Identifying Treatment Effects on Productivity*

Zhiyuan Chen† Moyu Liao‡

April 14, 2022

Abstract

This paper provides a novel econometric framework for analyzing the treatment effects on productivity. We embed the potential outcome framework into a dynamic model with firm-level productivity that can be affected by some discrete treatment. The treatment can be externally assigned or chosen by the firm. We first characterize conditions for non-parametrically identifying production functions. Then we provide conditions for identifying ATTs on the treated for absorbing treatment and generalize them to non-absorbing treatment. Our study has general implications for empirical studies on evaluating treatment effects on productivity.

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

*This is a preliminary draft, please do not circulate without our permission. 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.

1 Introduction

Productivity is one of the most important yet mysterious indicator in economic research. 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 destroy the identification of production functions, and more importantly, the desired effects of internal and external drivers on productivity. While this internal inconsistency has drawn 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 detecting treatment effects on productivity. We build a dynamic firm model which includes treatment influencing firms' productivity and/or production functions. The interested productivity driver is considered as a treatment that is either chosen by the firm (e.g., R&D investment, exporting, importing, etc.) or externally given (e.g., trade liberalization, environmental regulation, tax cuts, etc.), and can affect the productivity evolution and the production function. To lay the foundation for making causal inferences, we embed potential outcomes of productivity in the structural firm model. We call the productivity that are not realized after receiving the treatment as the productivity's potential outcomes.[3] As a result, firms make decisions based on the productivity's potential outcome. This modeling framework generalizes most of the fundamental firm behavior models considered in production function estimation as special cases. More importantly, it invites applying the powerful tool of causal inferences to detect the treatment effects on productivity. To the best of our knowledge, this is the first paper in considering potential outcomes of productivity in the

[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.

[3]We follow the statistical definition of potential outcomes. As pointed out by Donald B. Rubin in his R. A. Fisher Memorial Lecture (Rubin, 2005), Neyman (1923) is the first writer to use the potential outcome notation.

structural estimation of productivity.

We focus on a binary treatment variable, which is the most widely used in the causal inference literature. We consider a general non-parametric Markovian process for the potential productivity outcomes. The treatment variable is assumed to be mean-independent of the productivity shocks. We distinguish between two types of treatment effects on productivity. The first is a *trend effect* that changes the evolution of productivity permanently, the second is a *switching effect* that is active during the period of treatment status switching. While the trend effect is of the primary interest, the switching effect captures the initial change in the productivity during the period of treatment activation. Generally, the switching effect is non-zero and needs to be separated from the trend effect.[4] Using the proposed generic framework, we study the non-parametric identification and estimation of the interested treatment effects on productivity.

The first step to gauge the treatment effect on productivity is to recover it correctly. Our first identification result is on the production function's parameters. We show that if the treatment received time t is mean independent to the contemporary productivity shocks, the production functions in different treatment regimes can be non-parametrically identified up to a constant difference that depends on the treatment status. To acheive identification, one needs at least two groups of units. One group is treated for two periods, the other is untreated for two periods. This means that we do not require a group of never-treated units to identify the production functions. 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. We also show that if we impose more restrictions on the productivity's evolution rule or on the heterogeneity in treatment effects, the productivity can be recovered even for cases when the treatment status are switching constantly over periods.

In light of our identification strategy, we illustrate how the existing approaches may fail to identify production function parameters in the presence of productivity drivers. We examine two popular methods: the first is the ex-post regression approach which estimate productivity by assuming an exogenous productivity process and then regressing the estimated productivity on an interested treatment variable (e.g., Pavcnik (2003), Amiti and Konings (2007), Yu (2015)); 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 Jaumandreu (2013), Chen et al. (2021)).

[4]A simple reason is that the treatment effects accumulates over time and the initial timing of receiving treatment is somewhere in the middle of the treatment-switching period. Therefore, the treatment effects in the period of treatment switching differ from other periods.

We show that both methods suffer from a potential problem of model mis-specification about the productivity evolution, which causes biases in evaluating treatment effects productivity. The difference is that the ex-post method wrongly specify the productivity process both during the regime-switching period and ex-post periods, but the structural method is only subject to mis-specification in the regime-switching period. Crucially, the correctness of the structural method hinges on assuming away the switching effect during the treatment-switching period.

The identification of causal effects on productivity depends not only on the identification of production functions, but also the productivity process. Since the process is Markovian, even for a temporary action, its treatment effects on productivity will be carried on to following periods and become long-lasting. Under general treatment assignment rule, the challenge of identifying the full dynamic treatment effects is well acknowledged.[5] We show, however, under a conditional parallel trend assumption, the dynamic treatment effects on the treated units (ATTs) of absorbing treatment is point identified. The identification requires the existence of a group of control units that are not treated for at least two consecutive periods, which allows us to recover the productivity's evolution rule absent any treatment. Given the productivity of treated units prior to receiving the treatment, the non-treated productivity process can be used to infer the expected potential outcome. Similar to the identification of production functions, we do not need a group of never-treated units to identify the dynamic treatment effects. This implies that we still can identify ATTs if the treatment is assigned exclusively based on some firm characteristics such that no good comparison can be found in the never-treated group. This identification strategy motivates us to propose a simulation-based strategy to estimate ATTs. To compute the ATTs, one only needs to simulate the potential productivity outcomes for each treated units and take the difference between the observed productivity and the simulated potential outcomes. This flexible estimator allows rich treatment effect heterogeneity.

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.

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

[5] Vikström et al. (2018) point out that, because the treated units and control units drop out at different rates, the randomization only guarantees the comparability of treatment and controls at the time of randomization. This leads to the non-identification of long-run treatment effects.

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 for 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 a broad 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.

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 6 is the Monte Carlo simulation. Section 7 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 policy 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

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

where Q_{it} is the output, ω_{it} is the realized productivity known by the firm, 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 $\boldsymbol{\beta}$ is infinite when the production function is non-parametric. $\boldsymbol{\beta}$ 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,[6] which captures possible impacts on managerial efficiencies (Chen et al., 2021).

The firm knows its productivity when making decisions, but the econometrician does not. 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

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

To facilitate our exposition, we define an indicator for treatment changes.

