[s_{it} - \log\left(\frac{\partial f_0(k_{it}, l_{it}, m_{it}; \boldsymbol{\beta})}{\partial m_{it}}\right) \Big| k_{it}, l_{it}, m_{it}\right] = 0 \quad \forall t = 1, ..., T, \tag{5}$$

where s_{it} is the logged material share and $f_0(k_{it}, l_{it}, m_{it}; \boldsymbol{\beta}) \equiv f(k_{it}, l_{it}, m_{it}, D_{it} = 0; \boldsymbol{\beta})$. The estimation of other production function parameters relies on the productivity evolution process:

$$\mathbb{E}[\omega_{it}(\boldsymbol{\beta}) - h(\omega_{it-1}(\boldsymbol{\beta}))|\{k_{it}, l_{it}, k_{it-1}, l_{it-1}, m_{it-1}\}] = 0 \quad \forall t = 1, ..., T, \tag{6}$$

the productivity is recovered from $\omega_{it}(\boldsymbol{\beta}) = q_{it} - f_0(k_{it}, l_{it}, m_{it}; \boldsymbol{\beta})$.

ACF Value-added Approach. Consider the value-added production function $f_0(k_{it}, l_{it}; \boldsymbol{\beta})$. The material m_{it} is a strictly monotone function of ω_{it} and hence the non-parametric inversion $\omega_{it} = g(k_{it}, l_{it}, m_{it})$ exists. They first identify the non-parametric object

$$\Phi_{it-1}(k_{it-1}, l_{it-1}, m_{it-1}) \equiv \mathbb{E}[q_{it-1}|k_{it-1}, l_{it-1}, m_{it-1}], \tag{7}$$

and use the moment condition

$$E\left[\omega_{it}(\boldsymbol{\beta}) - h\left[\Phi_{it-1}(k_{it-1}, l_{it-1}, m_{it-1}) - f_0(k_{it-1}, l_{it-1}; \beta)\right] \big| \{k_{it}, l_{it}, k_{it-1}, l_{it-1}, m_{it-1}\}\right] = 0. \tag{8}$$

In the absence of a policy, both methods result in non-parametric identification of the production function.

Lemma 3.1. *If there is no treatment in the model, then: (1) The moment conditions (5) and (6) identify the gross production function $\boldsymbol{\beta}$ non-parametrically up to a constant difference; (2) The moment conditions (7) and (8) identify the value-added production function $\boldsymbol{\beta}$ non-parametrically up to a constant difference. Moreover, then h is identified non-parametrically in both the GNR and ACF cases.*

Proof. The proof of statement (1) is given in GNR. We use the techniques in GNR to prove statement (2). Let $\omega_{it-1}(\boldsymbol{\beta}) \equiv \Phi_{it-1}(k_{it-1}, l_{it-1}, m_{it-1}) - f_0(k_{it-1}, l_{it-1}; \boldsymbol{\beta})$. We first note that

$\mathbb{E}[q_{it}| \{k_{it}, l_{it}, k_{it-1}, l_{it-1}, m_{it-1}\}] = f_0(k_{it}, l_{it}; \boldsymbol{\beta}) - h(\omega_{it-1}(\boldsymbol{\beta}))$. Then we have:

$$\frac{\partial \mathbb{E}[q_{it}| \{k_{it}, l_{it}, k_{it-1}, l_{it-1}, m_{it-1}\}]}{\partial k_{it}} = \frac{\partial f_0(k_{it}, l_{it})}{\partial k_{it}}$$
$$\frac{\partial \mathbb{E}[q_{it}| \{k_{it}, l_{it}, k_{it-1}, l_{it-1}, m_{it-1}\}]}{\partial l_{it}} = \frac{\partial f_0(k_{it}, l_{it})}{\partial l_{it}}$$

Therefore, f_0 is identified up to a additive constant by the existence of solution to partial differential equations. □

It is important to note that Lemma 3.1 says that the production function is identified only up to a constant difference. Mathematically, if (F, h) is identified by the GNR or ACF method, then $(e^c F, \tilde{h})$ where $\tilde{h}(\omega) = h(\omega - c)$ also satisfy the GNR or ACF moment constraints for all $c \in \mathbb{R}$. We will come back to this scale non-identification and illustrate its importance in our econometric setting.

3.2 Recovering the Productivity with Variations in Treatment Status

We now extend the identification result to the case with a policy intervention. While the treatment can be chosen by the firm, we assume a conditional exogenous treatment, i.e. the treatment is exogenous to productivity shocks $(\epsilon^1_{it}, \epsilon^0_{it})$.

Assumption 3.2. *(Conditional Mean-Zero Shocks) The productivity shock $(\epsilon^0_{it}, \epsilon^1_{it})$ satisfies*

$$\mathbb{E}[(\epsilon^0_{is}, \epsilon^1_{is})|\mathcal{Z}_{it}] = \mathbf{0}, \quad \forall s \geq t.$$

Assumption 3.2 allows the treatment decision to be dependent of the past potential outcomes ω^0_{it-1} and ω^1_{it-1}. Consider a case where D_{it} is selected by the firm. A firm may observe its productivity $(\omega^0_{it-1}, \omega^1_{it-1})$ when making the decision on whether to adopt the treatment or not, and the productivity shocks $(\epsilon^0_{it}, \epsilon^1_{it})$ realize after the firm's choice of D_{it}. When the treatment is externally determined, this assumption implies that the assignment rule of treatment is independent of productivity shocks.

Assumption 3.3. *There exist two periods t_0, t_1 such that $Pr(D_{it_0} = D_{it_0-1} = 0) \neq 0$ and $Pr(D_{it_1} = D_{it_1-1} = 1) \neq 0$.*

Theorem 3.1. *Suppose Assumptions 2.1- 3.3 hold. The moment condition (5) (and respectively (7)) and*

$$\mathbb{E}[\omega_{it}(\boldsymbol{\beta}) - \bar{h}_0(\omega_{it-1}(\boldsymbol{\beta}))|\mathcal{Z}_{it}, D_{it} = D_{it-1} = 0] = 0, \tag{9}$$

$$\mathbb{E}[\omega_{it}(\boldsymbol{\beta}) - \bar{h}_1(\omega_{it-1}(\boldsymbol{\beta}))|\mathcal{Z}_{it}, D_{it} = D_{it-1} = 1] = 0, \tag{10}$$

identify the production function parameter $\boldsymbol{\beta}$ and the evolution process $\bar{h}_d$ non-parametrically up to a constant difference that depends on d.

Proof. We first look at equation (9), and the proof of expression (10) follows similarly. We can write

$$\begin{aligned} &\mathbb{E}[\omega_{it}(\boldsymbol{\beta}) - \bar{h}_0(\omega_{it-1}(\boldsymbol{\beta}))|\mathcal{Z}_{it}, D_{it} = D_{it-1} = 0] \\ &= \mathbb{E}[\omega_{it}^0(\boldsymbol{\beta}) - \bar{h}_0(\omega_{it-1}^0(\boldsymbol{\beta}))|\mathcal{Z}_{it}, D_{it} = D_{it-1} = 0] \\ &= \mathbb{E}[\epsilon_{it}^0|\mathcal{Z}_{it}, D_{it} = D_{it-1} = 0] = 0 \end{aligned} \tag{11}$$

where $\omega_{it}^0(\boldsymbol{\beta})$ denotes the potential productivity without treatment, recovered under parameter value $\boldsymbol{\beta}$ and $D_{it} = 0$. The first equality of (11) holds by the potential outcome equation and the last equality holds by Assumption 3.2. The moment condition (9) is well defined by Assumption 3.3. By Lemma 3.1, the result follows. □

The scalar non-identification result can contaminate the identification of treatment effects. If $(F(\cdot, 0; \boldsymbol{\beta}), \bar{h}_0)$ and $(F(\cdot, 1; \boldsymbol{\beta}), \bar{h}_1)$ satisfy the moment conditions in Proposition 3.1, then $(e^{c_0}F(\cdot, 0; \boldsymbol{\beta}), , \tilde{h}_0)$ and $(e^{c_1}F(\cdot, 1; \boldsymbol{\beta}), \tilde{h}_1)$ also satisfy the moment conditions in Proposition 3.1 for $\tilde{h}_d(\omega) = \bar{h}_d(\omega - c_d)$. This means that we cannot distinguish ω_{it}^1 recovered under $(F(\cdot, 1; \boldsymbol{\beta}), \bar{h}_1)$ from $\tilde{\omega}_{it}^1$ recovered under $(e^{c_1}F(\cdot, 1; \boldsymbol{\beta}), \tilde{h}_1)$. In particular, we have $\tilde{\omega}_{it}^d = \omega_{it}^d - c_d$.[12] Therefore, we recommend normalizing the production function constant $c_1 = c_0 = 0$[13].

Other Structural Objects In this firm model, there are many other interesting structural objects, such as the transition evolution function $\boldsymbol{h}_i^+, \boldsymbol{h}_i^-$ and the treatment adoption cost $C_D(\cdot)$ in equation (4). These structural objects are generally not identified under the assumptions in Theorem 3.1: The transition evolution function $\boldsymbol{h}_i^+, \boldsymbol{h}_i^-$ are not identified because we cannot observe ω_{it}^1 and ω_{it}^0 simultaneously; The treatment adoption cost $C_D(\cdot)$ is not identified because we do not fully specify what is the cost shock ζ_{it} and how it correlates with other variables in the dynamic firm model (4).[14] Leaving the firms' treatment adoption decision unspecified gives us more flexibility on the nature of treatment.

[12] This scale non-identification is also present in the multi-product context(Chen and Liao, 2021).

[13] For example, the standard Cobb-Douglas production function satisfies $c_d = 0$ for $d = 0, 1$.

[14] For example, a government policy that subsidize the treatment adoption for small firms will create a correlation between ζ_{it} and K_{it}.

3.3 Additional Moments for Restricted Productivity Processes

Our moment conditions in Proposition 3.1 impose no additional assumptions on the productivity evolution process (3). While implementing moment conditions in Proposition 3.1 requires minimal structural assumptions, we require a relatively large sample of two-year consecutive observations as in Assumption 3.3. Such data requirement can be satisfied when the panel satisfies a difference-in-difference type design. However, if the treatment variable is volatile over time, we may need to discard a substantial fraction of the firms to implement (9) and (10), which leads to inefficient use of data. We now consider several alternative assumptions on the evolution process that allow us to derive more flexible moment conditions and make use of firms with volatile treatment status.

3.3.1 Independent Evolution Process

Let's consider the case where the two potential productivity processes evolves independently as in Example 3. In this case, we may substitute the Markov process back several periods to form additional moment conditions. Even for a firm which changes its treatment status every period, we know the treatment status every two periods must coincide. To form moment conditions for independently-productivity process, we impose the following assumption:

Assumption 3.4. *For $d = 0, 1$, the Markov process ω_{it}^d satisfies*

$$\omega_{it}^d = h_d^{(s)}(\omega_{it-s}^d) + r(\epsilon_{it}^d, ..., \epsilon_{it-s+1}^d)$$

where $h_d^{(s)}$ is an s-period transition function and $r(\cdot)$ is linear in all arguments.

Assumption 3.4 is satisfied for the well-known AR(1) process. The linearity of $r(\cdot)$ ensures that we can generalize moment conditions (9) and (10) to an s-period lagged evolution process.

Corollary 3.1. *Suppose Assumption 3.4 holds and the productivity process satisfies Example 3, then the following two moment conditions hold:*

$$\mathbb{E}[\omega_{it}(\boldsymbol{\beta}) - h_0^{(s)}(\omega_{it-s}(\boldsymbol{\beta}))|\mathcal{Z}_{it-s+1}, D_{it} = D_{it-s} = 0] = 0, \tag{12}$$

$$\mathbb{E}[\omega_{it}(\boldsymbol{\beta}) - h_1^{(s)}(\omega_{it-s}(\boldsymbol{\beta}))|\mathcal{Z}_{it-s+1}, D_{it} = D_{it-s} = 1] = 0. \tag{13}$$

Moment conditions (12) and (13) allow us to use a larger fraction of firms in the dataset. However, we recommend combining moment conditions (12) and (13) with (9) and (10) to estimate the production functions unless Assumption 3.3 fails. Its unfortunate that we cannot show non-parametric identification of production function with moments (12) and (13) alone: The error terms ϵ_{it-s} is correlated with k_{it} and l_{it} for all $s \geq 1$ and thus they are not in instrument set $\mathcal{Z}_{it-s+1}$. Therefore, we cannot apply the GNR trick to differentiate both sides of (12) to identify the production function.

One may argue that k_{it-s+1} and l_{it-s+1} can serve as instruments for k_{it} and l_{it}. However, without solving the firms' dynamic optimization problem, we cannot establish the functional relationship between (k_{it}, l_{it}) and (k_{it-s+1}, l_{it-s+1}), and we cannot prove the non-parametric identification of production functions. When the production function is Cobb-Douglas, the log-linear form of the production function along with the valid instrument k_{it-s+1} and l_{it-s+1} allow us to identify the production function parameters and the evolution process.

3.3.2 Divergent Productivity Processes

Now we consider the productivity process in Example 2. We only consider homogeneous transition process, i.e. $\boldsymbol{h}_i^+ = \boldsymbol{h}^+$, $\boldsymbol{h}_i^- = \boldsymbol{h}^-$. To simplify the notation, we remove the unit subscripts and write the transition functions as $\boldsymbol{h}^+(\omega_{it}^0, \omega_{it}^1) = (h_0^+(\omega_{it}^0), h_1^+(\omega_{it}^1))$ and $\boldsymbol{h}^-(\omega_{it}^0, \omega_{it}^1) = (h_0^-(\omega_{it}^0), h_1^-(\omega_{it}^1))$. Since only the observed productivity matters for the evolution process, we can further derive the moment conditions at the transition periods.

Corollary 3.2. *Suppose Assumptions 2.1-3.3 hold and the productivity evolution process satisfies Example 2 with homogeneous transition process: $\boldsymbol{h}_i^+ = \boldsymbol{h}^+$, $\boldsymbol{h}_i^- = \boldsymbol{h}^-$. Then the moment condition (5) (and respectively (7)), (9), (10) and*

$$\mathbb{E}[\omega_{it}(\boldsymbol{\beta}) - h_1^+(\omega_{it-1}(\boldsymbol{\beta}))|\mathcal{Z}_{it}, D_{it} = 1, D_{it-1} = 0] = 0, \tag{14}$$

$$\mathbb{E}[\omega_{it}(\boldsymbol{\beta}) - h_0^-(\omega_{it-1}(\boldsymbol{\beta}))|\mathcal{Z}_{it}, D_{it} = 0, D_{it-1} = 1] = 0, \tag{15}$$

identify the production functions, and the evolution processes $\bar{\boldsymbol{h}}_d$, h_1^+, h_0^- non-parametrically up to a constant difference.

The additional moment conditions in Corollary 3.2 when the panel is short or when we only observe one period after the treatment status changes. Compared to the moment conditions in Doraszelski and Jaumandreu (2013), Corollary 3.2 requires the transition period to be treated separately from the consistent treatment status period. Moment conditions (14) and (15) are imposed to identify the positive transition process h_1^+ and h_0^-, separately.

3.4 A Revisit to Existing Methods

In this section, we use a simple example to illustrate the limitations of two commonly used methods in recovering the productivity with the presence of treatment: the ex-post regression method (Pavcnik, 2003; Amiti and Konings, 2007; Yu, 2015; Chen et al., 2021; He et al., 2020) and the endogenous productivity evolution method (De Loecker, 2007; Doraszelski and Jaumandreu, 2013; Chen et al., 2021). Without loss of generality, we assume that the production function is treatment-invariant.

We consider a simple "difference-in-difference" policy context: An exogenous policy shock happens at $t = T_0 + \Delta$ for $\Delta \in (0, 1)$. A random subset of firms is influenced by the policy while others are not, and firms are separated into treated and control groups. For the firms in the controlled group, $D_{it} = 0$ for all t. In this context, the policy variable D_{it} is fully exogenous to the productivity process.[15] For the firms in the treated group, $D_{it} = \mathbb{1}(t \geq T_0 + 1)$.

We use this empirical context to show that the ex-post regression method is invalid, and the endogenous productivity method can only accommodate very restricted empirical scenarios. We also define an alternative instrument set $\mathcal{Z}'_{it} = \{k_{it}, l_{it}, k_{it-1}, l_{it-1}, m_{it-1}\}$ that is used the expost-regression method.

The Ex-post Regression

The ex-post regression method consists of two steps: First, it estimates the firm model ignoring the existence of policy effect. To do so, it estimates the production function parameter β and the evolution process h using (5) and (6). Second, given the estimated parameter $\hat{\boldsymbol{\beta}}$ and $\hat{h}$, recover the pseudo firm-level productivity $\hat{\omega}_{it} = q_{it} - f(k_{it}, l_{it}, m_{it}; \hat{\boldsymbol{\beta}})$. They analyze the individual treatment effect based on $\hat{\omega}_{it}$. For example, ex-post regression method runs a two-way fixed effect regression by treating $\hat{\omega}_{it}$ as the outcome variable.

There are two problems in this procedure. First, the trend difference $\bar{h}_1 \neq \bar{h}_0$ is ignored in this model. Second, if there is any trend effect in the potential outcome process, the moment equality (6) fails.

We first note that for all $t \leq T_0$, the moment equation (6) becomes

$$\mathbb{E}[\omega^0_{it}(\boldsymbol{\beta}) - \bar{h}_0(\omega^0_{it-1}(\boldsymbol{\beta}))|\mathcal{Z}'_{it}] = 0 \quad \forall t \leq T_0.$$

By Proposition 3.1, this moment condition identifies $\boldsymbol{\beta}$ and $\bar{h}_0$. We now derive the incon-

[15]The policy exogeneity is only imposed here for illustration purpose.

sistency of (6). For $t \geq T_0 + 2$, the moment condition (6) becomes

$$
\begin{aligned}
(6) &= \mathbb{E}[\omega_{it}(\boldsymbol{\beta}) - \bar{h}_0(\omega_{it-1}(\boldsymbol{\beta}))|\mathcal{Z}'_{it}] \\
&=_{(1)} \mathbb{E}[\omega^0_{it}(\boldsymbol{\beta}) - \bar{h}_0(\omega^0_{it-1}(\boldsymbol{\beta}))|\mathcal{Z}'_{it}, D_{it} = 0]Pr(D_{it} = 0) \\
&+ \mathbb{E}[\omega^1_{it}(\boldsymbol{\beta}) - \bar{h}_0(\omega^1_{it-1}(\boldsymbol{\beta}))|\mathcal{Z}'_{it}, D_{it} = 1]Pr(D_{it} = 1) \\
&=_{(2)} \underbrace{\mathbb{E}[\omega^0_{it}(\boldsymbol{\beta}) - \bar{h}_0(\omega^0_{it-1}(\boldsymbol{\beta}))|\mathcal{Z}'_{it}]}_{\text{Part A}} Pr(D_{it} = 0) + \underbrace{\mathbb{E}[\omega^1_{it}(\boldsymbol{\beta}) - \bar{h}_0(\omega^1_{it-1}(\boldsymbol{\beta}))|\mathcal{Z}'_{it}]}_{\text{Part B}} Pr(D_{it} = 1)
\end{aligned}
\tag{16}
$$

where $\boldsymbol{\beta}$ and $\bar{h}_0$ are the quantities identified from moment conditions $t \leq T_0$, and we use the exogenous policy assumption to derive the equality (2). Part A in equation (16) is zero because it is consistent with the moment condition $t \leq T_0$. However, if $\bar{h}_1 \neq \bar{h}_0$, then Part B is not zero and the moment condition (6) fails for all $t \geq T_0 + 2$.