Definition 1. *(Treatment switching indicator) We define a 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 follow this tradition to further assume a generalized 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. 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*

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

[6]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. One can also think about some treatment only affects the optimal choices instead of affecting the production technology directly, e.g. Shenoy (2021). This case is also incorporated in our formulation since we can include the input constraints in the production function as the constraints in the Lagrangian function.

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 policy 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 as the case of no treatment, i.e. $D_{it} = 0$. The generalized productivity evolution process (3) itself also has interpretable 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* $\bar{\boldsymbol{h}}_1(\omega) = \bar{\boldsymbol{h}}_0(\omega + C)$*. These conditions lead to* $\omega_{it}^1 = \omega_{it}^0 + C$ *almost surely for all* t*.*

Example 2. *(Divergence of Productivity when Policy 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. In each period, firm 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.*

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 between 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 $(\omega^0_{it}, \omega^1_{it})$.*

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 variable and can be degenerate if no unobserved heterogeneity exists. Define $\Pi(S_{it})$ as the per-period indirect profit function, the Bellman equation for the firm's dynamic programming problem is

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

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.[7] If policy variable is externally assigned, D_{it} disappears from the firm's choice set. In this case, we impose that $C_D(\cdot) = 0$.[8] We summarize the timing assumption by formally stating firm i's information set $\mathcal{I}^F_{it}$:

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

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

The firm model is a natural generalization of dynamic discrete choice models following the seminal work by Rust (1987). Instead of focusing one dynamic discrete choice

[7] 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.

[8] 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.

(e.g., the replacement of GMC bus engines considered in Rust (1987)), our model includes choices of capital and labor, which are crucial firm decisions in identifying production functions. This adds to the complication of solving the structural model and analyze subjective treatment effects. However, the identification strategy does not depend on the functional forms of cost functions $C_I(\cdot)$ and $C_L(\cdot)$. One can include cost shifters into them to reflect the variations in capital and labor costs across firms and time. In the case of externally assigned treatment, this model 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 $C_D(\cdot)$ does play an important role in inferring the treatment effects on productivity. 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. In their setting, the hidden heterogeneity in costs of taking the treatment does not affect their estimation of the productivity evolution equation. However, we emphasize that it requires additional efforts to acheive causal interpretation for interested treatment effects.

2.2 Treatment-Effect Objects

The firm anticipates the effect of the treatment on productivity in period $t+1$ when making decisions. Moreover, a firm also make expectation on the evolution of the treatment when the policy is not controlled by the firm. A firm starts to receive the treatment when G_{it}=1. 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, 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 policy effect for firm i at time t is $\omega_{it}^1 - \omega_{it}^0$. The trend effect is given by the function $\bar{h}_1(\cdot) - \bar{h}_0(\cdot)$. The positive (resp. negative) switching effect is given by the function*

$\boldsymbol{h}_i^+$ *(resp.* $\boldsymbol{h}_i^-$*).*

The trend effect and switching effects come from the structure of the 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 hold 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* $\tilde{D}_i^T$*.*

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

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} is chosen by the firm according to (4). Indeed, from the firm's dynamic optimization problem, we know D_{it} is a function of $\mathcal{I}_{it}^F$, denoted by $D_{it} = g(\mathcal{I}_{it}^F)$. Then given the information set $\mathcal{I}_{it}^F$, D_{it} 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. $D_{it} = \tilde{g}(\mathcal{I}_{it}^F, \eta_{it})$ 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* $\mathcal{Z}_{it} = \mathcal{I}_{it}^F / \{(\omega_{is}^1, \omega_{is}^0, \zeta_{is})_{s \leq t}\}$.

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 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 adopt the convention that lower and upper case letters represent logs and levels, respectively.

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

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

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 is 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

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

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 (a) is given in GNR. We use the techniques in GNR to prove statement (b). Let $\omega_{it-1}(\boldsymbol{\beta}) \equiv \Phi_{it-1}(k_{it-1}, l_{it-1}, m_{it-1}) - f(k_{it-1}, l_{it-1}; \boldsymbol{\beta})$. We first note that

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

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

Therefore, f 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 in the identified set, then $(e^c F, \tilde{h})$ where $\tilde{h}(\omega) = h(\omega - c)$ is also in the identified set for all $c \in \mathbb{R}$.

3.2 Recovering the Productivity with Variations in Treatment Status

We now extend the identification result to the case with 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 $(\epsilon_{it}^1, \epsilon_{it}^0)$.

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

$$\mathbb{E}[(\epsilon_{it}^0, \epsilon_{it}^1)|\mathcal{Z}_{it}] = \mathbf{0}.$$

Assumption 3.2 allows the treatment decision to be dependent of the past potential outcomes ω_{it-1}^0 and ω_{it-1}^1. Consider a case where D_{it} is selected by the firm. A firm may observe its productivity $(\omega_{it-1}^0, \omega_{it-1}^1)$ when making the decision on whether to adopt the treatment or not, and the productivity shocks $(\epsilon_{it}^0, \epsilon_{it}^1)$ 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.

Proposition 3.1. *Suppose Assumptions 2.1-3.2 hold. Then 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 expression (10) follows similarly. We can write

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

where $\omega^0_{it}(\boldsymbol{\beta})$ 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. By Lemma 3.1, the result follows. □

The scalar non-identification result can ruin the estimation 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^1_{it} recovered under $(F(\cdot, 1; \boldsymbol{\beta}), \bar{h}_1)$ from $\tilde{\omega}^1_{it}$ recovered under $(e^{c_1}F(\cdot, 1; \boldsymbol{\beta}), \tilde{h}_1)$. In particular, we have $\tilde{\omega}^1_{it} = \omega^1_{it} - c_d$. Therefore, we recommend normalizing the production function constant $c_1 = c_0 = 0$.

3.3 Alternative Procedures for Restricted Productivity Processes

Our moment conditions in Proposition 3.1 requires only Assumption 3.2 and impose no additional assumptions on the productivity evolution process 3. While implementing moment conditions in Proposition 3.1 requires minimal structural assumptions, we typically require a relatively large sample of two-year consecutive observations satisfying $D_{it} = D_{it-1}$. Such data requirement can be satisfied when the panel satisfies a difference-in-difference type design. However, if the policy variable is volatile over time, we cannot implement Proposition 3.1 in practice.

We now consider several alternative assumptions on the evolution process that allow us to derive more flexible moment conditions that identify the production functions.