Under the mis-specified model, the estimator $\hat{\boldsymbol{\beta}}$ is not consistent for the true $\boldsymbol{\beta}$. As a consequence, $\hat{\omega}_{it}$ is not a consistent estimator of ω_{it}, and the subsequent treatment effect evaluation is incorrect.

The Endogenous Productivity Method

The endogenous productivity method in De Loecker (2007) and Doraszelski and Jaumandreu (2013) includes the interested treatment variable in the productivity process as:

$$
\omega_{it} = \tilde{h}(\omega_{it-1}, D_{it}) + \epsilon_{it}.
$$

This method solves the misspecification of the productivity process for treated and controlled group. Indeed, by defining $\bar{h}_d(\cdot) = \tilde{h}(\cdot, d)$ for $d = 0, 1$, we can show that moment condition (6) can be transforms to

$$
\begin{aligned}
&\mathbb{E}[\omega^0_{it}(\boldsymbol{\beta}) - \bar{h}_0(\omega^0_{it-1}(\boldsymbol{\beta}))|\mathcal{Z}_{it}] = 0 \quad \forall t \leq T_0, \quad \text{and} \\
&\mathbb{E}[\omega^0_{it}(\boldsymbol{\beta}) - \bar{h}_0(\omega^0_{it-1}(\boldsymbol{\beta}))|\mathcal{Z}_{it}, D_{it} = D_{it-1} = 0]Pr(D_{it} = D_{it-1} = 0) \\
+&\mathbb{E}[\omega^1_{it}(\boldsymbol{\beta}) - \bar{h}_1(\omega^1_{it-1}(\boldsymbol{\beta}))|\mathcal{Z}_{it}, D_{it} = D_{it-1} = 1]Pr(D_{it} = D_{it-1} = 1) \quad \forall t \geq T_0 + 2,
\end{aligned}
\tag{17}
$$

and the moment condition at the regime-switching period $T_0 + 1$:

$$
\begin{aligned}
&\underbrace{\mathbb{E}[\omega^0_{iT_0+1}(\boldsymbol{\beta}) - \bar{h}_0(\omega^0_{iT_0}(\boldsymbol{\beta}))|\mathcal{Z}'_{iT_0+1}, D_{iT_0+1} = D_{iT_0} = 0]}_{\text{Part A}} Pr(D_{iT_0+1} = D_{iT_0} = 0) + \\
&\underbrace{\mathbb{E}[\omega^1_{iT_0+1}(\boldsymbol{\beta}) - \bar{h}_1(\omega^0_{iT_0}(\boldsymbol{\beta}))|\mathcal{Z}'_{iT_0+1}, D_{iT_0+1} = 1, D_{iT_0} = 0]}_{\text{Part B}} Pr(D_{iT_0+1} = 1, D_{iT_0} = 0) = 0.
\end{aligned}
\tag{18}
$$

Moment condition (17) is correctly specified. In particular, by Proposition 3.1, $\boldsymbol{\beta}$, $\bar{h}_0$ are identified from the $t \leq T_0$ moment equality (17), and $\bar{h}_1$ is identified from the $t \geq T_0 + 2$ moment equality (17).

However, the moment condition at the regime switching period (18) is misspecified. Let $\bar{h}_0$ be identified from (17). Part A in (18) equals zero. However, the Part B may not equal zero. Given the evolution process (3), the transition process at the positive regime switching period should be $h_i^+(\omega_{it-1}^1, \omega_{it-1}^0)$, where in the Part B of (18), the transition process is $\bar{h}_1(\omega_{it-1}^0)$. This will lead to a possible misspecification issue. We now show that for the examples in Section 2, the structural evolution method only works with strong assumptions.

Let's first consider Example 1. At time $T_0 + 1$, the treated firm's observed last period productivity is the untreated potential outcome $\omega_{iT_0}^0$. In particular, consider the following productivity process: (1) $\omega_{it}^1 = \omega_{it-1}^1$; (2) $\omega_{it}^0 = \omega_{it-1}^0$; (3) $\omega_{it}^1 = \omega_{it}^0 + C$. In this case, productivity is constant over time, and both $\bar{h}_1$ and $\bar{h}_0$ are the identity map. The transition functions also satisfy $h_i^+ = h_i^- = \bar{h}$. Therefore, the Part B of (18) becomes $\mathbb{E}[\omega_{iT_0+1}^1(\boldsymbol{\beta}) - \omega_{iT_0}^0(\boldsymbol{\beta}) | \mathcal{Z}_{iT_0+1}, D_{iT_0+1} = D_{iT_0} = 1]$. The moment value of the Part B is C at the true production parameter rather than 0, so the model is misspecified.

In Example 2, the evolution at the transition period only depends on the observed outcome in the last period. If we impose $\boldsymbol{h}_i^+ = \boldsymbol{h}^+ = \bar{\boldsymbol{h}}$, then the Part B of (18) equals zero and the model is not misspecified. However, this is a strong assumption and may not be satisfied in some empirical contexts. Let's consider the regime switch happens at $T_0 + \Delta$ for some $\Delta < 1$. In this case, $\omega_{iT_0}^0$ first evolves to $\omega_{iT_0+\Delta}^0$ under the controlled process $\bar{h}_0$, and then the treatment status changes and the productivity evolves from $\omega_{iT_0+\Delta}^0$ to $\omega_{iT_0+1}^1$. In other words, the productivity only enjoys the benefit of the policy effects during the period $[T_0 + \Delta, T_0 + 1]$. If the policy variable D_{iT_0+1} affects the productivity process at the beginning of the period, then it is likely that $\bar{\boldsymbol{h}} \neq \boldsymbol{h}_i^+$.

4 Evaluating the Treatment Effect on Productivity

Recall that the treatment effect of interest are given in Definition 3. Since we only observe a firm either in the treated or non-treated state, the individual treatment effect $\omega_{it}^1 - \omega_{it}^0$ is typically not identified, and we instead focus on the average treatment effect (ATE) and the average treatment effect on the treated (ATT).

Corollary 4.1. *Under Assumption 2.1-3.3, if there exists a t such that $Pr(D_{it} = D_{it-1} = d) \neq 0$, we can recover the potential productivity $\omega_{it}^d + \eta_{it}$ for firms such that $D_{it} = d$.*

Proof. Recall that from Proposition 3.1, $\boldsymbol{\beta}$ and the evolution process $\bar{h}_d$ is identified. As a result, if firm i's treatment status is $D_{is} = d$, we can recover productivity $\omega_{is} + \eta_{is} = (q_{is} - f(k_{is}, l_{is}, m_{is}, D_{is}; \boldsymbol{\beta})$, which is $\omega_{is}^d + \eta_{is}$ since $D_{is} = d$. $\square$

Since the individual effective productivity is identified, the econometrician can view ω_{it} as 'observed' up to a mean zero random perturbation η_{it}. In many cases, η_{it} is purely random and cannot be separated from the firm productivity. We thus omit the η_{it} in our discussion below. We now define the econometrician's information set as below.

Definition 4. *The econometrician's information set is* $\mathcal{I}_{it}^E = \mathcal{Z}_{it} \cup \{\omega_{is}\}_{s \leq t-1} \subset \mathcal{I}_{it}^F$.

For the identification of ATE, we introduce a version of the sequential randomization condition based on the econometrician's information set. For ATT, we find it instructive to discuss the identification for absorbing treatment and non-absorbing treatment, separately.

4.1 ATT: Absorbing Treatment

The absorbing treatment is at the core of literature on estimating dynamic treatment effects (Sun and Abraham, 2021; Athey and Imbens, 2022). As a benchmark for analyzing ATT, we consider the absorbing policy for which the treatment indicator is non-decreasing $D_{it-1} \leq D_{it}$. For any treatment that is not absorbing, we can replace the treatment status D_{it} with an indicator for ever having received the treatment to obtain a new treatment being absorbing.[16]

Let $e_i > 1$ be the first period that firm i starts to receive treatment.[17] Since the treatment is absorbing, when firm i belongs to the treated group, we have $G_{it} = 1$ for $t = e_i$ and $D_{it} = 1$ for all $t \geq e_i$. We maintain Assumption 3.2 on the exogeneity of productivity shocks. Let g be a set of indicators for any subset of firms in the treated group, and $\ell \geq 0$ be the time relative to the first treatment period. We are interested in the ℓ-period-ahead ATT at time t for group g is given by

$$ATT_{g,\ell} = \mathbb{E}[\omega_{it}^1 - \omega_{it}^0 | t = e_i + \ell, i \in g]. \tag{19}$$

Failure of the Simple Parallel Trend Assumption Even the treatment is not randomly assigned, the Difference-in-Difference method allows us to identify the ATT if a parallel

[16] For example, Deryugina (2017) defines the treatment to be "having had any hurricane" and investigates its impact on the fiscal cost for a county.

[17] We exclude units who are always treated during the sample period due to a lack of an appropriate comparison group.

trend assumption is satisfied. We first look at a simple parallel trend assumption that is needed in the two-way fixed effect regression:

Assumption 4.1. *(Simple Parallel Trend) The following condition holds:*

$$\mathbb{E}[\omega_{it}^0 - \omega_{it-1}^0 | e_i = t] = \mathbb{E}[\omega_{it}^0 - \omega_{it-1}^0 | e_i > t]. \tag{20}$$

If condition (20) holds, then the $ATT_{g,0}$ is identified as $\mathbb{E}[\omega_{it}|e_i = t] - \mathbb{E}[\omega_{it-1}|e_i = t] - (\mathbb{E}[\omega_{it}|e_i > t] - \mathbb{E}[\omega_{it-1}|e_i > t])$. However, Assumption 4.1 is a high-level condition because it is imposed on the potential productivity before and after the treatment and can be hard to justify. To see it, note that from the productivity process (3), we can derive that:

$$\begin{aligned} \text{positive switchers: } \mathbb{E}[\omega_{it}^0 - \omega_{it-1}^0 | e_i = t] &= \mathbb{E}[h_{i0}^+(\omega_{it-1}^1, \omega_{it-1}^0) - \omega_{it-1}^0 | e_i = t], \\ \text{non switchers: } \mathbb{E}[\omega_{it}^0 - \omega_{it-1}^0 | e_i > t] &= \mathbb{E}[\bar{h}_0(\omega_{it-1}^0) - \omega_{it-1}^0 | e_i > t], \end{aligned} \tag{21}$$

where we use the condition (3.2) to derive (21). From (21) we see that the parallel trend condition can fail due two reasons: (1) The transition processes at the regime switch period can be different for the treated and controlled group; (2) Even when the two transition processes coincide, the treatment D_{it} can depend on the value of ω_{it-1}^0 and can correlate with the initial treatment time e_i. Consider Example 2 with a R&D decision, the firm chooses to invest in R&D only when ω_{it-1}^0 exceeds a certain level. In this case, D_{ie_i} is a function of ω_{it-1}^0, and (20) does not hold.

The Conditional Parallel Trend Assumption Now, we propose an alternative procedure that identifies the $ATT_{g,\ell}$ when the transition processes at the regime switch period coincide for the treated and controlled group. First we note that, by further conditional on the value of ω_{it-1}^0 in equation (21), we have

$$\begin{aligned} &\mathbb{E}[\omega_{it}^0 - \omega_{it-1}^0 | e_i = t, \omega_{it-1}^0] = h_{i0}^+(\omega_{it-1}^1, \omega_{it-1}^0) - \omega_{it-1}^0, \\ &\mathbb{E}[\omega_{it}^0 - \omega_{it-1}^0 | e_i > t, \omega_{it-1}^0] = \bar{h}_0(\omega_{it-1}^0) - \omega_{it-1}^0. \end{aligned} \tag{22}$$

The two equations in (22) coincide if $h_{i0}^+ = \bar{h}_0$. We call this the conditional parallel trend assumption.

Assumption 4.2. *(Conditional Parallel Trend)* $h_{i0}^+(\omega_{it}^0, \omega_{it}^1) = \bar{h}_0(\omega_{it}^0)$.

Assumption 4.2 is structural in the sense that it is imposed on the rule of productivity evolution rather than the cross-period potential outcome variables $(\omega_{it}^0, \omega_{it-1}^0)$. The structural parallel trend assumption 4.2 has the following economic meaning: Transition function for the untreated potential outcome is not influenced by the treatment status.

Proposition 4.1. *Under Assumption 4.2, the 0-period-ahead ATT is identified as* $ATT_{g,0} = \mathbb{E}[\omega_{ie_i} - \bar{h}_0(\omega_{ie_i-1})|i \in g]$.

Proof. Note that by further conditional on the group $e_i = t$,

$$\begin{aligned}(ATT_{g,0}|e_i = t) &=_{(1)} \mathbb{E}[\omega_{it}|e_i = t, i \in g] - \mathbb{E}[h^{+}_{i0}(\omega^0_{it-1}, \omega^1_{it-1}) + \epsilon^0_{it}|e_i = t, i \in g] \\ &=_{(2)} \mathbb{E}[\omega_{it}|e_i = t, i \in g] - \mathbb{E}[\bar{h}_0(\omega^0_{it-1})|e_i = t, i \in g] \\ &=_{(3)} \mathbb{E}[\omega_{it}|e_i = t, i \in g] - \mathbb{E}[\bar{h}_0(\omega_{it-1})|e_i = t, i \in g]\end{aligned}$$

where (1) by definition, (2) follows by Assumptions 3.2 and 4.2, (3) follows by the potential outcome (2). Further take the expectation with respect to the treatment time e_i to get the result. □

In general, the ℓ-period-ahead ATT is not identified for $\ell \geq 1$, because we cannot recover the untreated potential outcome $\omega^0_{ie_i+\ell-1}$. Moreover, the substitution of in Proposition 4.1 does not work without further restrictions. We now give several assumptions that help identify the ℓ-period-ahead ATT.

For notation purpose, let $\bar{h}^{\ell}_0$ be the ℓ-period productivity transition process, we can write $\omega^0_{ie_i+\ell} = \bar{h}^{(\ell)}_0(\omega^0_{ie_i}, (\epsilon^0_{is})^{e_i+l}_{s=e_i})$.

Assumption 4.3. *The Markov process* ω^0_{it} *satisfies*

$$\omega^0_{it} = \bar{h}^{(s)}_0(\omega^0_{it-s}) + r(\epsilon^d_{it}, ..., \epsilon^d_{it-s+1})$$

where $\bar{h}^{(s)}_0$ *is an* s*-period transition function and* $r(\cdot)$ *is linear in all its arguments.*

Proposition 4.2. *Under Assumption 3.2, 4.2, and 4.3, the* ℓ*-period-ahead ATT is identified as* $(ATT_{g,\ell}|e_i = t) = \mathbb{E}[\omega_{ie_i+\ell}|i \in g] - \mathbb{E}[\bar{h}^{(\ell)}_0(\omega_{ie_i-1})|i \in g]$.

Proof. Note that conditional on $e_i = t$,

$$\begin{aligned}ATT_{g,\ell} &=_{(1)} \mathbb{E}[\omega_{it+\ell}|e_i = t, i \in g] - \mathbb{E}[\bar{h}^{(\ell)}_0(\omega_{it-1}) + r(\epsilon^0_{it}, ..., \epsilon^0_{it-s+1})|e_i = t, i \in g] \\ &=_{(2)} \mathbb{E}[\omega_{it+\ell}|e_i = t, i \in g] - \mathbb{E}[\bar{h}^{(\ell)}_0(\omega_{it-1})|e_i = t, i \in g],\end{aligned}$$

where (1) by the conditional parallel trend assumption and Assumption 4.3. Note that the treatment is absorbing, so $E[r(\epsilon^0_{it}, ..., \epsilon^0_{it-s+1})|e_i = t, i \in g] = E[r(\epsilon^0_{it}, ..., \epsilon^0_{it-s+1})|D_{it} = 1, i \in g]$. As a result, (2) follows by Assumptions 3.2 and linearity of $r(\cdot)$. The result in the proposition follows by further taking expectation with respect e_i. □

Assumption 4.3 is satisfied for an AR(1) productivity process, but generally fails when non-linearity appears in the transition function $\bar{h}_0$. Therefore, Assumption 4.3 can be restrictive. We now consider a strong constraint on the productivity shocks but relax the constraint on the shape of $\bar{h}_0$.

Assumption 4.4. *There is a group-time pair (g', s) such that all firms i' such that $i' \in g'$ are untreated by l-periods since time s, i.e. $e_{i'} > s + l$. Moreover, the conditional distribution of $(\epsilon^0_{ie_i}, \ldots, \epsilon^0_{ie_i+l})|(i \in g, \omega^0_{is-1})$ is the same as the conditional distribution of $(\epsilon^0_{i's}, \ldots, \epsilon^0_{i's+l})|(i' \in g', \omega^0_{is-1})$.*

We will discuss an example for the group g and g' later. The group g' firms serve as the controlled match for the treated firms in g. We require more than the conditional mean-independence of the future productivity shocks with respect to the treatment time. Assumption 4.4 allows for nonlinearity in $\bar{h}^{(s)}_0(\cdot)$. The overall ATT can be derived by integrating out the group effect.

Proposition 4.3. *Suppose Assumption 4.2 and 4.4 hold. Let $\Delta(\ell, g, \omega) = \mathbb{E}[\omega_{ie_i+\ell}|i \in g, \omega_{ie_i-1} = \omega] - \mathbb{E}[\omega_{is+\ell}|i \in g', \omega_{is-1} = \omega]$. The ℓ-period-ahead ATT for group g is identified as*

$$ATT_{g,\ell} = \mathbb{E}[\Delta(\ell, g, \omega_{ie_i-1})|i \in g],$$

where the expectation is taken over the conditional distribution of ω_{ie_i-1} given $i \in g$.

Proof. Note that the average treatment effect conditional on the ω_{ig-1} is

$$\begin{aligned} ATT_{g,\ell}|_{\omega_{ie_i-1}=\omega} &=_{(a)} \mathbb{E}[\omega_{ie_i+\ell}|i \in g, \omega_{ie_i-1} = \omega] - \mathbb{E}[\bar{h}^{(\ell)}_0(\omega_{ie_i-1}, \epsilon^0_{ie_i}, \ldots, \epsilon^0_{ie_i+\ell})|i \in g, \omega_{ie_i-1} = \omega] \\ &=_{(b)} \mathbb{E}[\omega_{ie_i+\ell}|i \in g, \omega_{ie_i-1} = \omega] - \mathbb{E}[\bar{h}^{(\ell)}_0(\omega_{is-1}, \epsilon^0_{is}, \ldots, \epsilon^0_{is+\ell})|i \in g', \omega_{is-1} = \omega] \\ &=_{(c)} \mathbb{E}[\omega_{ie_i+\ell}|i \in g, \omega_{ie_i-1} = \omega] - \mathbb{E}[\omega_{is+\ell}|i \in g', \omega_{is-1} = \omega] = \Delta(l, g, \omega), \end{aligned} \tag{23}$$

where (a) follows by the conditional parallel trend assumption and the potential outcome equation, (b) follows by Assumptions 4.4, and (c) follows by the transition procedure (3) for untreated firms. The result follows by further integrating out the ω_{ig-1}. □

Proposition 4.3 requires us to match over the lagged productivity for each group g-firms with g'-firms since time s. This is because we cannot observe the untreated shocks ϵ^0_{it} for treated firms and the higher order moments of ϵ^0_{it} matters for the ℓ-period evolution process $\bar{h}^{(\ell)}_0$. On the other hand, Proposition 4.2 uses firm i's own lagged productivity as controls, because the linearity of the residual function $r(\cdot)$ in Assumption 4.3.

To identify $ATT_{g,l}$ in Proposition 4.3, we typically need to match a treated firm with a controlled firm with the same lagged productivity. This matching procedure can be hard to implement. We now impose a stronger version of Assumption 4.4 that allows us to find better match.

Assumption 4.5. *The conditional distribution of $(\epsilon^0_{ie_i}, ..., \epsilon^0_{ie_i+l})|(i \in g, \omega^0_{is-1})$ is the same as the distribution of $(\epsilon^0_{ie_i}, ..., \epsilon^0_{ie_i+l})|(i \in g)$.*

Assumption 4.5 requires the future shocks to be independent of the lagged productivity. Moreover, the distribution of productivity shocks is independent of the productivity.

Proposition 4.4. *Suppose Assumption 4.2, 4.4, and 4.5 hold. Let $G^{(l)}_{0|g'}$ be the conditional distribution of $(\epsilon^0_{i's}, ..., \epsilon^0_{i's+l})|(i' \in g')$. Then, the ℓ-period-ahead ATT for cohort g is identified as*

$$ATT_{g,\ell} = \mathbb{E}\left[\omega_{ie_i+l} - \mathbb{E}_{G^{(l)}_{0|g'}}\left[\bar{h}^{(l)}_0(\omega_{ie_i-1}, \epsilon_{i's}, ..., \epsilon_{i's+l})\middle| i' \in g'\right]\middle| i \in g\right]. \tag{24}$$

Instead of matching each treated firm with a controlled firm with the same lagged productivity, Assumption 4.5 combines the tricks in Proposition 4.2 and 4.3: We plug the treated firm i's lagged productivity ω_{ie_i-1} into the l-period evolution process $\bar{h}^{(l)}_0$, but we use the g' firms to deal with the non-linearity in the productivity process. Compared to Proposition 4.3, each treated firm is matched with all firms in the controlled group g'.

Example 4. *In many empirical setting, we are interested in a cohort of firms which start their treatment in period g_0: $g = \{i : e_i = g_0\}$. In this case, we can use the $g_0 + l + 1$-not-yet-treated firms as the control: $g' = \{i' : e_{i'} > g_0 + l\}$ and $s = g_0$.*

In this case, Assumption 4.4 hold under the following empirical context: Before g_0, no firms are treated. At time g_0, firms can decide whether to take the absorbing treatment. Between g_0 and $g_0 + l + 1$, firms cannot change their treatment status due to regulations or contracts. At time $g_0 + l + 1$, untreated firms can again decide whether to take the treatment or not.

Then at the time g_0, firm make the decision on whether their initial treatment time e_i is $e_i = g_0$ or $e_i > g_0 + l$, and firms can only make treatment choice based on their information set $\mathcal{I}^F_{ig_0}$, which does not contain information on future shocks $(\epsilon^0_{ig_0}, ..., \epsilon^0_{ig_0+\ell})$. This example can be seen in many government policy reforms that rolls out in several phases. For example, the privatization of Chinese State-Owned enterprise starts with a experiment phase in northeast provinces, and gradually roll out to the rest of the country.

However, if all firms can choose the initial treatment time freely after g_0, then Assumption 4.4 typically fails for the g_0+l+1-not-yet-treated firms: A firm chooses not to be treated until g_0+l+1 are likely the firms whose potential productivity $(\epsilon^0_{ie_i}, ..., \epsilon^0_{ie_i+l})$ are high and they are reluctant to *switch to the treated status. In this case, we can use $g' = \{all\ firms\}$, and $s = 1$: We use all firms' before the period of availability of treatment g_0 as the pool for match of firms treated at g_0 period.*

There are two shortcomings of Assumption 4.3: (1) We may not be able to find a (g', s) pair that satisfies the independence restriction; (2) Even when (g', s) is found, we may not have enough observation in group-time pair (g', s). Moreover, if all firms are treated at $g_0 + l$, we cannot make inference on the ATT_{g_0+l+s} for all $s > 0$. We now propose a stronger condition:

Assumption 4.6. *(i). The shocks satisfy $\epsilon^0_{is} \sim_{i.i.d.} G^0_\epsilon(\cdot)$, where the $i.i.d$ is over both firm index i and time index s. (ii). We can find a group-time pair (g', s) such that all firms in g' are untreated in period-s. (iii). There is no selection in shocks: $\epsilon^0_{i's}|i' \in g' \sim G^0_\epsilon(\cdot)$ and $\epsilon^0_{it}|i \in g \sim G^0_\epsilon(\cdot)$ for all $t \geq e_i$.*

There are two things in Assumption 4.6: First, we assume that the productivity shocks are $i.i.d$ across both firms and time. This assumption allow us to impute the unobserved productivity shocks for group-g firms using the distribution G^0_ϵ; Second, we can find a controlled-matching group g' and time s such that the marginal distribution of ϵ^0_{is} is identified.

Proposition 4.5. *Under Assumption 4.2, 4.4, and 4.5, 4.6, G^0_ϵ is identified, and the ℓ-period-ahead ATT for cohort-g is identified as*

$$ATT_{g,\ell} = \mathbb{E}[\omega_{ie_i+\ell}|i \in g] - \mathbb{E}_{(G^0_\epsilon)^\ell}[\bar{h}^{(\ell)}_0(\omega_{ie_i-1}, \epsilon^0_{is}, ..., \epsilon^0_{is+\ell})|i \in g],$$

where the second expectation is taken over the joint distribution of $(\epsilon^0_{ie_i}, ..., \epsilon^0_{ie_i+\ell})$.

Proof. With the identified $\bar{h}_0$ from Proposition 3.1, for any group-g' firm i at time s, we can recover its $\epsilon^0_{is} \equiv \omega_{is} - \bar{h}_0(\omega_{is-1})$, so the distribution G^0_ϵ is identified. By condition (iii) in Assumption 4.6, the joint distribution of $(\epsilon_{ie_i-1}, ..., \epsilon_{ie_i+l})$ is identified as $(G^0_\epsilon)^\ell$. The identification result follows by the evolution process (11). □

Remark 4.1. *We consider the group g as the cohort g_0 in Example 4. In this case, at time g_0, firms select into the treated and controlled groups based on the information set $\mathcal{I}^F_{ig_0}$. After deciding the treatment status in period g_0, the firms then get the $i.i.d$ draw of random shocks $\epsilon^0_{ig_0}$. This timing assumption implies that the distribution of $\epsilon^0_{ig_0}|e_i = g_0$ should be the same as the distribution of $\epsilon^0_{ig_0}|e_i > g_0$, and the distribution should be G^0_ϵ.*[18]

[18]However, if we look at the distribution of $\epsilon^0_{ig_0}|e_i = g_0 + 1$ and $\epsilon^0_{ig_0}|e_i > g_0 + 1$, these two distributions will not be the same. This is because the treatment decision at $g_0 + 1$ depends on the shock $\epsilon^0_{ig_0}$.

4.2 ATT: Non-absorbing Treatment

In some scenarios, the treatment is non-absorbing by nature. In reality, firms do participate in import, export, or R&D activities occasionally.[19] We now discuss the identification of effects of non-absorbing treatment. Since treatment can be volatile, the individual treatment effect can be influenced by a sequence of past treatment status[20].

Many dynamic treatment effects are not identified under the volatile treatment context. Instead, we focus on some treatment effect of firms that switch its treatment status at time g and maintain the status for ℓ-period. Formally, the ATT for the ℓ-period persistent treatment for a time g-positive/negative treatment switcher:

$$\begin{aligned} ATT^{+}_{g,\ell} &= \mathbb{E}[\omega^{1}_{ig+\ell} - \omega^{0}_{ig+\ell} | D_{ig-1} = 0, D_{ig} = ... = D_{ig+\ell} = 1] \\ ATT^{-}_{g,\ell} &= \mathbb{E}[\omega^{1}_{ig+\ell} - \omega^{0}_{ig+\ell} | D_{ig-1} = 1, D_{ig} = ... = D_{ig+\ell} = 0] \end{aligned} \tag{25}$$

Here the group is defined by the time when a firm switches its treatment status. We first show that the 0-period ahead treatment effect is identified under the conditional parallel trend assumption for both negative and positive switcher.

Proposition 4.6. *Under Assumption 4.2, the 0-period-ahead positive/negative switching ATT effects at time g are identified as* $ATT^{+}_{g,0} = \mathbb{E}[\omega_{ig} - \bar{h}_0(\omega_{ig-1}) | D_{ig-1} = 0, D_{ig} = 1]$, *and* $ATT^{-}_{g,0} = \mathbb{E}[\omega_{ig} - \bar{h}_1(\omega_{ig-1}) | D_{ig-1} = 1, D_{ig} = 0]$.

Proof. We prove the result for the positive switching effect $ATT^{+}_{g,0}$, and the negative switching ATT follows similarly. Note that for regime change indicator $G_{ig} = 1$,

$$\begin{aligned} ATT^{+}_{g,0} &=_{(a)} \mathbb{E}[\omega^{1}_{ig} - \omega^{0}_{ig} | D_{ig-1} = 0, D_{ig} = 1] \\ &=_{(b)} \mathbb{E}[\omega^{1}_{ig} | D_{ig-1} = 0, D_{ig} = 1] - \mathbb{E}[\bar{h}_0(\omega^{0}_{ig-1}) | D_{ig-1} = 0, D_{ig} = 1] \\ &=_{(c)} \mathbb{E}[\omega_{ig} | D_{ig-1} = 0, D_{ig} = 1] - \mathbb{E}[\bar{h}_0(\omega_{ig-1}) | D_{ig-1} = 0, D_{ig} = 1], \end{aligned}$$

where (a) by definition, (b) follows by Assumptions 3.2 and 4.2, (c) follows by the potential outcome (2). □

Similar to the absorbing-treatment case, evaluating the ℓ-period-ahead ATT requires additional structural assumption on the exogeneity of shocks.

[19] In the data on Taiwanese electronics industry employed by Aw et al. (2011), the annual transition probability from only R&D performer in year t to R&D performer in year $t+1$ is around 0.57, and the probability from only exporter in year t to exporter in year $t+1$ is around 0.78. In the Spanish data used by Doraszelski and Jaumandreu (2013), slightly more than 20% of firms are occasional performers that undertake R&D activities in some (but not all) years.

[20] See Heckman and Navarro (2007) for formal definition of the general dynamic treatment effects

Assumption 4.7. *There is a cohort group g' such that all firms i' such that $i' \in g'$ are untreated by l-periods since time g', i.e. $D_{ig'-1} = D_{ig'} = ... = D_{ig'+l} = 0$. Moreover, the conditional distribution of $(\epsilon^0_{ig}, ..., \epsilon^0_{ig+l})|(i \in g, \omega^0_{ig-1})$ is the same as the conditional distribution of $(\epsilon^0_{i'g'}, ..., \epsilon^0_{i'g'+l})|(i' \in g', \omega^0_{ig'-1})$.*

Assumption 4.7 generalizes Assumption 4.4 to the non-absorbing treatment case using firms that are not treated between g' and $g' + l$. Since treatment is not absorbing, we need to further conditional on the lagged treatment D_{it-1}.

Proposition 4.7. *Suppose Assumption 4.2 and 4.7 hold. The $ATT^+_{g,l}$ is identified by the same expression as (24).*

Remark 4.2. *Assumption 4.7 has a similar restriction as Assumption 4.4. However, if firms are allow to change the treatment status every period, then the g'-matching cohort is very hard to find: The l-period untreated firms are likely to face a very high $\epsilon^0_{ig'}$, and hence these firms are not good match for the g-cohort firms.*

However, Assumption 4.7 is likely to hold for treatment that must be maintained for several periods: For example, the treatment decision is whether to use a new technology, and the new technology is not available before time g. At time g, firms can decide whether to take the new technology, and the contract for adopting the new technology must last for at least l period. In this case, we can use all firms at $g' = 0$ as the match group.

4.3 ATE: The Sequential Randomization Condition

Given the econometrician's information set $\mathcal{I}^E_{it}$, we seek for conditions ensuring the identification of the average treatment effect for a particular group of firms. We use g to denote the set of firms of primary interest such as a cohort.

Assumption 4.8. *(SR-E) $D_{it} \perp (\omega^1_{is}, \omega^0_{is})_{s \geq t} | \mathcal{I}^E_{it}, i \in g$.*

The econometrician's sequential randomization Assumption 4.8 is imposed on the potential productivity and may or may not be satisfied in different empirical settings.

If D_{it} is externally imposed and absorbing, and the assigner randomize the treatment up to the econometrician's knowledge[21], i.e. $D_{it} = \breve{\psi}(\mathcal{I}^E_{it}, \eta_{it})$ for some η_{it} independent of $(\omega^0_{is}, \omega^1_{is})_{s \geq t}$, then SR-E holds, and we are able to estimate ATE using matching techniques. This condition is not restrictive in many external policy settings, since the treatment assigner can only observe limited information as the econometrician does. For example, a tax reduction policy can be assigned to firms with more capital stocks, up to a pure

[21] For example, the treatment is assigned to firms whose capital stock is greater than a threshold

randomization η_{it}. However, if the treatment assigner has more information than the econometrician, for example the treatment assigner has some knowledge on the potential productivity, sequential randomization still fails.

If D_{it} is chosen by the firm, the dynamic optimization problem implies $D_{it} = \breve{\psi}(\mathcal{I}_{it}^F)$ for some unknown function $\breve{\psi}$. The randomness of D_{it} after conditional on $\mathcal{I}_{it}^E$ comes from the unobserved cost heterogeneity ζ_{it}, and the potential productivity $\omega_{it-1}^0, \omega_{it-1}^1$. In our general framework, SR-E fails and the ATE are in not identified.

However, we can show that the SR-E condition is satisfied when we consider the productivity evolution process that satisfies Example 2, and the firm needs to decide whether to participate in an absorbing treatment[22]. Let e_i be the firm i's initial treatment time, i.e. $D_{it} = 1$ for all $t \geq e_i$. We focus on the group g that is not yet treated at time $g-1$, i.e. $e_i \geq g$. For example, a firm needs to decide whether to build a R&D research center at time g. For this group g, the productivity process in Example 2 implies that $\omega_{is}^1 = \omega_{is}^0$ for all $s < g$. This is similar to the empirical setting in Doraszelski and Jaumandreu (2013). For a firm at time g, based on the dynamic optimization problem (4), we can write $D_{ig} = \psi(K_{ig}, L_{ig}, D_{ig-1}, \omega_{ig-1}, \zeta_{ig})$ for some unknown function ψ, where we use the condition $\omega_{ig-1}^1 = \omega_{ig-1}^0 = \omega_{ig-1}$. On the other hand, the evolution process implies $\omega_{ig}^1 = h^+(\omega_{ig-1}) + \epsilon_{ig}^1$ and $\omega_{ig}^0 = \bar{h}_0(\omega_{ig-1}) + \epsilon_{ig}^0$. As long as $(\epsilon_{ig}^0, \epsilon_{ig}^1)$ is independent of the unknown cost ζ_{ig}, then the sequential randomization $(\omega_{ig}^1, \omega_{ig}^0) \perp D_{ig} | \mathcal{I}_{ig}^E, e_i \geq g$ holds.

Proposition 4.8. *Let $\kappa(\mathcal{I}_{it}^E) = \mathbb{E}[D_{it} | \mathcal{I}_{it}^E, i \in g]$ be the propensity score that lies strictly between 0 and 1. Then the average treatment effect for group-g firms at time t is identified as:*

$$ATE_{g,t} = \mathbb{E}\left[\frac{\omega_{it} D_{it}}{\kappa(\mathcal{I}_{it}^E)} - \frac{\omega_{it}(1 - D_{it})}{1 - \kappa(\mathcal{I}_{it}^E)} \Big| i \in g\right]$$

Proposition 4.8 follows directly by the propensity score matching method.

4.4 Counterfactual Treatment Effect

Treatment effect objects such as ATT and ATE are useful when we take a retrospective evaluation of the treatment or policy effect. However, in many settings, policy makers are deciding whether to apply the same treatment policy to a counterfactual group of firms based on the knowledge from the current available data.

In this section, we consider a program that rolls out in several phases and the treatment status is absorbing. We start with an initial full set of firms (denoted by $\mathcal{S}$) that are not treated. At time t_0, a subset of firms become treated (denoted by $\mathcal{S}^{tr}$) while the

[22]For illustration purpose, we only consider the absorbing treatment, i.e. $D_{it} \geq D_{it-1}$ for all t.

rest of firms remains untreated (denoted by $\mathcal{S}^{ut}$). Untreated firms cannot change their treatment status unless a new phase of the program begins. A policy maker stands at time period $t_0 + s$ have access to firm-level data up to time $t_0 + s - 1$ and need to decide whether to start a new phase of the program. There are many empirical examples where the treatment program rolls out in several phases: For example, the State-Owned Enterprise reform in China[23] first took place in Northeast provinces and rolled out to the rest of the country in several phases.

The policy maker is interested in the treatment effects on the untreated group $\mathcal{S}^{ut}$, while the treatment effects identified in previous sections are evaluated using the whole sample $\mathcal{S}$. These two quantities in general does not coincide even when the policy is fully a randomized controlled experiment. This is because the treatment effect objects at time t_0 depends on the distribution of potential outcome $\omega^1_{it_0-1}$. While a fully randomized treatment ensures that $F(\omega^1_{it_0-1}|i \in \mathcal{S}^{tr}) = F(\omega^1_{it_0-1})$, the s-period ahead distribution of potential outcome $\omega^1_{it_0+s-1}$ will not be the same as the $t_0 - 1$ period potential outcome distribution, i.e. $F(\omega^1_{it_0+s-1}|i \in \mathcal{S}^{tr}) \neq F(\omega^1_{it_0-1})$, unless the productivity distribution is stationary.

We therefore seek to characterize the counterfactual treatment effect objects that allows the policy maker to evaluate the value of extending the program to the rest of the firms at time $t + s$. In general, without imposing further structural assumptions other than Assumptions 2.2-4.8, it is almost impossible to identify the counterfactual treatment effect objects: The target treatment effect is defined as the difference $\omega^1_{it_0+s} - \omega^0_{it_0+s}$, but for the $\mathcal{S}^{ut}$ firms, we have at best the knowledge of $\omega^0_{it_0+s-1}$ but not $\omega^1_{it_0+s-1}$. We therefore consider several additional structural assumptions that allow us to evaluate the counterfactual treatment effects defined in the following:

$$ATE^{count}_{s,l} \equiv E[\omega^1_{it_0+s+l} - \omega^0_{it_0+s+l}|i \in \mathcal{G}], \tag{26}$$

which is the l-period ahead counterfactual treatment effect for group $\mathcal{G} \subseteq \mathcal{S}^{ut}$ firms if the treatment take place at time $t_0 + s$.

4.4.1 Divergent Productivity Process

Recall that the difficulty of characterizing the counterfactual treatment effect comes from the lack of knowledge of $\omega^1_{it_0+s-1}$. However, the divergent productivity process in Example 2 implies the coincidence of two potential outcomes before treatment status changes:

[23] This is known as the privatization process of state-owned enterprise.

$\omega^0_{it_0+s-1} = \omega^1_{it_0+s-1}$ for all $s \leq 0$. Therefore, we can characterize the counterfactual treatment effect.

Proposition 4.9. *Let the productivity evolution process satisfies Example 2. Moreover, suppose the conditional parallel trend assumption 4.2 holds. For a subset $\mathcal{G} \subseteq \mathcal{S}^{ut}$ of not-yet treated firms at time $t_0 + s$ that are assigned to take treatment at $t_0 + s$, the instantaneous counterfactual treatment effect is identified as*

$$ATE^{count}_{s,0} = E[h^+_1(\omega_{it_0+s-1}) - \bar{h}_0(\omega_{it_0+s-1})\big|i \in \mathcal{G}],$$

where h^+_1 is identified from Proposition 3.2.