3.3.1 Independent Evolution Process

Let's consider the case where the two productivity processes evolves independently as in Example 3. In this case, we may substitute the Markov process back several periods to identify the production function. In this case, we can simplify the notation and use h_1 and h_0 to denote the transition functions for treated and non-treated units, respectively.

Assumption 3.3. *For $d = 0, 1$, the Markov process ω_{it}^d satisfies*

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

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

Assumption 3.3 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.

Proposition 3.2. *Suppose Assumption 3.3 hold and the productivity process satisfies Example 3, then the following two moment conditions hold:*

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

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

Its unfortunate that we cannot show the non-parametric point identification of the production function $\boldsymbol{\beta}$. The reason is the following: the error terms ϵ_{it-s} is correlated with k_{it} and l_{it} for all $s \geq 1$. Without solving firms' dynamic optimization problem, we cannot directly show k_{it-s+1} and l_{it-s+1} can serve as good instruments for k_{it} and l_{it}. 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. For the sake of brevity, we only consider homogeneous transition process, i.e. $\boldsymbol{h}_i^+ = \boldsymbol{h}^+$, $\boldsymbol{h}_i^- = \boldsymbol{h}^-$. To simplify 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.

Proposition 3.3. *Suppose Assumptions 2.1-3.2 hold and the productivity evolution process satisfies Example 2 with homogeneous transition process. 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 function parameter $\boldsymbol{\beta}$, and the evolution process $\bar{\boldsymbol{h}}_d$, $\boldsymbol{h}^+$, $\boldsymbol{h}^-$ non-parametrically up to a constant difference.

Compared to the moment conditions in Doraszelski and Jaumandreu (2013), Proposition 3.3 requires the transition period to be treated separately. Moment condition (14) and (15) are imposed to identify the positive transition process h_1^+ and h_0^-, separately. Moreover, they are extra moments that identifies the production functions. When the treatment variable is volatile, these two moment conditions allow us to make better use of the limited data.

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 in 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 policy-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. For the firms in the treated group, $D_{it} = \mathbb{1}(t \leq T_0)$.

We use this economic 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 $\mathcal{Z}'_{it} = \{k_{it}, l_{it}, k_{it-1}, l_{it-1}, m_{it-1}\}$.

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, we estimate 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 $\hat{\omega}_{it}$.

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

$$\mathbb{E}[\omega_{it}^0(\boldsymbol{\beta}) - \bar{h}_0(\omega_{it-1}^0(\boldsymbol{\beta}))|\mathcal{Z}_{it}] = 0 \quad \forall t \leq T_0.$$

By Proposition 3.1, this moment condition identifies $\boldsymbol{\beta}$ and $\bar{h}_0$. We now derive the inconsistency of (6). For $t \geq T_0 + 2$, the moment condition the moment equation (6) becomes

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

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 a consistent estimator of the true $\boldsymbol{\beta}$. As a consequence, $\hat{\omega}_{it}$ is not a consistent estimator of ω_{it}, and the subsequent policy 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

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

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

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

Moment condition (17) is correctly specified. In particular, by Proposition 3.1, β, $\bar{h}_0$ are identified from the $t \leq T_0$ moment equality of (17), and $\bar{h}_1$ is identified from the $t \geq T_0 + 2$ moment equality of (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, 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 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, 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 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 $h_i^+ = h^+ = \bar{h}$, then 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 policy realizes 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 $\bar{h} \neq h_i^+$.

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 not identified, and we instead focus on the average treatment effect (ATE) and the average treatment effect on the treated (ATT).

Corollary 1. *Under Assumption 2.1-3.2, if there exists a t such that $Pr(D_{it} = D_{it-1} = d) \neq 0$, we can recover the unobserved potential productivity ω_{is}^d for firms such that $D_{is} = d$.*

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 $\omega_{is} = (q_{is} - f(k_{is}, l_{is}, m_{is}, D_{is}; \boldsymbol{\beta})$, which is ω_{is}^d since $D_{is} = d$. $\square$

Since the individual effective productivity is identified, the econometrician can view ω_{it} as 'observed'. We 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. We first show that the conditional distribution of potential productivity is identified given our assumptions in previous sections.

4.1 ATE: The Sequential Randomization Condition

Given the econometrician's information set $\mathcal{I}_{it}^E$, 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 firm indicators for interested firms.

Assumption 4.1. *(SR-E) $D_{it} \perp (\omega_{is}^1, \omega_{is}^0)_{s \geq t} | \mathcal{I}_{it}^E, i \in g$.*

The econometrician's sequential randomization assumption 4.1 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, i.e. $D_{it} = \breve{\psi}(\mathcal{I}_{it}^E, \eta_{it})$ for some η_{it} independent of

$(\omega_{is}^0, \omega_{is}^1)_{s\geq t}$, 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 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 satisfies Example 2, and the firm needs to decide whether to participate in an absorbing treatment[9]. 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 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 4.1. *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, $ATE_{g,t} \equiv \mathbb{E}[\omega_{it}^1 - \omega_{it}^0 | i \in g]$ 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 4.1 follows directly by the propensity score matching method. We now discuss whether Assumption 4.1 holds or not in different empirical settings.

On the other hand,

Given the difficulty of justifying the sequential randomization condition SR-E, we turn to the identification of ATT, which typically requires fewer assumptions on the structural model. We start with a simple absorbing treatment environment and then generalize the results to non-absorbing treatment.

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

4.2 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 $D_{it-1} \leq D_{it}$. The absorbing treatment is more general than it appears. 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.[10]

Let $e_i > 1$ be the first period that firm i starts to receive treatment.[11] 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 trend assumption is satisfied. In the analysis here, we focus on the 0-period-head ATT. The following parallel trend assumption is needed.

Assumption 4.2. *(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 4.2 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

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

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

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 hence influence 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 4.3. *(Conditional Parallel Trend)* $h_{id}^+(\omega_{it}^d, \omega_{it}^{1-d}) = \bar{h}_d(\omega_{it}^d)$ *for* $d = 0, 1$.

Assumption 4.3 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 4.3 has the following economic meaning: Transition function for the untreated potential outcome is not influenced by the treatment status.

Proposition 4.2. *Under Assumption 4.3, the 0-period-ahead ATT is identified as* $ATT_{g,0} = \mathbb{E}[\mathbb{E}[\omega_{it} - \bar{h}_0(\omega_{it-1}) | e_i = t, i \in g] | i \in g]$.