Proof. By the divergent productivity process assumption, $\omega^1_{it_0+s} = h^+_1(\omega_{it_0+s-1}) + \epsilon^1_{it+s}$ and $\omega^0_{it_0+s} = h^+_0(\omega_{it_0+s-1}) + \epsilon^0_{it+s}$. The parallel trend assumption implies that $h^+_0 = \bar{h}_0$. The result follows by the conditional mean zero condition: $E[\epsilon^d_{it+s}|D_{it+s}] = 0$ for $d \in \{0, 1\}$. $\square$

For l period ahead counterfactual treatment effect, we need additional structural assumptions on the productivity process shocks so that we can simulate the productivity process several periods ahead.

Assumption 4.9. *(i). The shocks satisfy $\epsilon^d_{it} \sim_{i.i.d.} G^d_\epsilon(\cdot)$ for $d \in \{0, 1\}$, where the $i.i.d$ is over both firm index i and time index t. (ii). No selection in pre-treatment shocks: $\epsilon^0_{it} \sim_{i.i.d} G^0_\epsilon(\cdot)$ for $t < t_0$. (iii) No selection in already-treated group shocks: $\epsilon^1_{it}|i \in \mathcal{S}^{tr} \sim_{i.i.d} G^1_\epsilon(\cdot)$ for $t_0 \leq t < t_0 + s$.*

Assumption 4.9 is similar to Assumption 4.6 except that we also require the distribution $G^1_\epsilon(\cdot)$ is identified from the already treated firms $\mathcal{S}^{tr}$. This is because for the factual treatment, we can observe the ω^1_{it+s} once the firms are treated. However, for the counterfactual treatment effect, we need to simulate both the treated and untreated future productivity.

Proposition 4.10. *Under Assumption 4.2, 4.4, and 4.9, $G^0_\epsilon, G^1_\epsilon$ are identified, and the ℓ-period-ahead counterfactual treatment effect at period $t_0 + s$ is identified as*

$$\begin{aligned} ATE^{count}_{s,\ell} = &\mathbb{E}_{(G^1_\epsilon)^\ell}[\bar{h}^{(l-1)}_1(h^+_1(\omega_{it_0+s-1}, \epsilon^1_{it_0+s}), \epsilon^1_{it_0+s+1}, ..., \epsilon^1_{it_0+s+l})|i \in \mathcal{G}] \\ &- \mathbb{E}_{(G^0_\epsilon)^\ell}[\bar{h}^{(\ell)}_0(\omega_{it_0+s-1}, \epsilon^0_{is}, ..., \epsilon^0_{is+\ell})|i \in \mathcal{G}], \end{aligned}$$

where the expectation on $(G^d_\epsilon)^l$ is taken over the joint distribution of $(\epsilon^d_{it_0+s}, ..., \epsilon^d_{it_0+s+\ell})$.

Remark 4.3. *The characterization of counterfactual treatment effect also highlights another reason in favor of the potential productivity process over the endogenous productivity method (Doraszelski and Jaumandreu, 2013). Recall that Doraszelski and Jaumandreu (2013) do not model the* *transition period and implicitly assume that $h_1^+ = \bar{h}_1$ in the identifying moment condition. While imposing $h_1^+ = \bar{h}_1$ may not lead to large bias in the production function estimates when the panel is long, it does lead to a bias in the counterfactual treatment effect, especially the instantaneous treatment effect $ATE_{s,0}^{count}$.*

4.4.2 Stationary Conditional Potential Outcome Moments

In more general models, we do not have information on the ω_{it+s-1}^1 for the not-yet-treated group $\mathcal{S}^{ut}$. We now investigate conditions where we can transfer the knowledge of the factual treatment effect to the counterfactual treatment effect. In particular, we want stationary conditional distribution of potential outcomes:

Assumption 4.10. *The distribution of $\omega_{it_0-1}^1|(\omega_{it_0-1}^0 = w,\ i \in \mathcal{S}^{tr})$ is the same as the distribution of $\omega_{it_0+s-1}^1|(\omega_{it_0+s-1}^0 = w,\ i \in \mathcal{G})$.*

Assumption 4.10 is the high-level assumption on potential productivity distribution. There are two constraints embedded in 4.10: 1. No selection in the potential outcome. This is reflected in the requirement that we condition on the different firm groups $\mathcal{S}^{tr}$ and $\mathcal{S}^{ut}$; 2.The conditional distribution of ω_{it}^1 is stationary at time $t_0 - 1$ and $t_0 + s - 1$.

Proposition 4.11. *Suppose Assumptions 4.9 and 4.10 hold. The counterfactual treatment effect is identified as*

$$ATE_{s,l}^{count} = \mathbb{E}\left\{\mathbb{E}[\omega_{it_0+\ell}|i \in \mathcal{S}^{tr}, \omega_{it_0-1} = \omega_{it_0+s-1}]\middle| i \in \mathcal{G}\right\} - \mathbb{E}_{(G_\epsilon^0)^\ell}[\bar{h}_0^{(\ell)}(\omega_{it_0+s-1}, \epsilon_{is}^0, ..., \epsilon_{is+\ell}^0)|i \in \mathcal{G}].$$

Proof. We first note that

$$\begin{aligned}
&\mathbb{E}[\omega_{it_0+\ell}|i \in \mathcal{S}^{tr}, \omega_{it_0-1} = \omega_{it_0+s-1}] \\
&= \mathbb{E}_{(G_\epsilon^1)^\ell, \omega_{it_0-1}^1}[\bar{h}_1^{(l-1)}(h_1^+(\omega_{it_0-1}^1, \omega_{it_0-1}^0, \epsilon_{it_0}^1), \epsilon_{it_0+1}^1, ..., \epsilon_{it_0+l}^1)|i \in \mathcal{S}^{tr}, \omega_{it_0-1} = \omega_{it_0+s-1}] \\
&=_{(*)} \mathbb{E}_{(G_\epsilon^1)^\ell, \omega_{it_0+s-1}^1}[\bar{h}_1^{(l-1)}(h_1^+(\omega_{it_0+s-1}^1, \omega_{it_0+s-1}^0, \epsilon_{it_0+s}^1), \epsilon_{it_0+s+1}^1, ..., \epsilon_{it_0+s+l}^1)|i \in \mathcal{G}, \omega_{it_0+s-1}] \\
&= \mathbb{E}[\omega_{it_0+s+l}^1|i \in \mathcal{G}, \omega_{it_0+s-1}]
\end{aligned}$$

where we use Assumptions Assumptions 4.9 and 4.10 in the $(*)$ step.

Then the counterfactual treatment effect is identified as

$$
\begin{aligned}
ATE^{count}_{s,l} &= \mathbb{E}\left\{ \mathbb{E}[\omega^1_{it_0+s+l} | i \in \mathcal{G}, \omega_{it_0+s-1}] - \mathbb{E}[\omega^0_{it_0+s+l} | i \in \mathcal{G}, \omega_{it_0+s-1}] \middle| i \in \mathcal{G} \right\} \\
&= \mathbb{E}\left\{ \mathbb{E}[\omega_{it_0+\ell} | i \in \mathcal{S}^{tr}, \omega_{it_0-1} = \omega_{it_0+s-1}] \middle| i \in \mathcal{G} \right\} \\
&- \mathbb{E}_{(G^0_\epsilon)^\ell}[\bar{h}^{(\ell)}_0(\omega_{it_0+s-1}, \epsilon^0_{is}, ..., \epsilon^0_{is+\ell}) | i \in \mathcal{G}].
\end{aligned}
$$

The result follows. □

The identified counterfactual treatment effect in Proposition 4.11 use two different approaches to impute the unrealized future potential productivities. For the treated future potential productivity, we use the stationary distribution assumption and use the already treated firms to impute the ω_{it_0+s+l}. In particular, we match each not-yet-treated firm at time t_0+s-1 with a already-treated firm at time t_0-1 with the same realized productivity. For the untreated potential future productivity, we simulate the productivity process into the future.

5 Empirical Study

The rise of AI technology in various industries have caught much attention from the academia and policymakers (Gans, 2022a,b). Although the AI technology comes with great potential, many empirical studies have documented a negative impact of artificial intelligence on productivity, at least in the short run (Brynjolfsson et al., 2017). Empirically, whether the AI technology and/or AI products has spurred productivity growth remains an open question. Existing findings on the influence of AI on productivity are mixed, which calls for more empirical studies. In light of our econometric framework, we evaluate the effect of AI products and related technologies on firms' productivity and contribute new evidence on the literature on productivity effects of AI.

5.1 Data

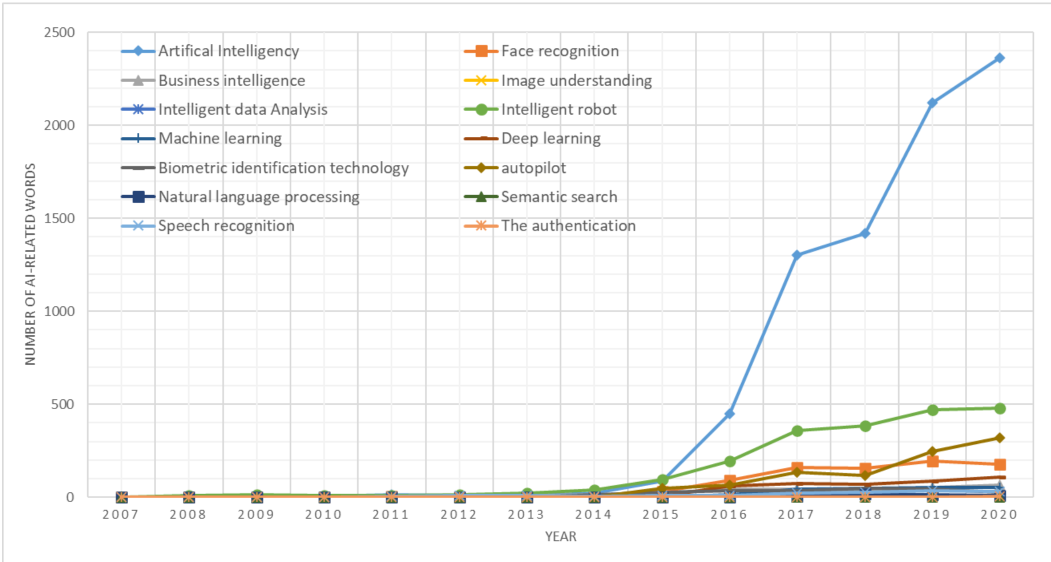
The empirical study combines two datasets. The first dataset is on publicly traded manufacturing firms in China stock market between 2005 and 2019. This dataset is collected by CSMAR (equivalent to the US Compustat) and contains rich information on firms' production activities. The second database is the Database for Research on China's Digital Economy (DRCDE hereafter) provided by CSMAR, from which we extract the information on firms' adoption of AI. This database contains information on firms' reporting of using digital technologies including AI, cloud computing, blockchain, big data, etc. CSMAR obtains such information by searching in the annual reports of each firm to detect keywords that are related to different technologies. For AI-related keywords, CSMAR uses a group of words to capture representative AI technologies. These keywords include *"artificial intelligence", "business intelligence", "intelligent data analysis", "intelligent robot", "machine learning", "deep learning", "biometric identification technology", "autopilot", "natural language processing", "natural language processing", "speech recognition", and "the authentication"*. Note that the word "artificial intelligence" is a general notion and not related to a specific type of AI technology.[24] In Appendix A, we show that the word frequency of "artificial intelligence" is disproportionately higher than other keywords in the CSMAR database. To better capture the firm's AI-adopting strategy, we identify firms as AI-product producers if they report at least "artificial intelligence" and other AI keywords simultaneously in their annual reports. To show the validity of this strategy, we have performed several checks. First, we have manually read a random sample of the annual reports and find that almost all the keywords refer to AI products rather than the adoption of AI firms. We provide text examples in Appendix A. Second, we show that the industry with higher intensity of AI words tend to be industries that are likely to produce AI products.

Figure 2 Panel (a) displays the strong growing trend between 2007 and 2020, indicating that more and more AI technology related words had been reported during the sample period. The surge in AI technology adoption is a relatively new phenomenon. In the data, during 2007 and 2013, the total number of AI keywords is 292 which is smaller than the total number of keywords in 2015. Between 2015 and 2020, the average annual growth rate of the word frequency has been around 97%. The frequency of the word "artificial intelligence" is still the mostly detected compared to other AI keywords. This may imply that Chinese public firms may also use the catchy words to attract investors' attention.

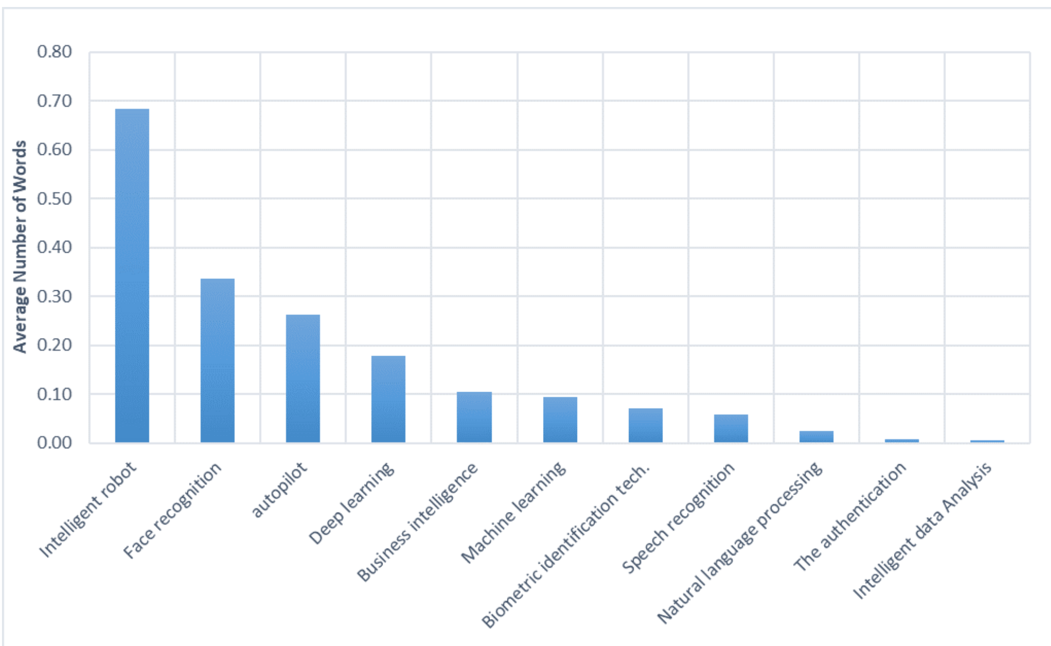
[24] In Appendix A, we present example texts to illustrate that only using the word "AI" to define the firm's adoption of AI technology can be misleading.

Figure 2: Trend of AI Adoption

(a) Trend of the Number of AI Keywords



(b) Average Frequency of Specific AI Technology Words

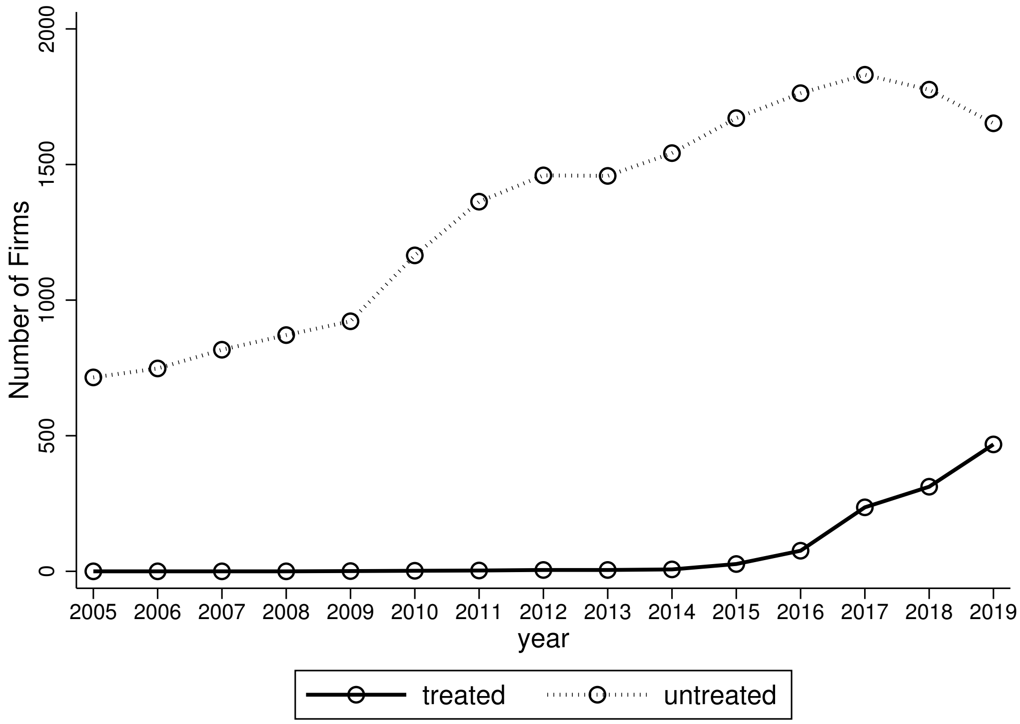


Note: Based on authors' calculation using DRCDE from CSMAR. In Panel (b),

To understand the relative prevalence of different types of AI technologies, in Panel (b) we display the average number of different AI keywords in the sample of annual reports documenting AI technologies. The top three keywords are "intelligent robot", "face recognition" and "autopilot", and the bottom three keywords are "natural language pro-

H

Figure 3: Number of AI firms and non-AI firms



Note: "treated" means firms that have reported "artificial intelligence" and other AI words simultaneously, and "untreated" refers to firms that have never reported that.

cessing", "the authentication", and "intelligent data analysis". The ranking is consistent with the usual products of manufacturing firms.

Figure 3 shows that the absolute number of firms that have reported AI-related words has been increasing steadily, indicating that more and more firms have produced AI products. The growing trend of AI adoption in our sample provides us a suitable empirical background to study the impact of introducing AI products on productivity. The number of untreated firms grows during the sample period, this is because an increasing number of firms have become public firms due to the development of China's stock market. In contrast, we do not find any firm exit in the data. This is because only well-performing firms are allowed to be listed firms, and they tend to survive well after they become public firms. We merge these two datasets to perform the empirical analysis of AI products on productivity. To maintain sufficient observations for both the treated and untreated group, we focus on the industries with at least 50 observations reporting AI-related words. Our final sample for estimation contains 8 manufacturing industries and 9314 observations. The final sample accounts for around 81% of the total AI word counts, 77% of the number of annual reports reporting AI words, and around 54% of the total observations of public manufacturing firms. Details of data processing and variable construction are reported in Appendix A.

5.2 Estimation Procedure

We estimate the treatment effects of AI on productivity using a two-step procedure. In the first step, we employ the approach suggested by Ackerberg et al. (2015) to estimate a value-added production function, with the extension that the productivity process is a controlled Markov process. We choose the Cobb-Douglas specification as the benchmark. The logged Cobb-Douglas production function has the following form:

$$y_{it} = \beta_t t + \beta_l l_{it} + \beta_k k_{it} + \omega_{it} + \varepsilon_{it} \tag{27}$$

where y_{it} is the logged value added, $\beta_t t$ captures the exogenous trend in the production function, and ε_{it} is the exogenous idiosyncratic output shocks. Consistent with our econometric framework, the realized productivity ω_{it} can be expressed as $\omega_{it} = AI_{it} \times \omega_{it}^1 + (1 - AI_{it})\omega_{it}^0$, where $AI_{it} \in \{0, 1\}$ is the indicator for AI adoption, with $AI_{it} = 1$ if firm i reported AI adoption in the year $s \leq t$, and $AI_{it} = 0$ otherwise. The Cobb-Douglas form is restrictive in the sense that the output elasticity for each input is a constant, and the elasticity of substitution between two inputs is equal to one. As a robustness check, we also estimate the Translog production function. We specify the following dynamic equation for the productivity process:

$$\omega_{it}^d = \rho_0^d + \rho_1^d \omega_{it-1}^d + \rho_2^d (\omega_{it-1}^d)^2 + \rho_3^d (\omega_{it-1}^d)^3 + \xi_{it}^d, \quad d \in \{0,\, 1\}, \tag{28}$$

where $d = 1$ indicating firms that have adopted AI, and $d = 0$ otherwise. Therefore, AI_{it} captures whether the firm has adopted AI or not. Consistent with the specification in Equation (28), we use the following empirical model for the productivity process:

$$\begin{aligned} \omega_{it} =& \rho_0 + \rho_1 \omega_{it-1} + \rho_2 \omega_{it-1}^2 + \rho_3 \omega_{it-1} \\ &+ \rho_4 AI_{it-1} + \rho_5 AI_{it} \times \omega_{it-1} + \rho_6 AI_{it} \times \omega_{it-1}^2 + \rho_7 AI_{it} \times \omega_{it-1}^3 + \xi_{it} \end{aligned} \tag{29}$$

As we have emphasized in the current paper, the productivity process in the switching period is not well characterized by the above specification. To deal with this issue, we drop the switching period when estimating the production function in which firms start to report AI technologies in their annual reports.[25] Based on (29), we form moment condition using instruments suggested by Ackerberg et al. (2015). To account for industrial heterogeneity, we estimate the production function by industry so that production function parameters and the productivity process are allowed to be completely different across different industries.

In the second step, with the estimated production function, we compute the productivity and back out the productivity evolution process. To obtain the potential productivity of not being treated for the treated firms that have adopted AI technology, we randomly draw productivity shocks ξ_{it}^0 from the pre-treatment period in which all firms had not been treated separately for each industry and simulate the counterfactual productivity path for each treated firm.[26] Note that this resampling strategy avoids the bias originating from the firm's selection into treatment based on ξ_{it}^0. For example, if firms which expect lower ξ_{it}^0 in the future are more likely to produce AI products, using all the not-yet-treated observations would squeeze the density of ξ_{it}^0 to the left. Denote $\hat{\omega}_{it}^0$ as the potential productivity outcome of treated firm i in year t. Then the ATT on productivity in period ℓ after AI adoption is calculated as:

$$\widehat{ATT}_\ell = \frac{1}{N_\ell} \sum_{i \in \mathcal{G}_\ell} \left(\hat{\omega}_{it} - \hat{\omega}_{it}^0 \right), \tag{30}$$

where $\mathcal{G}_\ell$ is the set of firms that have adopted AI technology for ℓ years, and N_ℓ is the number of firms in set $\mathcal{G}_\ell$.

For the sake of sample size, we choose $\ell = 0, 1, 2, 3$. We use the bootstrap method the confidence interval for the ATT estimates. To account for the impact of first-step estimation on the second step, we resample firms for each industry by firm-level clusters and repeat the two-step estimation procedure. Considering that different industries have different intensity of AI technology adoption, we also bootstrap the sample by strata of treated firms and untreated firms. We bootstrap 300 times to obtain the confidence interval of the ATT estimates.

5.3 Empirical Results

Baseline Results. We report the estimation results in Table 1. For the specification of Cobb-Douglas production function, we find positive estimates on treatment effects on produc-

[25] Given that the panel is relatively long, we anticipate that dropping the transition period would not affect much the estimation outcomes.

[26] Because very few firms were treated before the year of 2011, we use all untreated observations before year 2011 to construct the candidates for drawing ξ_{it}^0.

tivity, with magnitudes varying across periods. However, none of the estimates are statistically significant at the 5% significance level, which means that AI has not caused significant productivity growth among these public firms in China. These results are robust when using the Translog production function. The wide confidence interval may reflect that only a handful of firms gained productivity premium from introducing AI products to the market.

Table 1: Treatment Effects of AI Products on Productivity

	Cobb-Douglas		Translog		
Periods	ATT	95% CI	ATT	95% CI	Obs.
0	0.124	(-1.677, 2.247)	0.104	(-1.408, 1.726)	255
1	0.378	(-1.721, 2.819)	0.327	(-1.579, 2.188)	164
2	1.577	(0.492, 5.567)	0.290	(-1.490, 2.159)	115
3	0.310	(-2.124, 4.386)	0.203	(-3.148, 3.338)	40
Total	0.501	(-1.369, 2.611)	0.212	(-1.616, 2.021)	574

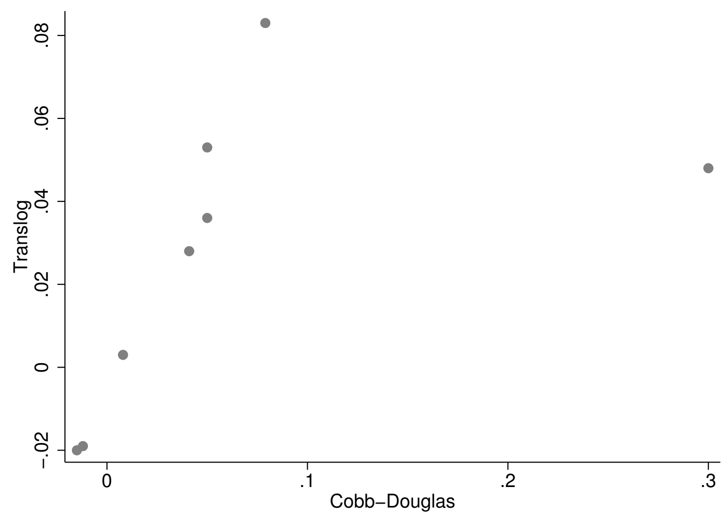
Note: Confidence intervals are obtains by bootstrapping 300 times.

Table 2: Industry-level Treatment Effects on Productivity

	Cobb-Douglas		Translog		
2-digit Industry	ATT	Contribution	ATT	Contribution	Obs.
chemical products	-0.191	-0.031	-0.251	-0.095	46
electronic components	0.245	0.100	0.175	0.170	118
other electronics	0.463	0.100	0.486	0.248	62
general machinery	0.115	0.015	0.048	0.015	38
special equipments	0.244	0.081	0.169	0.133	96
transportation e	-0.159	-0.025	-0.245	-0.091	45
Electrical machinery	0.368	0.159	0.387	0.394	124
instrumentation	3.832	0.600	0.607	0.225	45

Industrial Heterogeneity. The flexibility of our method allows us to explore the industrial heterogeneity in treatment effects on productivity by calculating the treatment effects by 2-digit industry. Table 2 reports the industrial treatment effects on productivity. Whether we use Cobb-Douglas production function or the Translog production function, the industry of instrumentation and cultural products shows the highest ATT of productivity ($ATT = 3.832$), while the industry of chemical products show the lowest ATT of productivity ($ATT = -0.191$). This detected industrial heterogeneity reflects that firms obtain different productivity gains when introducing AI products. To gauge the importance of each industry in contributing the overall treatment effects of AI products on productivity, we multiply each industry's ATT estimate with the share of the industry's observations in the sample. Column "Contribution" in Table 2 reports the outcome. For the Cobb-Douglas specification, the electrical machinery industry contributes around 8% of the total ATT, ranking as the largest contributor among all eight industries. The relative importance of different industry in affecting the total ATT remains robust if we use the Translog specification. Figure 4 shows the correlation of the relative importance of each industry's ATT between the Cobb-Douglas production function and the Translog production function. We see a strong positive correlation between these two specifications.

Figure 4: Correlation of Contributions in ATT between Cobb-Douglas and Translog Production Function



Structural Change in the Production Function. One may wonder that introducing new products would trigger changes in production lines, management and sales teams. Consider the possible structural changes in the production function, only allowing the treatment variable to affect the productivity fails to capture the change in the production function's parameters. Recall that our econometric framework permits us to separately identify the influence of the treatment on the production function's parameters from that on the productivity. To additionally evaluate the impact of AI products on the production function parameters, we specify the following production function:

$$y_{it} = \beta_t + \beta_l l_{it} + \beta_l^s l_{it} \times AI_{it} + \beta_k k_{it} + \beta_k^s k_{it} \times AI_{it} + \varepsilon_{it}, \tag{31}$$

where β_l^s and β_k^s capture the treatment effects of AI product on labor coefficient and capital coefficeint, separately. We treat the input term $l_{it} \times AI_{it}$ as similar to l_{it}, and $k_{it}\times$ similar to k_{it}. This allows us to invoke the ACF timing assumption to perform the estimation. Table 3 reports the production function parameter estimates for Equation (31). Strikingly, for most industries, we see most of the firms increased their labor shares after introducing AI products. However, the impact of the treatment effects on capital share is more negative. Industries such as general machinery (C71), other electronic products (C57), and electrical machinery (C76) significantly decreased the capital share, and only the industry of instrumentation and cultural products (C78) significantly increased the capital share.

Table 3: Production Function Estimates with Changes in the Parameters

Indus. Code	β_l		β_l^s		β_k		β_k^s		β_t	
C43	0.432	(0.028)	-0.143	(0.044)	0.391	(0.042)	0.088	(0.082)	0.057	(0.023)
C51	0.506	(0.024)	0.052	(0.021)	0.283	(0.020)	-0.014	(0.033)	0.086	(0.024)
C57	0.779	(0.278)	0.123	(0.043)	0.094	(0.033)	-0.054	(0.024)	0.107	(0.040)
C71	0.417	(0.070)	0.481	(0.077)	0.372	(0.061)	-0.175	(0.043)	0.062	(0.016)
C73	0.591	(0.082)	0.012	(0.025)	0.279	(0.059)	0.005	(0.037)	0.062	(0.011)
C75	0.558	(0.051)	0.024	(0.097)	0.363	(0.100)	0.013	(0.117)	0.064	(0.078)
C76	0.458	(0.009)	0.268	(0.006)	0.426	(0.015)	-0.086	(0.029)	0.057	(0.010)
C78	0.643	(0.267)	-0.235	(0.099)	0.183	(0.080)	0.096	(0.043)	0.153	(0.071)

Note: Industry codes classifications are C43: chemical products; C51: electronic components; C57: other electronic products; C71: general machinery; C73: special equipment; C75: transportation equipments; C76: electrical machinery; C78: instrumentation and cultural products. Bootstrapped standard errors are reported in the parenthesis.

Table 4 reports the dynamic treatment effects of introducing AI products on productivity. We find that ATTs in period 0 and period 3 after receiving the treatment are negative but not significant. ATTs in period 1 and 2 are positive with magnitudes approximating our baseline results. Importantly, the positive estimate of the treatment effect on productivity is statistically significant at 5% significance level, indicating that accounting for the functional change in the production function lead to stronger productivity effects from the treatment.

Table 4: Dynamic Treatment Effects: Structural Change in the Production Function

Periods	ATT	95% CI		Obs.
0	-0.111,	(-2.080,	2.002)	255
1	0.142	(-2.038,	2.609)	164
2	1.996	(1.490,	6.738)	115
3	-0.036	(-2.646,	5.096)	40
Total	0.389	-1.401	3.329	574

6 Conclusion

In this paper, we studied the identification and estimation of treatment effects on productivity. We generalize the standard firm-level investment model by incorporating binary treatment which affects the productivity evolution and/or production functions. The treatment reflects either the change in the macro environment or individual action. The treatment effects of productivity is the difference between the realized productivity and the potential outcome of productivity. As the productivity is unobservable to the econometrician, the detection the treatment effects on productivity requires recovering the productivity and its evolution rule. We examine the underlying assumptions that lead to the identification of treatment effects on the structurally estimated productivity. Taking advantage of the Markovian productivity process, we propose a new approach for estimating the full dynamic treatment effects on productivity.

References

Abbring, J. H. and J. J. Heckman (2007). Econometric evaluation of social programs, part III: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation. *Handbook of econometrics 6*, 5145–5303.

Ackerberg, D. A., K. Caves, and G. Frazer (2015). Identification properties of recent production function estimators. *Econometrica 83*(6), 2411–2451.

Amiti, M. and J. Konings (2007). Trade liberalization, intermediate inputs, and productivity: Evidence from indonesia. *American Economic Review 97*(5), 1611–1638.

Athey, S. and G. W. Imbens (2022). Design-based analysis in difference-in-differences settings with staggered adoption. *Journal of Econometrics 226*(1), 62–79.

Aw, B. Y., M. J. Roberts, and D. Y. Xu (2011). R&D Investment, Exporting, and Productivity Dynamics. *American Economic Review 101*(4), 1312–1344.

Braguinsky, S., A. Ohyama, T. Okazaki, and C. Syverson (2015). Acquisitions, productivity, and profitability: evidence from the japanese cotton spinning industry. *American Economic Review 105*(7), 2086–2119.

Brandt, L., J. Van Biesebroeck, L. Wang, and Y. Zhang (2017). Wto accession and performance of chinese manufacturing firms. *American Economic Review 107*(9), 2784–2820.

Brynjolfsson, E., D. Rock, and C. Syverson (2017, November). Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics. Technical Report w24001, National Bureau of Economic Research, Cambridge, MA.

Chen, Y., M. Igami, M. Sawada, and M. Xiao (2021). Privatization and productivity in china. *The RAND Journal of Economics 52*(4), 884–916.

Chen, Z., X. Jin, and X. Xu (2021). Is a corruption crackdown really good for the economy? firm-level evidence from china. *The Journal of Law, Economics, and Organization 37*(2), 314–357.

Chen, Z. and M. Liao (2021). Identification and estimation of production functions for multiproduct firms. *Available at SSRN 3968432*.

De Loecker, J. (2007). Do exports generate higher productivity? evidence from slovenia. *Journal of international economics 73*(1), 69–98.

De Loecker, J. (2013). Detecting learning by exporting. *American Economic Journal: Microeconomics 5*(3), 1–21.

De Loecker, J. and C. Syverson (2021, September). An industrial organization perspective on productivity. Working Paper 29229, National Bureau of Economic Research.

Deryugina, T. (2017). The fiscal cost of hurricanes: Disaster aid versus social insurance. *American Economic Journal: Economic Policy 9*(3), 168–98.

Doraszelski, U. and J. Jaumandreu (2013). R&d and productivity: Estimating endogenous productivity. *Review of Economic Studies 80*(4), 1338–1383.

Gandhi, A., S. Navarro, and D. A. Rivers (2020). On the identification of gross output production functions. *Journal of Political Economy 128*(8), 2973–3016.

Gans, J. S. (2022a). Ai adoption in a competitive market. Technical report, National Bureau of Economic Research.

Gans, J. S. (2022b). Ai adoption in a monopoly market. Technical report, National Bureau of Economic Research.

Gill, R. D. and J. M. Robins (2001). Causal inference for complex longitudinal data: the continuous case. *Annals of Statistics*, 1785–1811.

He, G., S. Wang, and B. Zhang (2020). Watering down environmental regulation in china. *The Quarterly Journal of Economics 135*(4), 2135–2185.

Heckman, J. J. and S. Navarro (2007). Dynamic discrete choice and dynamic treatment effects. *Journal of Econometrics 136*(2), 341–396.

Kim, K. i., Y. Luo, and Y. Su (2019). A robust approach to estimating production functions: Replication of the acf procedure. *Journal of Applied Econometrics 34*(4), 612–619.

Levinsohn, J. and A. Petrin (2003). Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies 70*(2), 317–341.

Liu, Y. and J. Mao (2019). How do tax incentives affect investment and productivity? firm-level evidence from china. *American Economic Journal: Economic Policy 11*(3), 261–91.

Lok, J. J. (2008). Statistical modeling of causal effects in continuous time. *The Annals of Statistics 36*(3), 1464–1507.

Olley, G. S. and A. Pakes (1996). The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica 64*(6), 1263–1297.

Pavcnik, N. (2003). What explains skill upgrading in less developed countries? *Journal of Development Economics 71*(2), 311–328.

Peters, B., M. J. Roberts, V. A. Vuong, and H. Fryges (2017). Estimating dynamic R&D choice: an analysis of costs and longrun benefits. *The RAND Journal of Economics 48*(2), 409–437.

Robins, J. M. (1997). Causal inference from complex longitudinal data. In *Latent variable modeling and applications to causality*, pp. 69–117. Springer.

Rovigatti, G. and V. Mollisi (2018). Theory and practice of total-factor productivity estimation: The control function approach using stata. *The Stata Journal 18*(3), 618–662.

Sun, L. and S. Abraham (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics 225*(2), 175–199.

Vikström, J., G. Ridder, and M. Weidner (2018). Bounds on treatment effects on transitions. *Journal of Econometrics 205*(2), 448–469.

Wooldridge, J. M. (2009). On estimating firm-level production functions using proxy variables to control for unobservables. *Economics Letters 104*(3), 112–114.

Yu, M. (2015). Processing trade, tariff reductions and firm productivity: Evidence from chinese firms. *The Economic Journal 125*(585), 943–988.

Appendices

A Data Appendix

A.1 Variables Construction

We construct the main variables for production function estimation as follows.

Materials: Costs of goods sold plus selling, general and administrative expenses minus labor costs. Labor costs are measured using the payroll payable, deflated using industry-year level input price index.

Capital: Fixed assets including property, plant, and equipment (PP&E) deflated by province-year level investment price index.

Labor: Total number of registered working employees reported in the annual report.

Value Added: Operational revenue minus materials, deflated by province-year level output price index.

Annual Sales: Total operational revenue, deflated province-year level output price index.

All the price indices are extracted from China's Statistical Yearbook. The summary statistics of these variables are displayed in the following table.

Table A.1: Summary Statistics of Main Variables

Variables	mean	s.d.	p5	p25	p50	p75	p95
m	21.02	1.21	19.19	20.13	20.93	21.79	23.19
l	7.60	1.01	6.05	6.88	7.55	8.26	9.35
k	19.84	1.24	17.91	19.00	19.75	20.61	21.98
y	18.85	1.23	16.90	18.04	18.81	19.63	20.99
$log(sale)$	21.06	1.18	19.31	20.20	20.98	21.82	23.17

The industrial classification is based on the two-digit China's National Industrial Classification. We choose the manufacturing industries and perform the estimation by 2-digit industry. We drop some industries which contain too few observations to conduct meaningful analysis or contain few treated observation. The final sample of industries and number of observations for treated and control groups are listed as follows.