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

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

where (1) by definition, (2) follows by Assumptions 3.2 and 4.3, (3) follows by the potential outcome (2). Further take the expectation with respect to the treatment time to get the result. $\square$

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 4.2 does not work anymore. Let $\bar{h}^\ell_0$ be the ℓ period productivity transition process, we can write $\omega^0_{ie_i+\ell} = \bar{h}^{(\ell)}_0(\omega^0_{ie_i}, (\epsilon^0_{is})^{e_i+l}_{s=e_i})$. We now state two sufficient conditions such that the ℓ-period-ahead ATT is identified.

Assumption 4.4. *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 arguments. Moreover, there is no mean-selection of $(\epsilon^0_{is})_{s>e_i}$ on D_{ie_i}, i.e. $\mathbb{E}[\epsilon^0_{is}|D_{ie_i} = 1] = 0$ for all $s \geq e_i$.

Proposition 4.3. *Under Assumption 3.2, 4.3, and 4.4, the ℓ-period-ahead ATT is identified as $ATT_{g,\ell} = \mathbb{E}[\omega_{it+\ell}|e_i = t, i \in g] - \mathbb{E}[\bar{h}^{(\ell)}_0(\omega_{it})|e_i = t, i \in g]$.*

Proof. Note that

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

where (1) by the conditional parallel trend assumption and Assumption 4.4, (2) follows by Assumptions 3.2 and linearity of $r(\cdot)$. $\square$

Assumption 4.4 is satisfied for an AR(1) productivity process, but generally fails when non-linearity appears in the transition function $\bar{h}_0$. Therefore, Assumption 4.4 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 4.5. *The ℓ-period-ahead shock independence $(\epsilon^0_{ie_i}, ..., \epsilon^0_{ie_i+l}) \perp e_i | i \in g, \omega^0_{ie_i-1}$.*

We require more than the conditional mean-independence of the future productivity shocks with respect to the treatment time. Assumption 4.5 allows for nonlinearity in $\bar{h}^{(s)}_0(\cdot)$. We first derive the average treatment effect for a cohort, i.e. g indicates the treatment-cohort year of the firm, and $i \in g$ if and only if $e_i = g$. The overall ATT can be derived by integrating out the cohort effect. We define the ℓ-period-ahead treatment effect for cohort g as

$$ATT_{g,\ell} = \mathbb{E}[\omega^1_{it} - \omega^0_{it}|t = e_i + \ell, e_i = g]. \tag{23}$$

Proposition 4.4. *Suppose Assumption 4.3 and 4.5 hold. Let* $\Delta(\ell, g, \omega) = \mathbb{E}[\omega_{ig+\ell}|e_i = g, \omega_{ig-1} = \omega] - \mathbb{E}[\omega_{ig+\ell}|e_i > g + \ell, \omega_{ig-1} = \omega]$*. The* ℓ*-period-ahead ATT for cohort* g *is identified as*

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

where the expectation is taken over the conditional distribution of ω_{ig-1} *given* $e_i = g$*.*

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

$$\begin{aligned} ATT_{g,\ell}|_{\omega_{ig-1}=\omega} &=_{(a)} \mathbb{E}[\omega_{ig+\ell}|e_i = g, \omega_{ig-1} = \omega] - \mathbb{E}[\bar{h}_0^{(\ell)}(\omega_{ig-1}, \epsilon_{ig}^0, ..., \epsilon_{ig+\ell}^0)|e_i = g, \omega_{ig-1} = \omega] \\ &=_{(b)} \mathbb{E}[\omega_{ig+\ell}|e_i = g, \omega_{ig-1}] - \mathbb{E}[\bar{h}_0^{(\ell)}(\omega_{ig-1}, \epsilon_{ig}^0, ..., \epsilon_{ig+\ell}^0)|e_i > g + l, \omega_{ig-1} = \omega] \\ &=_{(c)} \mathbb{E}[\omega_{ig+\ell}|e_i = g, \omega_{ig-1} = \omega] - \mathbb{E}[\omega_{ig+\ell}|e_i > g + \ell, \omega_{it-1} = \omega] = \Delta(l, g, \omega), \end{aligned} \tag{24}$$

where (a) follows by the conditional parallel trend assumption and the potential outcome equation, (b) follows by Assumptions 4.5, and (c) follows by the transition procedure (3) for untreated firms. The result follows by further integrating out the ω_{ig-1}. □

Proposition 4.4 requires us to match over the lagged productivity for each g-cohort firms with $g + \ell$-not-yet-treated firms. This is because we cannot observe the untreated shocks ϵ_{it}^0 for treated firms and the higher moments of ϵ_{it}^0 matters for the ℓ-period evolution process $\bar{h}_0^{(\ell)}$. On the other hand, Proposition 4.3 uses firm i's own lagged productivity as controls, because the linearity of the residual function $r(\cdot)$ in Assumption 4.4.

However, if treatment time is staggered, we may have only a fraction of firms that are treated after $g + l$ period, and $g + 1$ to $g + \ell$ cohort firms are ignored. This can lead to inefficient use of information. We now propose a strong condition that allows us to use the information from $g + 1$ to $g + \ell$ cohort firms.

Assumption 4.6. *The productivity shocks* $(\epsilon_{is}^1, \epsilon_{is}^0)_{s \geq e_i} \perp (e_i, \mathbb{1}(i \in g), \omega_{ie_i-1}^0)$ *and* $\epsilon_{it}^d \sim_{i.i.d} G_\epsilon^d(\cdot)$ *for* $d = 0, 1$*, where the* $i.i.d$ *is both across firm index* i *and time index* t*.*

Assumption 4.6 is stronger than 4.5. However, the strong condition in Assumption 4.6 allow us to identify $ATT_{g,l}$ using results similar to Proposition 4.6.

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

$$ATT_{g,\ell} = \mathbb{E}[\omega_{ig+\ell}|e_i = g] - \mathbb{E}_{(G_\epsilon^0)^\ell}[\bar{h}_0^{(\ell)}(\omega_{ig-1}, \epsilon_{ig}^0, ..., \epsilon_{ig+\ell}^0)|e_i = g],$$

where the second expectation is taken over the joint distribution of $(\epsilon_{ig}^0, ..., \epsilon_{ig+\ell}^0)$*.*

Proof. Let g be the target cohort in Proposition 4.5, and let g' be any cohort such that $g' > g$.