Table A.2: Number of Treated and Untreated Observations for AI-intensive Industries

Indus. Code	Industry	untreated	treated	Total
C43	chemical products	1,944	49	1,993
C51	electronic components	1,048	132	1,180
C57	other electronic products	293	108	401
C71	general machinery	939	52	991
C73	special equipment	1,464	146	1,610
C75	transportation equipment	1,225	56	1,281
C76	electrical machinery	1,396	133	1,529
C78	instrumentation	283	46	329
Total		8,592	722	9,314

A.2 Empirical Results using the Original Sample

Figure A.1: AI Words Frequency in the Original CSMAR Database

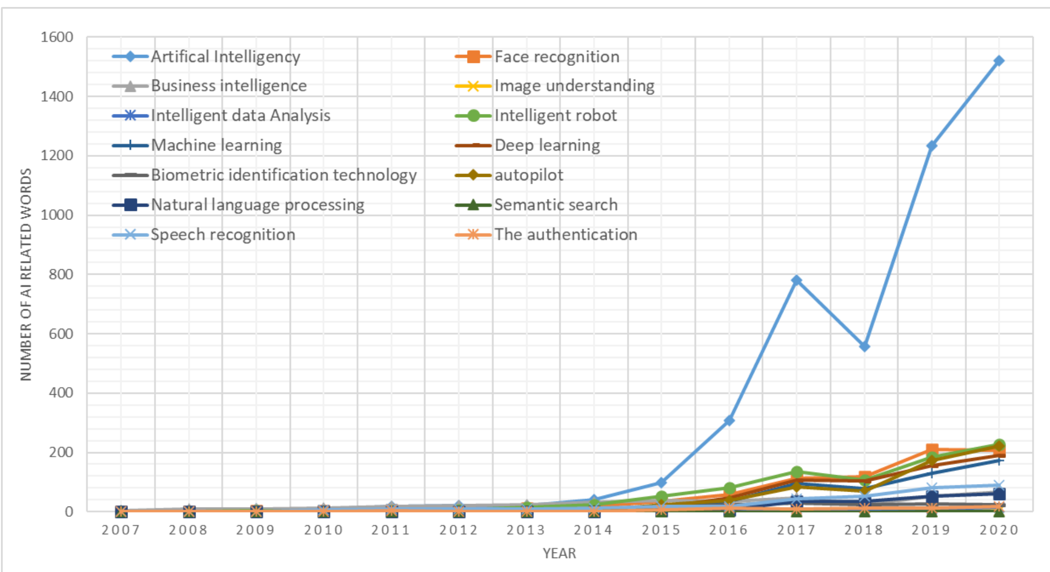


Table A.3: Treatment Effects of AI on Productivity: Original AI-product Definition

	Cobb-Douglas			Translog			
Periods	ATT	95% CI		ATT	95% CI		Obs.
0	0.090	(-1.631,	1.782)	0.035	(-1.717,	1.718)	625
1	0.221	(-1.678,	2.220)	0.160	(-1.819,	2.236)	419
2	0.212	(-1.797,	2.455)	0.113	(-1.988,	2.425)	313
3	0.201	(-2.034,	2.630)	0.060	(-2.293,	2.479)	150
4	0.094	(-2.065,	2.567)	-0.061	(-2.372,	2.377)	93
Total	0.159	(-1.784,	2.183)	0.080	(-1.932,	2.145)	1600

Note: the 95% confidence intervals are obtained by bootstrapping 300 times.

B Monte-Carlo Setup

The Monte-Carlo setup is similar to Ackerberg et al. (2015) to a large extend except that we consider an extended productivity process with exogenous policy shocks. We consider a panel of 1000 firms over T (T=3, 5, 10) periods to gauge the performance of our method against alternative approaches. The parameters are chosen to match the key aspects of the Chilean data. In the description, we focus on the productivity process and the implied choice of inputs.

B.1 (Conditionally) Exogenous Productivity Process with Policy Interventions

B.1.1 Production Function and Potential Productivity Shocks

The production function is Leontief in the material input, i.e.

$$Y_{it} = \min\left\{\beta_0 K_{it}^{\beta_k} L_{it}^{\beta_l} e^{\omega_{it}}, \beta_m M_{it}\right\} \tag{B.1}$$

We choose $\beta_0 = 1, \beta_k = 0.4, \beta_l = 0.6$ and $\beta_m = 1$. The variable for the binary policy shock is $D_{it} \in \{0, 1\}$. The potential productivity shocks $\omega_{it}^d, d \in \{0, 1\}$, follow the following AR(1) process:

$$\omega_{it}^0 = \rho_0 \omega_{it-1}^0 + \epsilon_{it}^0, \tag{B.2}$$

$$\omega_{it}^1 = \rho_1 \omega_{it-1}^1 + \gamma + \epsilon_{it}^1, \tag{B.3}$$

where $d = 1$ refers to the productivity process of treated units after they received the treatment and $d = 0$ for the untreated units. We choose $\rho_0 = 0.7, \rho_1 = 0.8, \gamma = 0.2$, and the exogenous shocks $(\epsilon_{it}^0, \epsilon_{it}^1) \sim \mathbf{N}(0, 0.3\mathbb{I})$.

B.1.2 (Conditionally) Exogenous Policy Shocks

There is an exogenous policy shock that is captured by $D_{it} \in \{0, 1\}$. The policy shock can be either purely exogenous or exogenous conditional on the firm's current productivity process. The realized productivity is therefore:

$$\omega_{it} = \omega_{it}^1 D_{it} + \omega_{it}^0 (1 - D_{it}) \tag{B.4}$$

The policy shock arrives at $t_0 - b$, where $1 \leq t_0 < T$ and $b \in [-1, 0]$. The timing of labor choice and the policy intervention is more intricate, and we will discuss them later. We distinguish the purely exogenous policy and conditionally exogenous policy. For the *purely exogenous* policy shock, we randomly assign 50% firms to be treated by the policy. While for the *conditionally exogenous* policy, we set a cutoff value of the productivity $\bar{\omega}$ such that only firms whose productivity level $\omega_{it_0-1} > \bar{\omega}$ are exposed to the policy shock. This simple selection criterion captures a large class of models in which there is a strict sorting pattern for the considered firm decision and firm productivity.

Before the arrival of policy shock, the productivity evolution is just (B.2). Upon the arrival of the policy, we follow Ackerberg et al. (2015) and think of decomposing the productivity evolution into two sub-processes. First, ω_{it_0-1} evolves to ω_{it_0-b}; the evolution rule is about to be specified later. After the policy shock, ω_{it_0-b} evolves to ω_{it0}. We use the following model of the evolution of ω between sub-periods:

$$\omega_{it_0} = \rho_1^b \omega_{it_0-b} + b \times \gamma + \sqrt{1-\rho_1^{2b}}\epsilon_{it}^1 \tag{B.5}$$

Thus, when $b = 0$, firms receive the policy shocks at t_0, which triggers the evolution of productivity to switch to the regime of treated. When $b \in (0, 1)$, the regime switching is happening between $t_0 - 1$ and t_0, and the evolution of productivity from t_0 to t_1 is a mixture of the above two productivity processes. The term $b \times \gamma$ reflects the treatment effects on the level of productivity for a duration of b.[27] After t_0, the productivity evolution is rendered to be (B.3).

B.1.3 Choice of Labor and Material Inputs

The choice of labor and material inputs are static. There are firm specific wage shocks. The logged wage for firm i follows an AR(1) process:

$$\ln(W_{it}) = 0.3\ln(W_{it-1}) + \xi_{it}^W \tag{B.6}$$

The variance of ξ_{it}^W and $\ln(W_{i0}$ is chosen such that the standard deviation of $\ln(W_{it})$ is equal to 0.1.

In all periods except between $t_0 - 1$ to t_0, we follow DGP1 in Ackerberg et al. (2015) in assuming the timing of choosing inputs, except that we additionally consider the timing of policy intervention. That is, we assume that during this period, labor is chosen at time $t_0 - b_l$. The productivity process is decomposed two sub-processes. First, productivity

[27] The assumption that the treatment effects is cumulative is consistent with the evolution rule introduced by (B.3)

evolves to $\omega_{it_0-b_l}$ at which point the firm chooses labor input. Then, after L_{it} is chosen, $\omega_{it_0-b_l}$ evolves to ω_{it_0}. In all periods but t_0-1, the following model is to characterize the evolution of ω_{it}^0 and ω_{it}^1 between sub-periods in time periods:

$$\omega_{it-b_l}(d) = \rho_d^{1-b_l}\omega_{it-1}(d) + \frac{1-b_l}{\rho_d^{b_l}} \times \gamma \times d + \underbrace{\sqrt{1-\rho_d^{2-2b_l}}\epsilon_{it-b_l}^d}_{\epsilon_{it}^d(B)}, \text{ for } d \in \{0,1\},\, t \neq t_0 \tag{B.7}$$

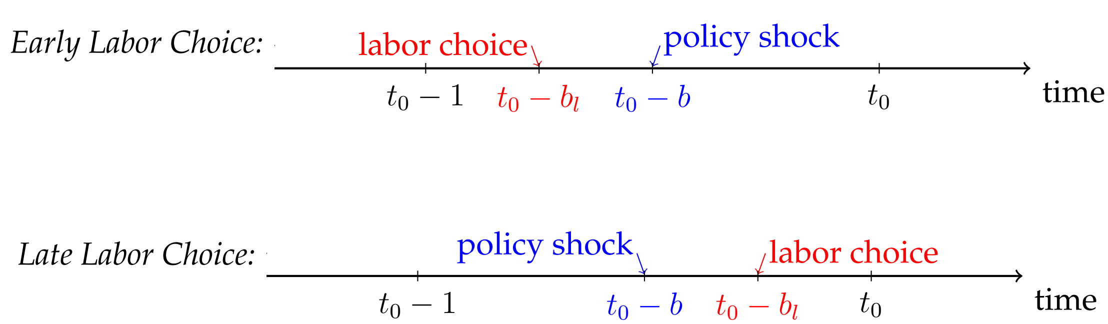
$$\omega_{it}^d = \rho_d^{b_l}\omega_{it-b_l}(d) + b_l \times \gamma \times d + \underbrace{\sqrt{1-\rho_d^{2b_l}}\epsilon_{it}^d}_{\epsilon_{it}^d(A)}, \text{ for } d \in \{0,1\},\, t \neq t_0 \tag{B.8}$$

Combining (B.7) with (B.8), we obtain the evolution rule of productivity:

$$\omega_{it}^d = \rho_d\omega_{it-1}(d) + \gamma \times d + \rho_d^{b_l}\sqrt{1-\rho_d^{2-2b_l}}\epsilon_{it-b_l}^d + \sqrt{1-\rho_d^{2b_l}}\epsilon_{it}^d,\; t \neq t_0 \tag{B.9}$$

Note that Equation (B.9) is consistent with the AR(1) coefficient in (B.2) and (B.3) because $\rho_d^{1-b}\rho_d^b = \rho_d$. The variance of ω_{it-b_l} is constant over time if we impose that $Var(\epsilon_{it-b_l}^d) = Var(\epsilon_{it}^d)$ so that $Var(\rho_d^{b_l}\sqrt{1-\rho_d^{2-2b_l}}\epsilon_{it-b_l}^d + \sqrt{1-\rho_d^{2b_l}}\epsilon_{it}^d) = Var(\epsilon_{it}^d)$. If $b = 0$, then the above evolution rule also applies to the period between t_0-1 and t_0. If $b \in (0,1)$, The timing of labor choice during the arrival period of the policy shock is more subtle. We need to consider two cases (See Figure B.1): first, $b_l \geq b$ such that the labor is chosen no later than the arrival of policy; second, $b_l < b$ so that labor is chosen later than the policy shock's arrival.

Figure B.1: Relative Timing of Labor Choice and Policy Intervention Between t_0-1 and t_0



Early labor choice: $b_l \geq b$. In this case, before the policy's arrival, the evolution of productivity from $t_0 - 1$ to $t_0 - b$ breaks into two stages:

$$\omega_{it_0-b_l}(0) = \rho_0^{1-b_l}\omega_{it_0-1}(0) + \sqrt{1-\rho_0^{2-2b_l}}\epsilon^0_{it_0-b_l} \tag{B.10}$$

$$\omega_{it_0-b}(0) = \rho_0^{b_l-b}\omega_{it_0-b_l}(0) + \sqrt{1-\rho_0^{2b_l-2b}}\epsilon^0_{it_0-b} \tag{B.11}$$

This guarantees that the variance of productivity is constant. Then the productivity evolves to t_0 according to the following process:

$$Treated: \quad \omega_{it_0}(1) = \rho_1^b\omega_{it_0-b}(0) + b\times\gamma + \sqrt{1-\rho_1^{2b}}\epsilon^1_{it} \tag{B.12}$$

$$Non-treated: \quad \omega_{it_0}(0) = \rho_0^b\omega_{it_0-b}(0) + \sqrt{1-\rho_0^{2b}}\epsilon^0_{it} \tag{B.13}$$

In this case, the firm's labor choice will only be adjusted after t_0.

Late labor choice: $b_l < b$. At the time of policy's arrival, the productivity has evolved to ω_{it_0-b} according to the following process:

$$\omega_{it_0-b}(0) = \rho_0^{1-b}\omega_{it_0-1}(0) + \sqrt{1-\rho_0^{2(1-b)}}\epsilon^0_{it} \tag{B.14}$$

The evolution of productivity of the treated units from $t_0 - b$ to t_0 could further break into two sub-processes:

$$\omega_{it_0-b_l}(1) = \rho_1^{b-b_l}\omega_{it_0-b}(0) + \frac{b-b_l}{\rho_1^{b_l}}\times\gamma + \sqrt{1-\rho_1^{2b-2b_l}}\epsilon^1_{it} \tag{B.15}$$

$$\omega_{it_0}(1) = \rho_1^{b_l}\omega_{it_0-b_l}(1) + b_l\times\gamma + \sqrt{1-\rho_1^{2b_l}}\epsilon^1_{it} \tag{B.16}$$

Firms do not anticipate the arrival of policy. For untreated units, the optimal labor choice would always be

$$L_{it} = \theta_{it}K_{it}^{\frac{\beta_k}{1-\beta_l}}e^{(1-\beta_l)^{-1}\left(\rho_0^{b_l}\omega_{it-b_l}+\frac{1-\rho_0^{2b_l}}{2}\sigma^2_{\epsilon^0}\right)}. \tag{B.17}$$

But after the arrival of the policy, treated firms commit its labor choice to the new productivity process. Therefore, the optimal labor choice for treated units are given by:

$$
L_{it} = \begin{cases} \theta_{it} K_{it}^{\frac{\beta_k}{1-\beta_l}} e^{(1-\beta_l)^{-1}\left(\rho_0^{b_l}\omega_{it-b_l} + \frac{1-\rho_0^{2b_l}}{2}\sigma_{\epsilon 0}^2\right)}, & if \quad t \leq t_0 - 1 \\ \theta_{it} K_{it}^{\frac{\beta_k}{1-\beta_l}} e^{(1-\beta_l)^{-1}\left(\rho_0^{b_l}\omega_{it-b_l} + \frac{1-\rho_0^{2b_l}}{2}\sigma_{\epsilon 0}^2\right)}, & if \quad t = t_0 \quad and \quad b_l \geq b \\ \hat{\theta}_{it} K_{it}^{\frac{\beta_k}{1-\beta_l}} e^{(1-\beta_l)^{-1}\left(\rho_1^{b_l}\hat{\omega}_{it-b_l} + \frac{1-\rho_1^{2b_l}}{2}\sigma_{\epsilon 0}^2\right)}, & otherwise \end{cases} \tag{B.18}
$$

where $\theta_{it} \equiv \beta_0^{\frac{1}{1-\beta_l}} \beta_l^{\frac{1}{1-\beta_l}} W_{it}^{\frac{-1}{1-\beta_l}}$, $\hat{\theta}_{it} \equiv (\beta_0 e^{\frac{\gamma b_l}{1-\rho_1^{b_l}}})^{\frac{1}{1-\beta_l}} \beta_l^{\frac{1}{1-\beta_l}} W_{it}^{\frac{-1}{1-\beta_l}}$, and $\hat{\omega}_{it-b_l} \equiv \omega_{it-b_l} - \frac{\gamma b_l}{1-\rho_1^{b_l}}$.[28]

B.1.4 Investment Choice and Steady State

Capital is a dynamic input, which is accumulated through investment according to

$$
K_{it} = (1-\delta)K_{it-1} + I_{it-1},
$$

where the depreciation rate is $\delta = 0.2$. Investment is chosen at the initial time of each period. The adjustment costs in investment are given by

$$
c_i(I_{it}) = \frac{\phi_i}{2} I_{it}^2
$$

where $\phi_i/2$ is distributed lognormally across firms (but constant over time) with standard deviation 0.6. In the presence of policy shock, the investment rule can be characterized by two regimes: (1) ex-ante regime where the policy shock has not been introduced and (2) ex-post regime where the policy has come to effect. Specifically, for treated firms, if $t \leq t_0 - 1$, the investment rule is given by:

[28] This is because, after the policy shock, we can re-write the productivity's evolution equation as: $\hat{\omega}_{it_0}(1) = \rho_1^{b_l}\hat{\omega}_{it-b_l} + \sqrt{1-\rho_1^{2b_l}}\epsilon_{it}^1$.

$$
\begin{aligned}
I_{it}(t \leq t_0 - 1) =& \frac{\beta}{\phi_i}\left(\frac{\beta_k}{1-\beta_l}\right)\beta_0^{\frac{1}{1-\beta_l}} \times \left(\beta_l^{\frac{\beta_l}{1-\beta_l}} - \beta_l^{\frac{1}{1-\beta_l}}\right) \\
& \sum_{\tau=0}^{\infty}[\beta(1-\delta)]^{\tau} \times \exp\Big\{\Big[\frac{\rho_0^{\tau+1}\omega_{it}}{1-\beta_l} - \frac{\beta_l\rho_W^{\tau+1}\ln(W_{it})}{1-\beta_l} \\
& + \frac{1}{2}\left(\frac{\beta_l}{1-\beta_l}\right)^2 \sigma_{\xi W}^2 \sum_{s=0}^{\tau}\rho_W^{2(\tau-s)} + \frac{1}{2}\left(\frac{1}{1-\beta_l}\right)^2 \rho^{2b_l}\Big(\rho_0^{2\tau}\sigma_{\epsilon^0 A}^2 \\
& + \sum_{s=1}^{\tau}\rho_0^{2(\tau-s)\sigma_{\epsilon^0}^2}\Big) + \frac{\sigma_{\epsilon^0(B)}^2}{2(1-\beta_l)}\Big]\Big\}.
\end{aligned} \tag{B.19}
$$

The above investment rule applies to all periods for the untreated units. Note that after the arrival of policy shock, there is a structural change in the productivity evolution. The new productivity process is known by the firm. To obtain the investment rule, we can rewrite the productivity process (B.3) as:

$$
\underbrace{\omega_{it}^1 - \frac{\gamma}{1-\rho_1}}_{\tilde{\omega}_{it}} = \rho_1 \underbrace{\left(\omega_{it-1}^1 - \frac{\gamma}{1-\rho_1}\right)}_{\tilde{\omega}_{it-1}} + \epsilon_{it}^1 \tag{B.20}
$$

We can also re-write the production function as:

$$
Y_{it} = \min\{\underbrace{\beta_0 e^{\frac{\gamma}{1-\rho_1}}}_{\tilde{\beta}_0} K_{it}^{\beta_k} L_{it}^{\beta_l} e^{\tilde{\omega}_{it}}, \beta_m M_{it}\} \tag{B.21}
$$

As a result, we can obtain the investment rule for $t \geq t_0$ from (B.19) by replacing β_0 , ω_{it}, ρ_0, and $\sigma_{\epsilon^0}^2$ with $\tilde{\beta}_0$, $\tilde{\omega}_{it}$, ρ_1, and $\sigma_{\epsilon(1)}^2$, respectively:

$$
\begin{aligned}
I_{it}(t \geq t_0) =& \frac{\beta}{\phi_i}\sum_{\tau=0}^{\infty}[\beta(1-\delta)]^{\tau}\left(\frac{\beta_k}{1-\beta_l}\right)\tilde{\beta}_0^{\frac{1}{1-\beta_l}} \times \left(\beta_l^{\frac{\beta_l}{1-\beta_l}} - \beta_l^{\frac{1}{1-\beta_l}}\right) \\
& \times \exp\Big\{\Big[\frac{\rho_1^{\tau+1}\tilde{\omega}_{it}}{1-\beta_l} - \frac{\beta_l\rho_W^{\tau+1}\ln(W_{it})}{1-\beta_l} \\
& + \frac{1}{2}\left(\frac{\beta_l}{1-\beta_l}\right)^2 \sigma_{\xi W}^2 \sum_{s=0}^{\tau}\rho_W^{2(\tau-s)} + \frac{1}{2}\left(\frac{1}{1-\beta_l}\right)^2 \rho^{2b_l}\Big(\rho_1^{2\tau}\frac{1-b_l}{\rho_1^{2b_l}}\sigma_{\epsilon(1)}^2 \\
& + \sum_{s=1}^{\tau}\rho_1^{2(\tau-s)\sigma_{\epsilon^0}^2}\Big) + \frac{\sigma_{\epsilon^0}^2}{2(1-\beta_l)}\Big]\Big\}.
\end{aligned} \tag{B.22}
$$

To avoid the influence of initial distribution of capital stocks and other variables (ω_{i0},

ϕ_i, $\ln(W_{i0})$), we simulate the firm model forward in the absence of policy intervention and take $t_0 - 1$ periods ($t_0 = 2, 3, 6$ respectively for $T = 3, 5, 10$) of the data from the steady state. We then introduce the policy shock and re-solve the input decisions using the new productivity process to obtain the data for the rest of periods. The key difference between our simulation and that obtained from Ackerberg et al. (2015) is that our simulated data contain the transition periods from the original steady state to the new steady state. Therefore, our simulation reflects the pattern of data in the midst of policy change, which is usually the case for empirical studies on policy evaluations. We display the parameter values and targeted moments in the following table.