With the identified $\bar{h}_0$ from Proposition 3.1, for any g'-cohort firm i at time $t < g'$, we can recover its $\epsilon^0_{it} \equiv \omega_{it} - \bar{h}_0(\omega_{it-1})$, so the distribution of ϵ^0_{it} conditional on $(e_i, \mathbb{1}(i \in g), \omega^0_{ie_i-1})$ is identified. However, the productivity shocks are independent of $(e_i, \mathbb{1}(i \in g), \omega^0_{ie_i-1})$, so $G^0_\epsilon(\cdot)$ is identified.

□

4.3 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.[12] 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. Instead, we focus on the 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 at a time g positive/negative treatment switcher:

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

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 4.3, 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 $e_i = g$,

$$\begin{aligned} ATT^+_{g,0} &=_{(1)} \mathbb{E}[\omega^1_{ig} - \omega^0_{ig} | D_{ig-1} = 0, D_{ig} = 1] \\ &=_{(2)} \mathbb{E}[\omega^1_{ig} | D_{ig-1} = 0, D_{ig} = 1] - \mathbb{E}[\bar{h}_0(\omega^0_{ig-1}) | D_{ig-1} = 0, D_{ig} = 1] \\ &=_{(3)} \mathbb{E}[\omega_{ig} | D_{ig-1} = 0, D_{ig} = 1] - \mathbb{E}[\bar{h}_0(\omega_{it-1}) | D_{ig-1} = 0, D_{ig} = 1], \end{aligned}$$

[12] 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.

where (1) by definition, (2) follows by Assumptions 3.2 and 4.3, (3) follows by the potential outcome (2). $\square$

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

Assumption 4.7. *The ℓ-period-ahead shock independence holds for all $t > 1$:*

$$(\epsilon_{it}^0, ..., \epsilon_{it+\ell}^0) \perp (D_{it}, ..., D_{it+\ell}) | \omega_{it-1}, D_{it-1}.$$

Assumption 4.7 generalizes Assumption 4.5 in the absorbing treatment case. Since treatment is not absorbing, we need to further conditional on the lagged treatment D_{it-1}.

Proposition 4.7. *Suppose Assumption 4.3 and 4.7 hold. Let*

$$\begin{aligned}\tilde{\Delta}(\ell, g, \omega) = \mathbb{E}[\omega_{ig+\ell} | D_{ig-1} = 0, D_{ig} = ... = D_{ig+\ell} = 1, \omega_{ig-1} = \omega] \\ - \mathbb{E}[\omega_{ig+\ell} | D_{ig-1} = D_{ig} = ... = D_{ig+\ell} = 0, \omega_{ig-1} = \omega].\end{aligned}$$

The ℓ-period-ahead ATT for time-g positive switcher is identified as

$$ATT_{g,\ell}^+ = \mathbb{E}[\tilde{\Delta}(l, g, \omega_{ig-1}) | D_{ig-1} = 0, D_{ig} = ... = D_{ig+\ell} = 1],$$

where the expectation is taken over the conditional distribution of ω_{ig-1} given $D_{ig-1} = 0, D_{ig} = ... = D_{ig+\ell} = 1$.

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

$$\begin{aligned} ATT_{g,\ell}^+|_{\omega_{ig-1}=\omega} =_{(a)} & \mathbb{E}[\omega_{ig+\ell} | D_{ig-1} = 0, D_{ig} = ... = D_{ig+\ell} = 1, \omega_{ig-1} = \omega] \\ & - \mathbb{E}[\bar{h}_0^{(\ell)}(\omega_{ig-1}, \epsilon_{ig}^0, ..., \epsilon_{ig+\ell}^0) | D_{ig-1} = 0, D_{ig} = ... = D_{ig+\ell} = 1, \omega_{ig-1} = \omega] \\ & =_{(b)} \mathbb{E}[\omega_{ig+\ell} | D_{ig-1} = 0, D_{ig} = ... = D_{ig+\ell} = 1, \omega_{ig-1} = \omega] \\ & - \mathbb{E}[\bar{h}_0^{(\ell)}(\omega_{ig-1}, \epsilon_{ig}^0, ..., \epsilon_{ig+\ell}^0) | D_{ig-1} = D_{ig} = ... = D_{ig+\ell} = 0, \omega_{ig-1} = \omega] \\ & =_{(c)} \mathbb{E}[\omega_{ig+\ell} | D_{ig-1} = 0, D_{ig} = ... = D_{ig+\ell} = 1, \omega_{ig-1} = \omega] \\ & - \mathbb{E}[\omega_{ig+\ell} | D_{ig-1} = D_{ig} = ... = D_{ig+\ell} = 0, \omega_{ig-1} = \omega] = \tilde{\Delta}(l, g, \omega), \end{aligned} \tag{26}$$

where (a) follows by the conditional parallel trend assumption, (b) follows by Assumptions 4.7, and (c) follows by the transition procedure (3) for untreated firms. The result follows by further integrating out the ω_{ig-1}. $\square$

Remark 4.1. *Assumption 4.7 rules out some empirical settings. In particular, if the treatment decision is a per-period decision as in (4), then the value of ϵ_{it}^0 will influences ω_{it}^0 and hence influence* *the value of D_{is} for $s \geq t+1$. However, if the treatment must be maintained for l periods due to exogenous constraints (say contract or legal constraints), then $D_{it}, D_{it+1}, ..., D_{it+l}$ is determined based on the time-t information $\mathcal{I}_{it}^F$, and Assumption 4.7 can be satisfied.*

5 Estimation

6 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.

6.1 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.[13]

Table 1 about here

Figure 1 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, 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 2, 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 2 about here

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

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 1, we report the regression outcome of the productivity process using the productivity estimates. Table 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 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 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 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 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.[14] Then the treatment effect of the policy on productivity for any period is calculated by take

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

the difference between the observed productivity and the simulated potential productivity. In Figure 3 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 3 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 3 about here

6.2 Endogenous Firm-level Action

We continue to analyze the estimation results for the datasets generated by conditionally exogenous policy.[15]

Figure 4 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 4 about here

[15] 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 fewer coefficients.

Figure 5 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 5 about here

We report the regression outcome for the productivity's evolution process in Table 2. Table 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 2 about here

Figure 6 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 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.

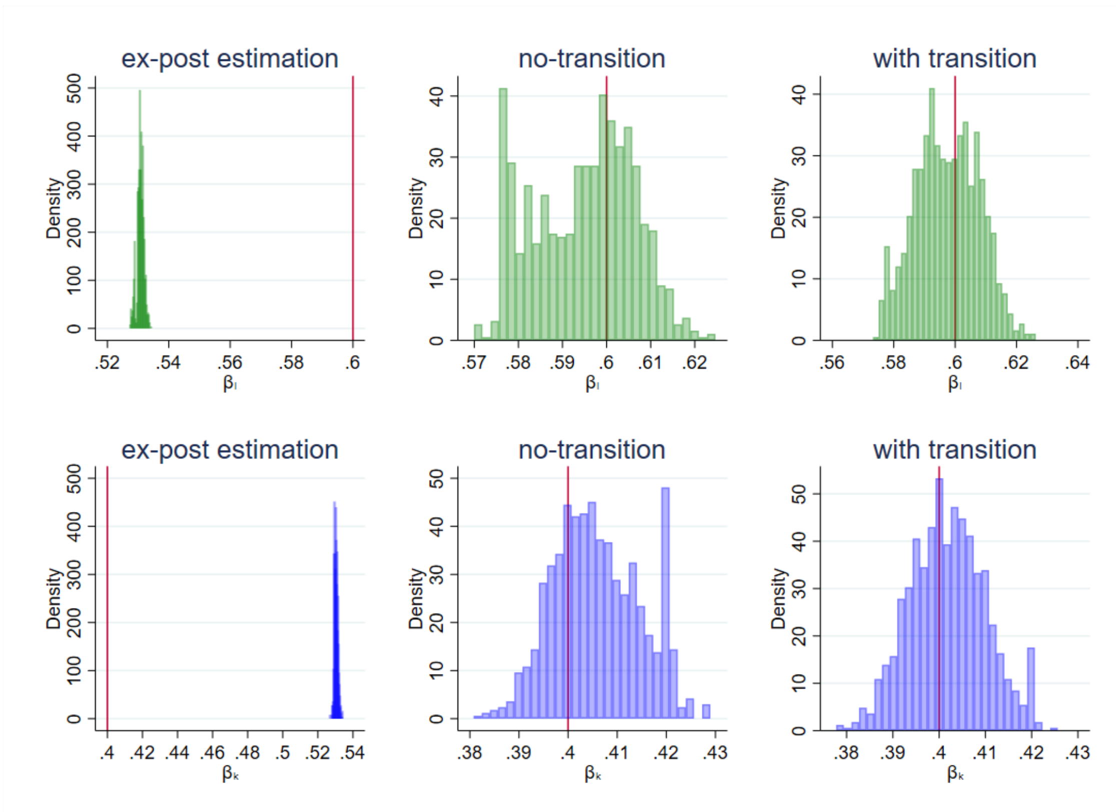
Figure 6 about here

7 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 reflect 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.

Tables and Figures

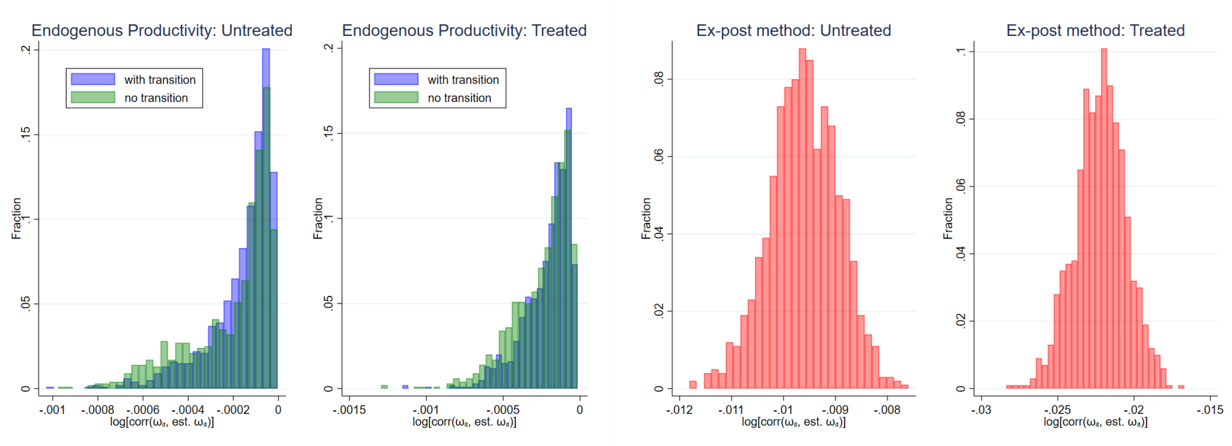
Figure 1: Production Function Parameter Estimates



Note: Vertical lines indicate the true values.

Figure 2: 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.

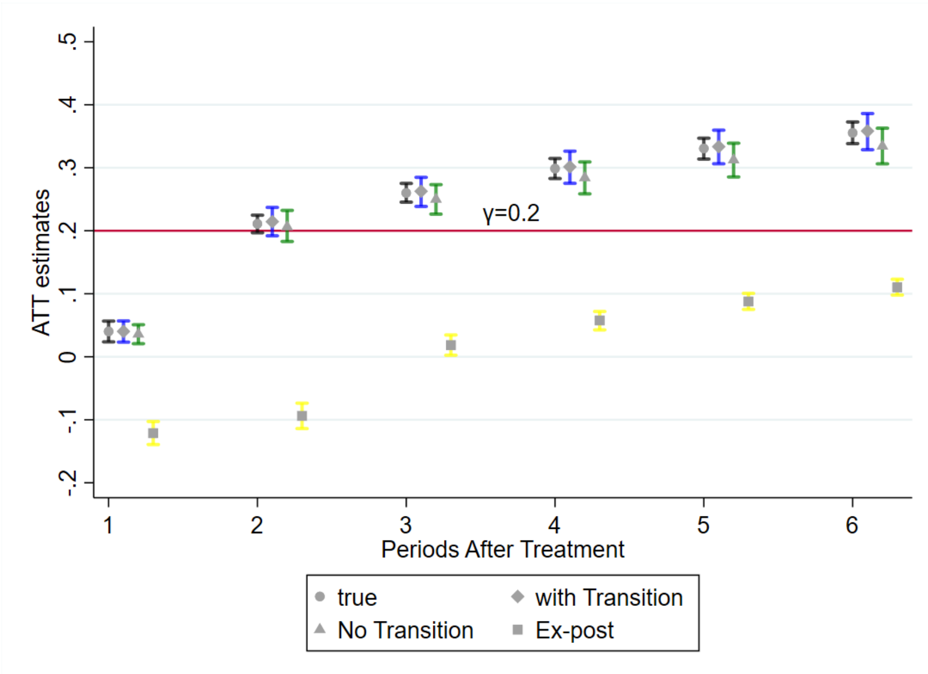
Table 1: Regression Outcome for the Productivity's Evolution Process under Strictly Exogenous Policy Shock