Table B.1: Summary of Parameter Values

Parameters	Value	Source	Targeted Moments
β_0	1	ACF (2016)	n.a.
β_k	0.4	ACF (2016)	n.a.
β_l	0.6	ACF (2016)	n.a.
β_m	1	ACF (2016)	n.a.
β	0.95	ACF (2016)	n.a.
ρ_0	0.7	ACF (2016)	n.a.
ρ_1	0.8	Our Choice	n.a.
γ	0.2	Our Choice	n.a.
σ_{ϵ^0}	0.3	ACF (2016)	$std(\omega_{it}) = 0.3$
$\sigma_{\epsilon(1)}$	0.3	ACF (2016)	$std(\omega_{it}) = 0.3$
b_l	0.5	ACF (2016)	n.a.
b	$\{0, 0.4, 0.6\}$	Our Choice	n.a.
σ_{ξ^W}	0.1	ACF (2016)	$std(W_{it}) = 0.1$
δ	0.2	ACF (2016)	n.a.
ϕ_i	$\ln(\phi_i) \sim \mathbf{N}(0, 0.36)$	ACF (2016)	n.a.
T/t_0	10/5	Our choice	n.a.

C Estimation

The key to the estimation strategy is to consider that the productivity evolves differently during the transitioning period. Define $Post_t$ to be the indicator for periods after the arrival of policy shock. Therefore, $Post_t = 1$ if $t \geq t_0$. Define Mid_t to be the indicator for whether the firm is during the first period of receiving treatment, i.e., $Mid_t = 1$ if $t = t_0 - 1$ and zero otherwise for all firms. Lastly, we define $Treat_i$ as a firm indicator for whether the firm is eventually treated in the sample: $Treat_i = 1$ means firm i belongs to the treated group. According to our definition, we have $D_{it} = Treat_i \times Post_t$, which is a time-varying indicator for the firm's treatment status. We use an extended Diff-in-Diff equation to model the productivity process as follows:

$$\begin{aligned}\omega_{it} =& h_0(\omega_{it-1}) + h_1(Treat_i \times Post_t \times \omega_{it-1}) + h_2(Treat_i \times Mid_t \times \omega_{it-1}) \\ &+ \gamma_1 Treat_i \times Post_t + \gamma_2 Treat_i \times Mid_t + \eta_{it},\end{aligned} \tag{C.1}$$

where η_{it} is the error with a mean of zero. We approximate $h_0(\omega_{it-1})$ using a linear equation of ω_{it-1}:

$$\begin{aligned}\omega_{it} =& \rho_1 \omega_{it-1} + \theta_1(Treat_i \times Post_t \times \omega_{it-1}) + \theta_2(Treat_i \times Mid_t \times \omega_{it-1}) \\ &+ \gamma_1 Treat_i \times Post_t + \gamma_2 Treat_i \times Mid_t + \eta_{it},\end{aligned} \tag{C.2}$$

To estimate the treatment effects of productivity, the productivity process is incorporated into the standard production function estimation methods including Levinsohn and Petrin (2003), Wooldridge (2009), and Ackerberg et al. (2015). For the choice of instruments and algorithms, we refer to Kim et al. (2019) to choose instruments and conduct the estimation using Stata. To compare our methods with the traditional methods, we also perform estimation using the following productivity process without considering the transitioning period:

$$\omega_{it} = \rho_1 \omega_{it-1} + \theta_1(Treat_i \times Post_t \times \omega_{it-1}) + \gamma_1 Treat_i \times Post_t + \eta_{it}, \tag{C.3}$$

D Monte Carlo Simulations

We consider two scenarios: purely exogenous policy (See Assumption **??**) and conditional exogenous policy (See Assumption 3.2). The timing assumption on input choices in the Monte-Carlo setup is similar to DGP1 in Ackerberg et al. (2015) to a large extent. Critically, our setting differs from theirs by considering policy interventions that incur regime-switching in the productivity's evolution. We generate a balanced panel consisting of 1000 firms spanning over 10 periods. During the fourth period, we introduce a policy shock that is unanticipated by the firm, and keep track of firms for five more periods. We call the time between the triggering of policy shock and the beginning of the next period as the transition period. And we naturally assume that the policy impact during the transition period is different from subsequent periods. We repeat the experiments by generating 1,000 datasets and analyze the estimation outcomes.

D.1 Strictly Exogenous Policy

In our first Monte Carlo experiment, we consider firms encounter a strictly exogenous policy shock during the fourth period. Firms choose capital investment I_{it} at the beginning of each period on observing productivity ω_{it-1}, while hire labor L_{it} at $t - b_l$ (b_l is set to be 0.5) when observing ω_{it-b}. We introduce a policy intervention at time $t_0 - b$ (we set $b = 0.2 < b_l$) after the choice of L_{it_0}. This policy shock is not anticipated by firms, therefore firms' choices of inputs are not affected by the policy before period t_0. After the realization of the policy shock, firms take it into account when making choices of investment, labor, and capital. The policy shock generates a transitional path for the productivity until it reaches a new steady state. Details of the DGP and estimation are explained in the Appendix.

We use three methods to estimate the productivity. The first is the ex-post method, which ignores the policy intervention in the productivity process when estimating the production function, but accounts for the policy shock in the regression after obtaining productivity estimates. The second approach is to consider the policy shock in the evolution process, but the variable indicating the policy is imperfect. Specifically, the transition right after policy shock is not separately controlled for. Lastly, we consider estimating the production function with a perfect control for the possible timing varying effect of the policy shock on productivity's evolution. These three methods only differ by the productivity evolution process; We use the ACF approach to estimate the production function.[29]

Table D.3 about here

Figure D.3 displays the distribution of production function parameter estimates by using different methods. The true values for labor elasticity (β_l) and capital elasticity (β_k) are 0.6 and 0.4, respectively. We find that estimates using the "ex-post" method results concentrates on values that are very different from the true values, implying substantial mistakes in productivity estimates. The reason is that the policy has altered firms' labor and capital choices since its initiation. However, in the ex-post method, the policy variable is not controlled in the evolution process, and hence it is contained in the error term. This causes a correlation between the productivity shocks and input choices. Given our specific parameterization, the correlation is negative for labor and positive for capital, which generates downward (*resp.*) biased estimate for labor (*resp.* capital). For a similar reason, when the regime-switching is not controlled (See Panels titled "no-transition"), we see same directions of bias in the labor and capital coefficients' estimates. Lastly,

[29] We refer to Kim et al. (2019) to add lagged capital and constant in the instrument set to avoid "spurious" minimization problem.

for the productivity process with perfect control of the transitioning effect of the policy, we obtain the most accurate production function estimates that center around the true value with a density function close to the normal distribution. These results indicate that despite that the policy is exogenous, failing to control the policy impact in the production function estimation would lead to biases in the production function estimation.

In Figure D.4, we display the empirical distribution of the logged correlation coefficients for different productivity estimates. The productivity estimates obtained without considering the regime-switching period is highly correlated with the true productivity, so as the productivity estimates with transitional period included. The average simple correlation coefficient for both of them is one for treated and non-treated units. We find that the bias of the productivity estimates is more serious for the treated units for the ex-post estimation method, with an average correlation coefficient to be 0.978 for both treated and non-treated units.

Figure D.4 about here

As we have illustrated, the parameters of the productivity's persistence in the absence of policy, i.e., $h_0(\cdot)$, is the key to understanding the policy's treatment effects on productivity. In Table D.1, we report the regression outcome of the productivity process using the productivity estimates. Table D.1 Column (1) reports the regression outcome using the true productivity estimates, which serves as the baseline result. It shows that the coefficient of current productivity ω_t is 0.719 for the untreated observations, standing close to the true value 0.7. We add dummies indicating the treated group and the periods post to the happening of policy shock. Importantly, we also add a dummy Mid_t to capture the difference of the productivity's evolutionary dynamics during the transitioning period. The estimation results show that the policy has increased the productivity's persistence by 0.081 (vs. true value 0.1), and the level by 0.200 (vs. true value 0.200). Column (2) in Table D.1 shows the estimation results for the productivity estimates obtained by assuming exogenous productivity process. The estimate of the productivity's persistence is 0.742. Also, the estimate of the impact of the policy shock on the changes in the persistence is 0.103, which is higher than that in the baseline group. Moreover, the ex-post regression tend to result in a larger positive impact of the policy shock on the level of productivity. Table D.1 Column (3) reports the estimation results without considering the possible transitioning dynamics of the productivity process triggered by the exogenous policy shock. The productivity's persistence parameter is 0.719, which is the same as that in the benchmark group. This means that despite the full structural change in the productivity process is not accounted for, this method can still generate a reliable estimate for the productivity's persistence. However, it does not deliver accurate estimates for the policy-induced changes in levels and persistence of productivity. The last column in Table D.1 show that the estimates are quite close to the baseline results using the true productivity. Given the high correlation between the productivity estimates, it is not a surprise to see that considering the structural changes in the productivity would generate more accurate estimates for both the productivity and the productivity's evolutionary process.

Table D.1 about here

To evaluate how different methods could lead to different outcomes in estimating the ATTs on productivity. We estimate ATT by period using our proposed method. That is, we simulate the potential outcome productivity using the productivity's persistence parameter and the recovered distribution of productivity shocks for the untreated units.[30] Then the treatment effect of the policy on productivity for any period is calculated by take the difference between the observed productivity and the simulated potential productivity. In Figure D.5 Panel (a), we show the estimates of ATTs using our proposed structural estimates. We find that including the policy shock in the productivity process and considering the regime-switching period leads to estimates very close to the ATT estimates using the true productivity. If the regime-switching period is not considered, the ATTs are underestimated. This is because the productivity jumps during the transition period is not taken into account when comparing between the treated firms and the non-treated firms. Moreover, the ex-post method systematically lead to much smaller ATT estimates. In first two periods after the policy shock, the ex-post method leads to negative estimates for the positive ATTs. Figure D.5 Panel (b) displays the estimation results using conventional event study designs without including the lagged productivity as regressor. Ignoring the dynamic feature of the productivity's evolution equation leads to quite different estimates of ATTs. On average, the estimated treatment effects are larger and the growing trend is more pronounced.

Figure D.5 about here

D.2 Endogenous Firm-level Action

We continue to analyze the estimation results for the datasets generated by conditionally exogenous policy.[31]

[30]This is different from the traditional event studies in which the outcome variable usually does not depend on the its past realizations.

[31]We find that when we consider a full control for the policy effect, the production function estimates have two distinctive modes. We thus drop the transition period to estimate the productivity process with

Figure D.6 reports the histograms of the estimates for capital and labor coefficients in the production function. The ex-post method still generates biased estimates for β_l and β_k. To a large extent, the estimates are close to the case when the policy shock is purely exogenous. This is because the ex-post method does not take the policy shock into account when incorporating the productivity process into the estimation procedure. The reason for the directions in biases is similar to that for the purely exogenous policy shock. For the approach without considering the transition, we now see more biases in the estimates of labor coefficient compared to the case of purely exogenous policy. This is because the conditionally endogenous policy may generate a correlation between the productivity shocks and the lagged labor choice during the transitioning period, which bias the labor coefficient downward. The estimation approach which drop the transitioning period generate reasonable production function estimates that center around the true value.

Figure D.6 about here

Figure D.7 shows the distribution of logged correlation coefficient for different pairs of productivity. Both for the treated and untreated units, we find that the estimation approach which accounts for the full structural change in the productivity's evolution performs the best. Compared to the case of purely exogenous policy, the approach that does not consider the structural change during the transitioning period leads to more biased productivity estimates for untreated observations. Still, the ex-post method performs the worst in recovering the productivity.

Figure D.7 about here

We report the regression outcome for the productivity's evolution process in Table D.2. Table D.2 Column (1) is the baseline result using true productivity. We find that the coefficient for ω_t and $Treat_t \times Post_t$ are close to the true values in DGP. In Column (4), we see that the coefficient estimates obtained from controlling the full structural change in the productivity evolution generates very similar estimates to the baseline result using true productivity in the regression.

Table D.2 about here

Figure D.8 shows the estimated ATTs using different methods. It is clear that considering the structural change in the productivity's evolution process leads to the most accurate estimates of the ATT for each period. And the ex-post method leads to the least

fewer coefficients.

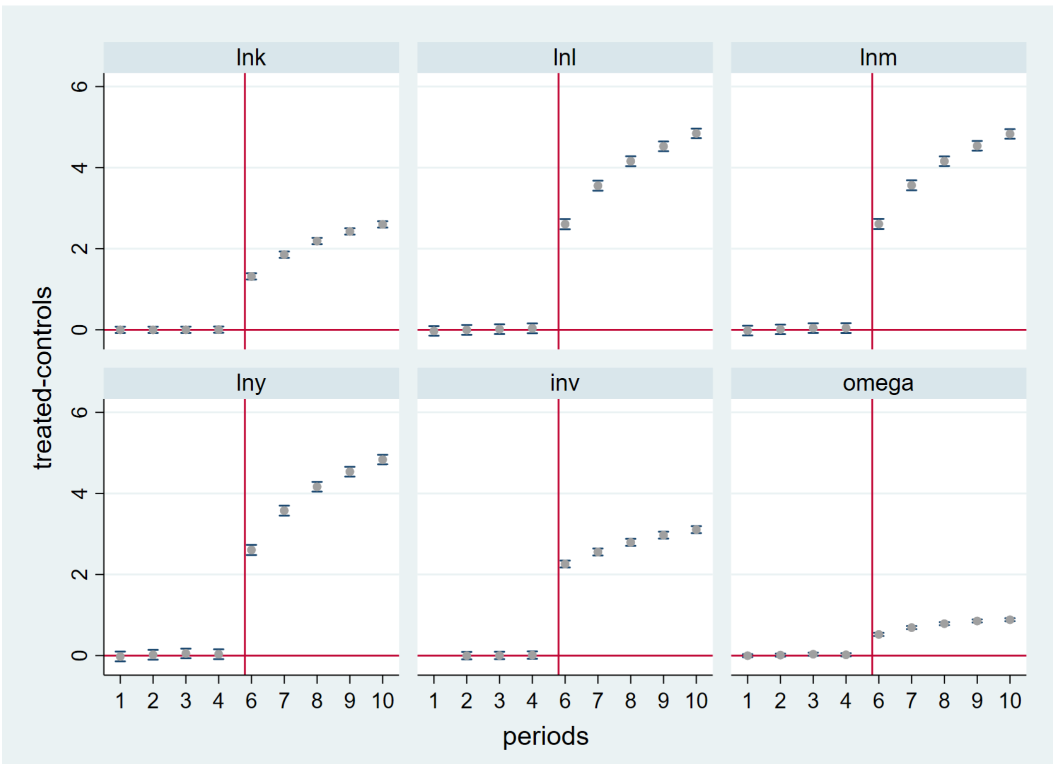
accurate estimates for ATTs. But in the case of conditional exogenous policy, failing to consider the structural change in the productivity's evolution process leads to an upward bias in evaluating the treatment effects on productivity.

Figure D.8 about here

D.2.1 Auxiliary Simulation Results

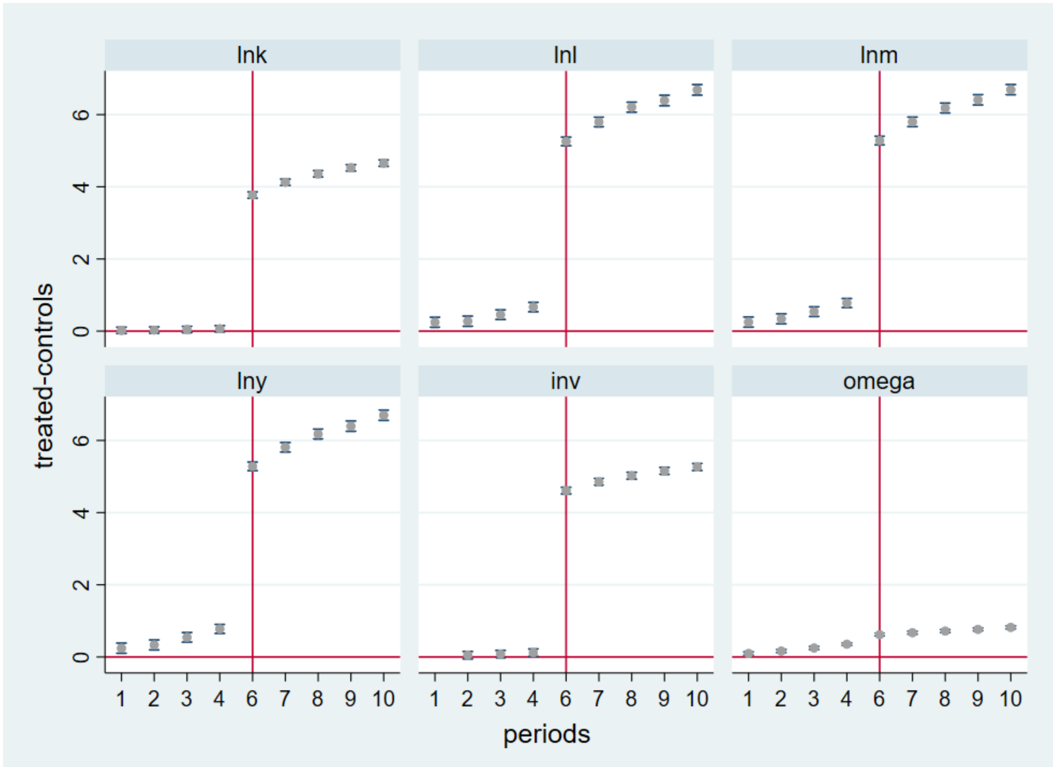
Figure D.1 and Figure D.2 plot the pattern of simulated data for strictly exogenous policy and conditional exogenous policy, respectively.

Figure D.1: Data Pattern for Strictly Exogenous Policy Shock



Note: We conduct an event study by regression the variables on a full set of time dummies and its interaction with treatment group indicator.