	(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)

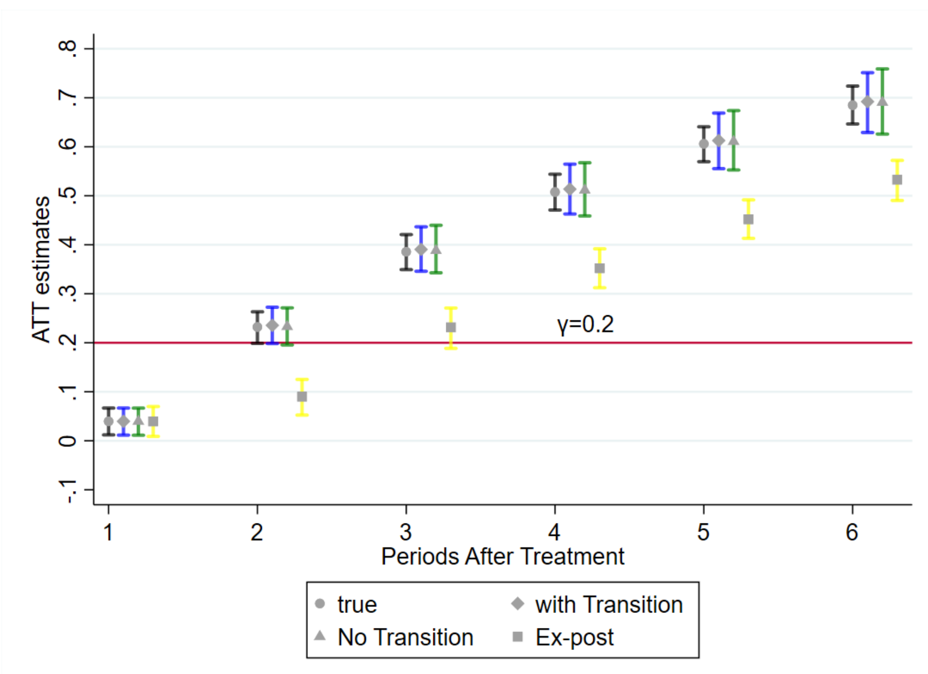
Note: 5th percentile and 95th percentiles are in the brackets.