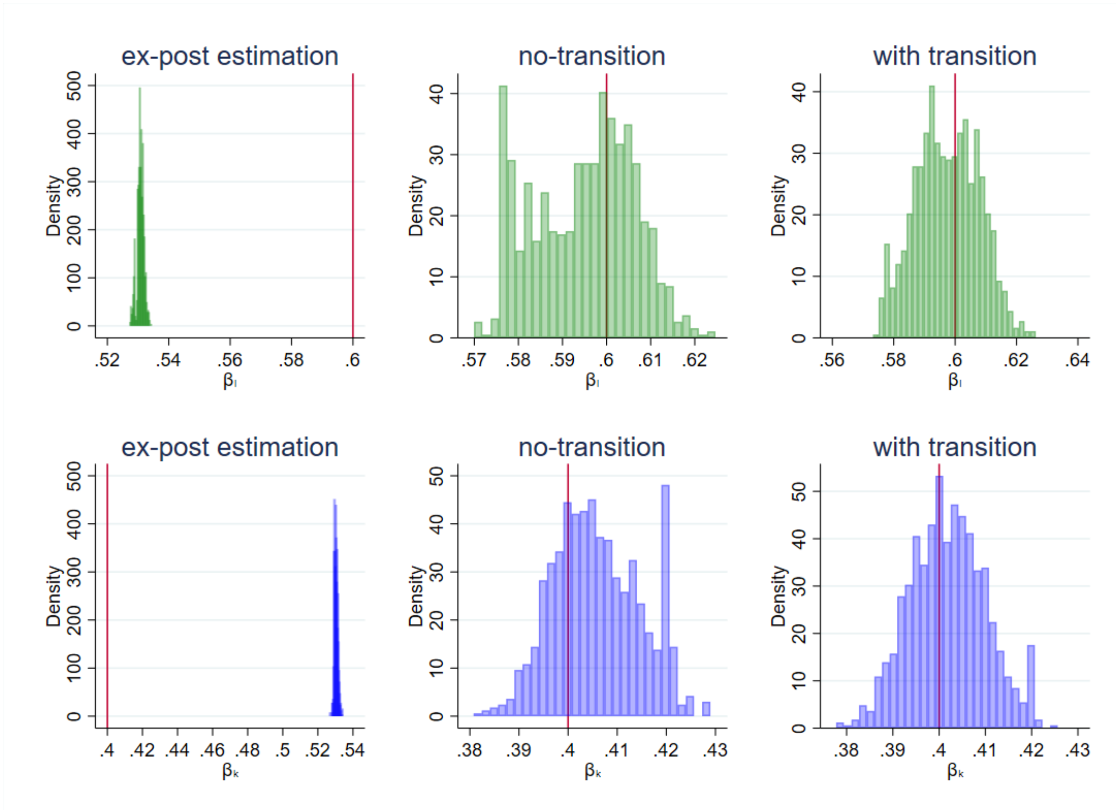
Figure D.2: Data Pattern for Conditionally Exogenous Policy Shock



Note: We conduct an event study by regression the variables on a full set of time dummies and its interaction with treatment group indicator.

Tables and Figures in Appendix

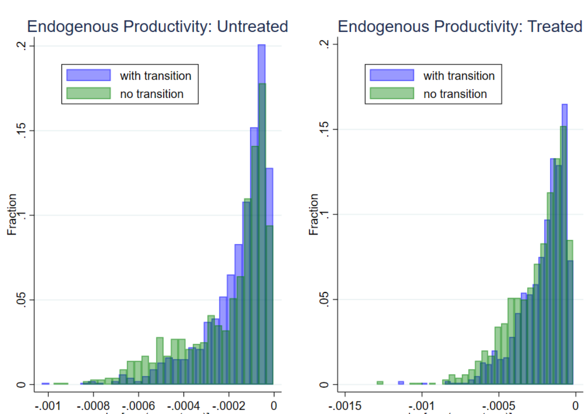
Figure D.3: Production Function Parameter Estimates



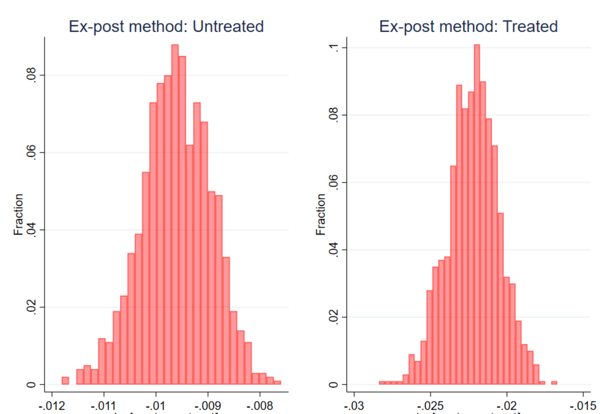
Note: Vertical lines indicate the true values.

Figure D.4: Correlations between Productivity Estimates and the True Productivity: Purely Exogenous Policy

(a) Endogenous Productivity



(b) Exogenous Productivity



Note: Untreated units include never treated units and not-yet-treated units.

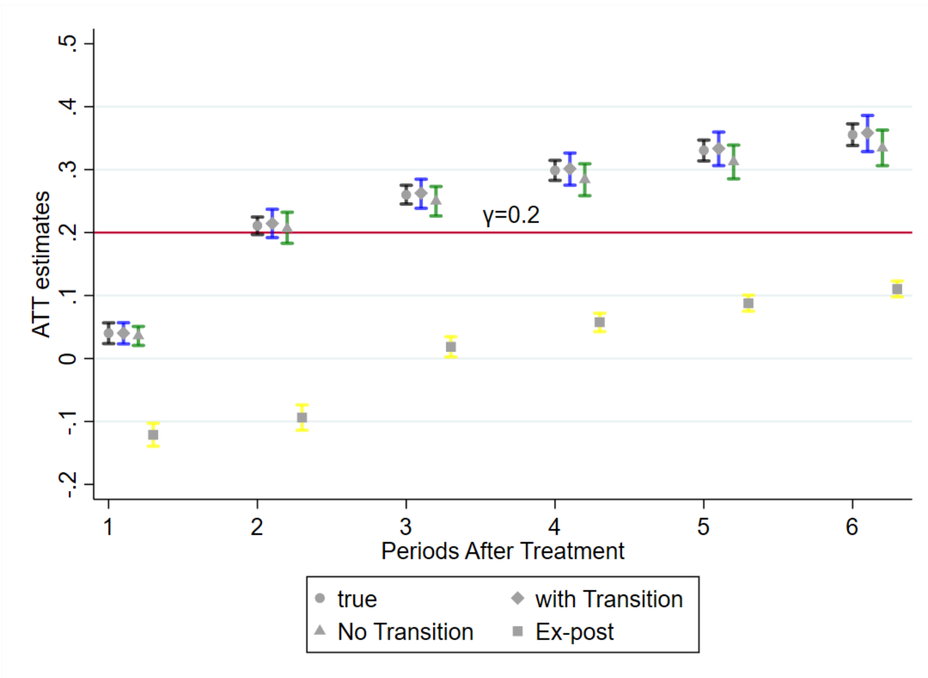
Table D.1: Regression Outcome for the Productivity's Evolution Process under Strictly Exogenous Policy Shock

	(1) True	(2) Ex-post	(3) No Transition	(4) With Transition
ω_t	0.699	0.723	0.700	0.701
	(0.684, 0.715)	(0.709, 0.738)	(0.683, 0.717)	(0.683, 0.718)
$\omega_t \times Treat_t \times Post_t$	0.100	0.120	0.101	0.098
	(0.078, 0.122)	(0.098, 0.142)	(0.077,0.125)	(0.074, 0.123)
$\omega_t \times Treat_t \times Mid_t$	0.021	0.026		0.021
	(-0.040, 0.079)	(-0.032, 0.085)		(-0.039, 0.080)
$Treat_t \times Post_t$	0.201	0.181	0.199	0.203
	(0.192, 0.211)	(0.171, 0.191)	(0.180,0.218)	(0.185, 0.220)
$Treat_t \times Mid_t$	0.040	0.031		0.040
	(0.024, 0.057)	(0.005, 0.059)		(0.024, 0.057)

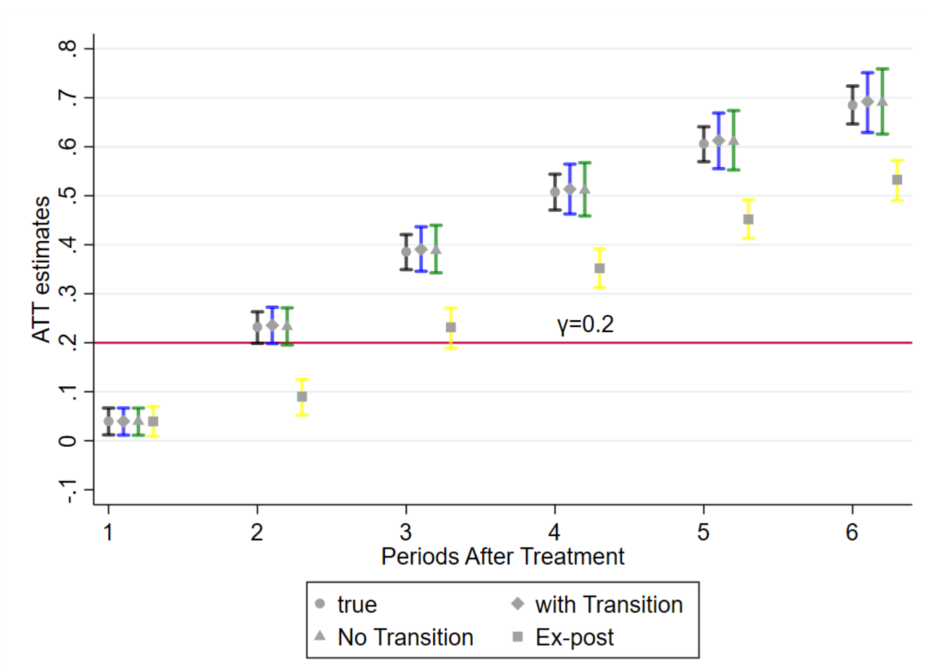
Note: 5th percentile and 95th percentiles are in the brackets.

Figure D.5: Estimates of Dynamic ATTs

(a) Structural Estimates of ATTs

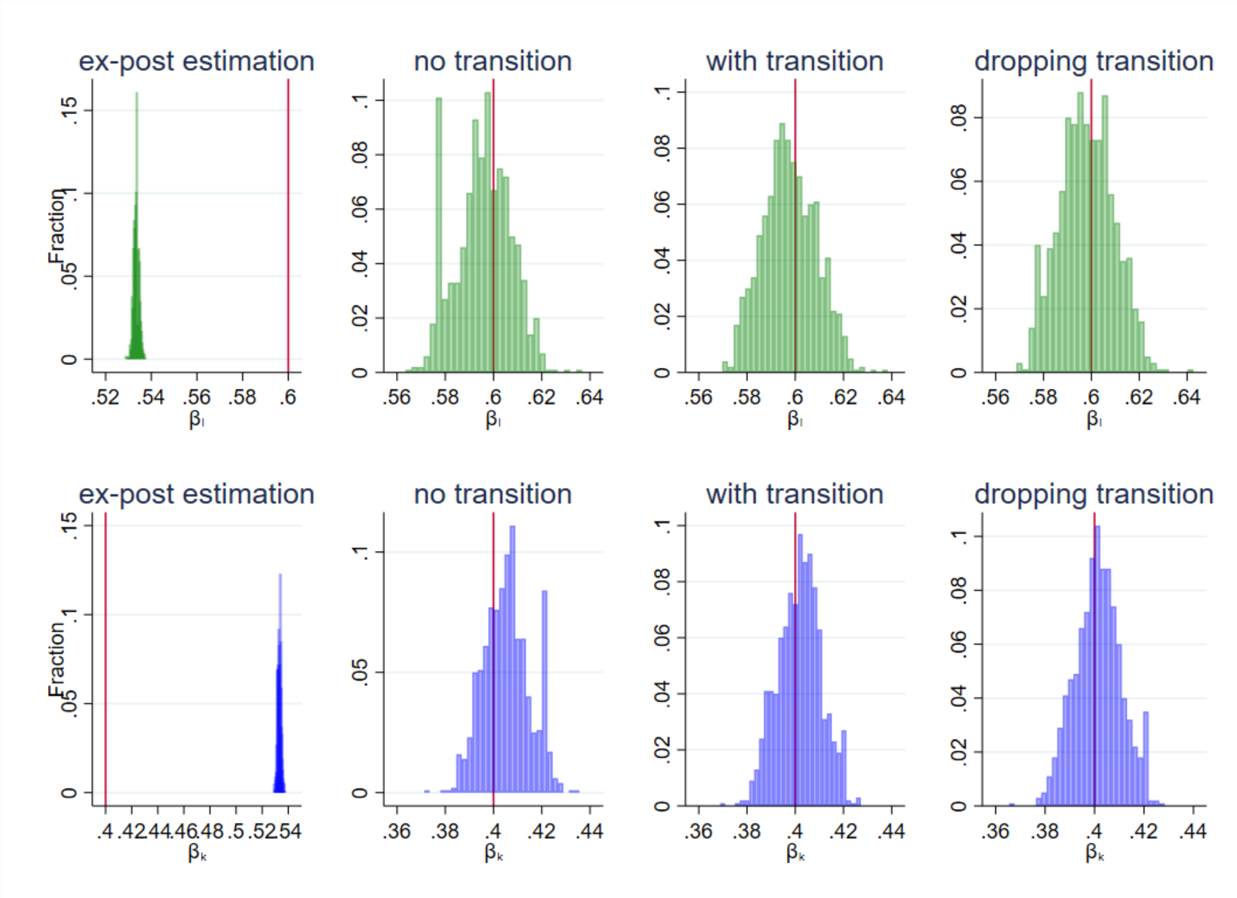


(b) Event-study Type Estimates of ATTs



Note: The capped vertical lines indicate the 5th and 95th percentiles of all 1000 experiments.

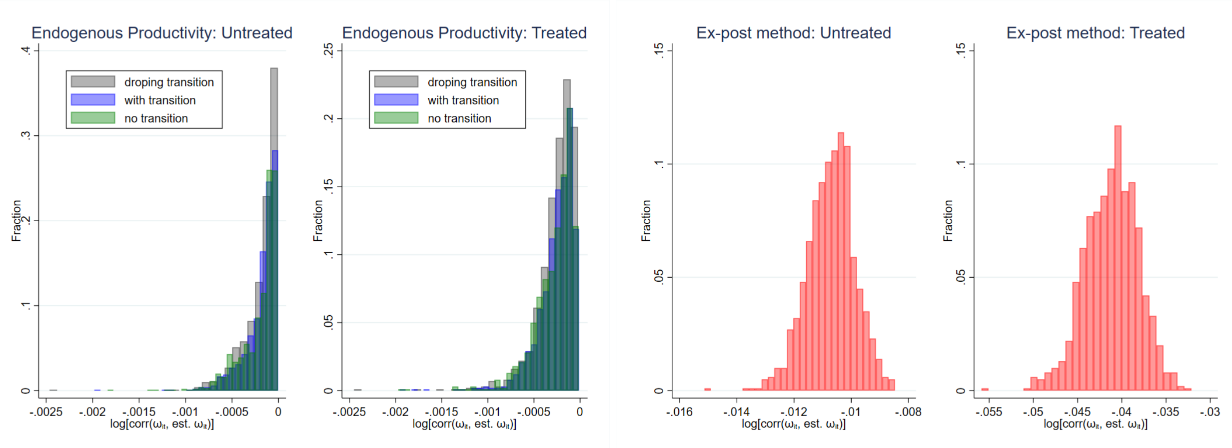
Figure D.6: Production Function Parameter Estimates for Conditionally Exogenous Policy



Note: Vertical lines indicate the true values.

Figure D.7: Correlations between Productivity Estimates and the True Productivity: Conditionally Exogenous Policy

(a) Endogenous Productivity (b) Exogenous Productivity



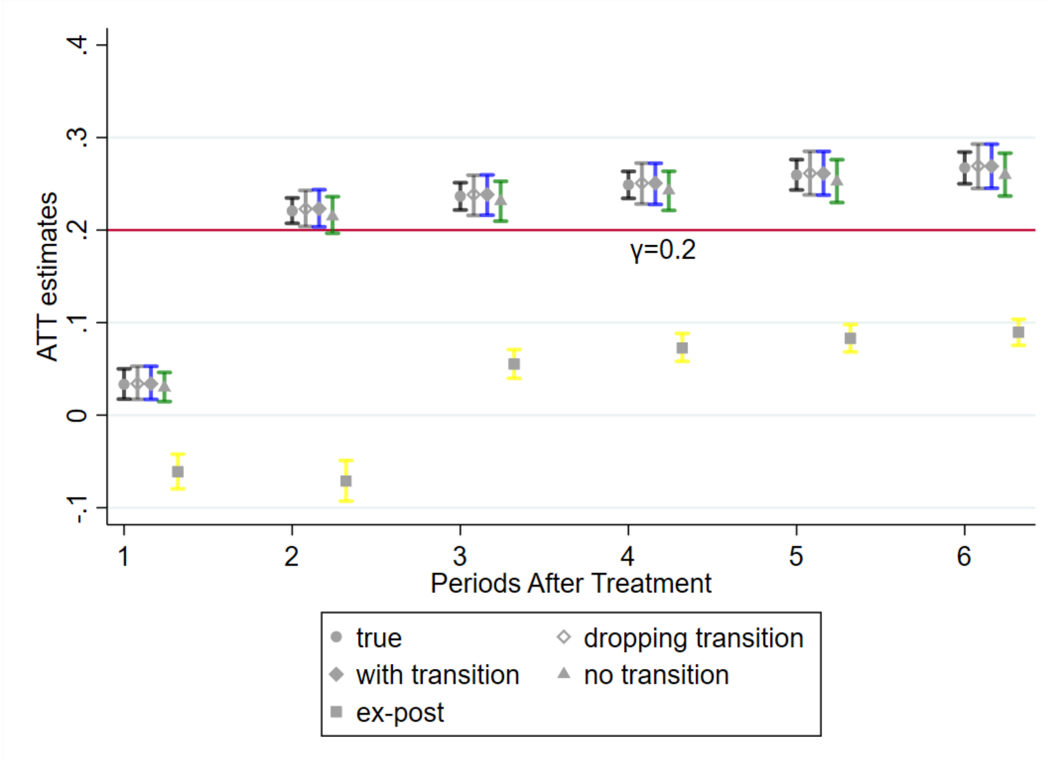
Note: Untreated units include never treated units and not-yet-treated units.

Table D.2: Regression Outcome for the Productivity's Evolution Process under Conditionally Exogenous Policy Shock

	Baseline	Ex-post	No transition	With transition	Dropping transition
ω_t	.700	.724	.709	.701	.701
	(.684, .715)	(.709, .738)	(.690, .727)	(.681, .719)	(.681, .719)
$\omega_t \times treat_i \times post_t$	.100	.117	.093	.099	.099
	(.076, .123)	(.094, .141)	(.069, .118)	(.075, .124)	(.074, .126)
$\omega_t \times Treat_t \times Mid_t$	0.021	0.020		0.018	
	(-0.039, 0.080)	(-0.069, 0.110)		(-0.074, 0.112)	
$treat_i \times post_t$	.200	.164	.198	.202	.202
	(.189, .212)	(.153, .175)	(.180, .217)	(.183, .220)	(.183, .220)
$treat_i \times mid_t$	.040	.047		.041	
	(.013, .067)	(.024, .071)		(.014, .069)	
Obs.	9,000	9,000	9,000	9,000	8,000

Note: 5th percentile and 95th percentiles are in the brackets.

Figure D.8: Estimates of Dynamic ATT for Conditionally Exogenous Shocks



Note: capped lines indicate the 5th and 95th percentile of the ATTs for all the treated units, and the white diamond indicate the mean value of the ATT for treated units.