Figure 3: Estimates of Dynamic ATTs

(a) Structural Estimates of ATTs

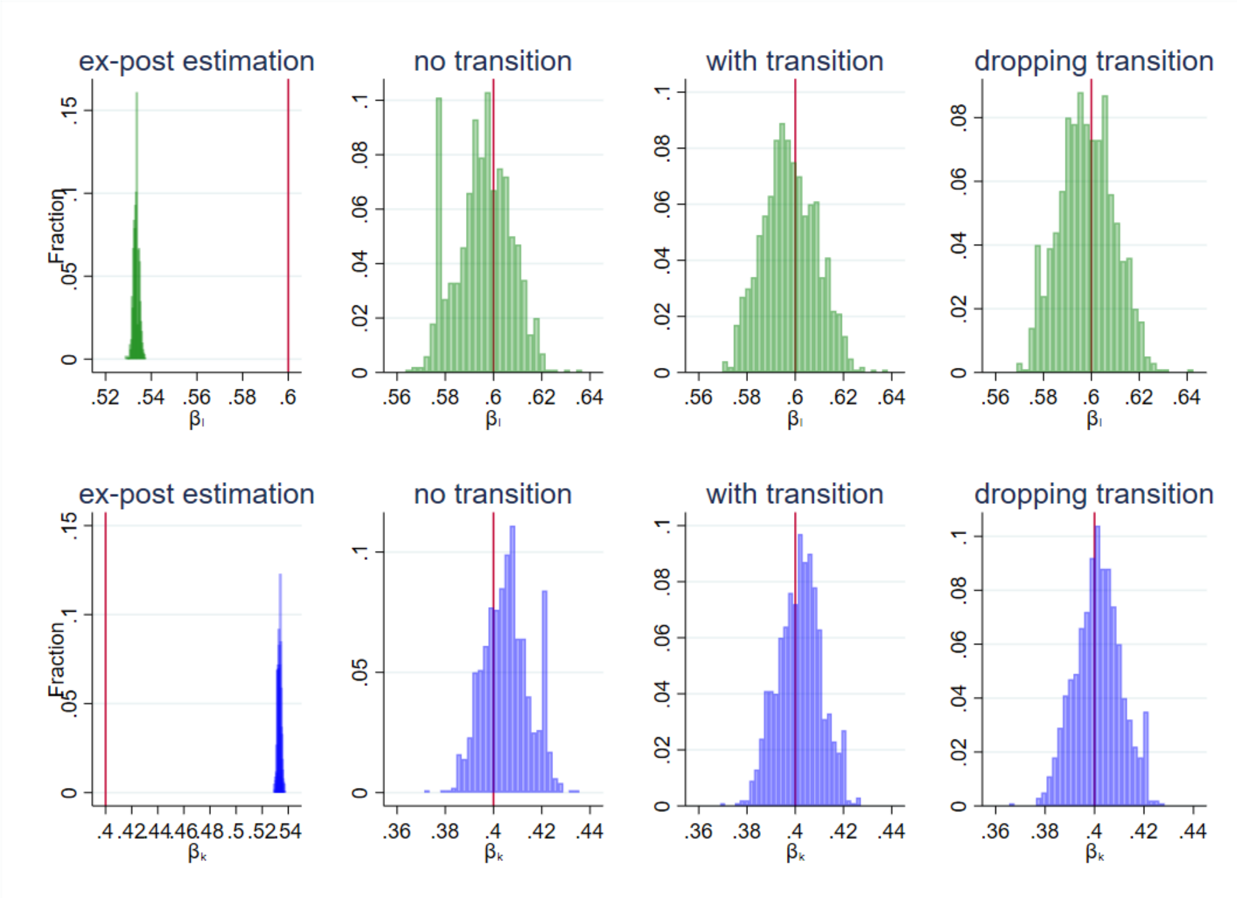


(b) Event-study Type Estimates of ATTs



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

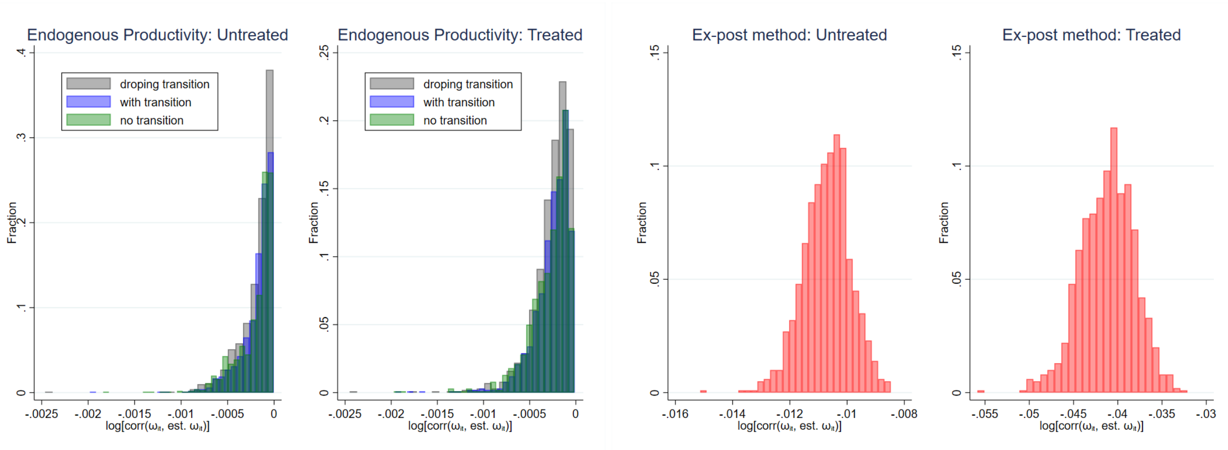
Figure 4: Production Function Parameter Estimates for Conditionally Exogenous Policy



Note: Vertical lines indicate the true values.

Figure 5: Correlations between Productivity Estimates and the True Productivity: Conditionally Exogenous Policy

(a) Endogenous Productivity (b) Exogenous Productivity



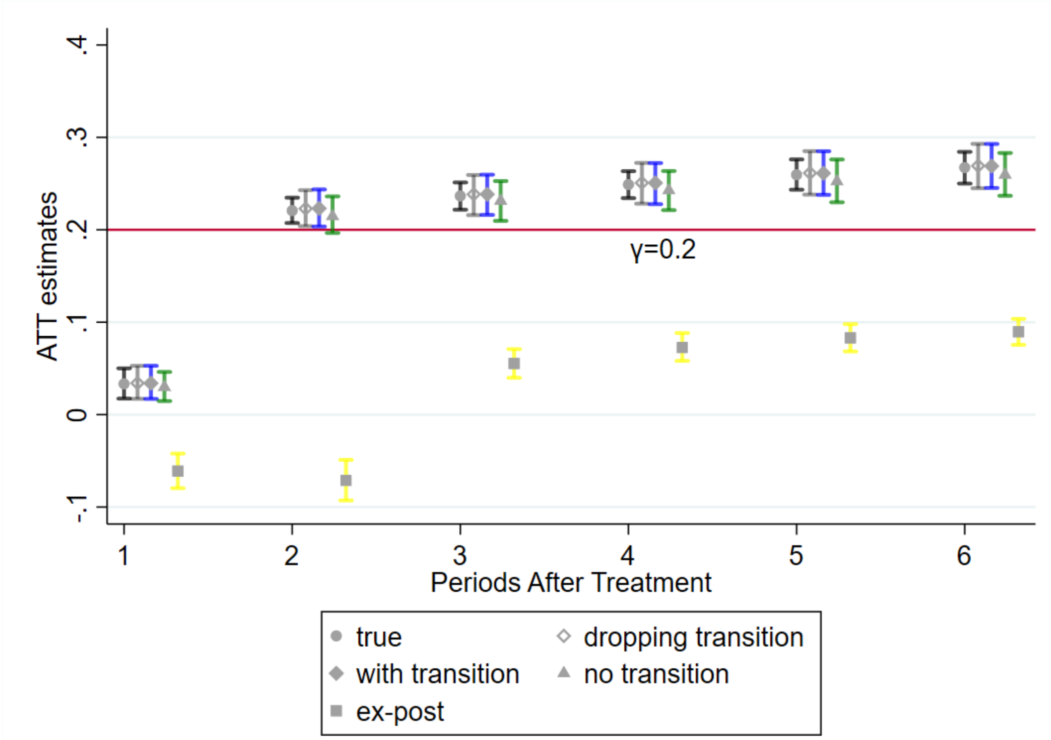
Note: Untreated units include never treated units and not-yet-treated units.

Table 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 6: 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.

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.

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.

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.

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.

Neyman, J. S. (1923). On the application of probability theory to agricultural experiments. essay on principles. section 9.(tlanslated and edited by dm dabrowska and tp speed, statistical science (1990), 5, 465-480). *Annals of Agricultural Sciences 10*, 1–51.

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.

Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association 100*(469), 322–331.

Rust, J. (1987). Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica: Journal of the Econometric Society*, 999–1033.

Shenoy, A. (2021). Estimating the production function under input market frictions. *Review of Economics and Statistics 103*(4), 666–679.

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 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.

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

A.1.1 Production Function and Potential Productivity Shocks

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

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

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:

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

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

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 $(\epsilon_{it}^0, \epsilon_{it}^1) \sim \mathbf{N}(0, 0.3\mathbb{I})$.

A.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:

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

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 between 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 (A.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:

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

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.[16] After t_0, the productivity evolution is rendered to be (A.3).

A.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:

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

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.

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

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 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^0_{it} and ω^1_{it} between sub-periods in time periods:

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

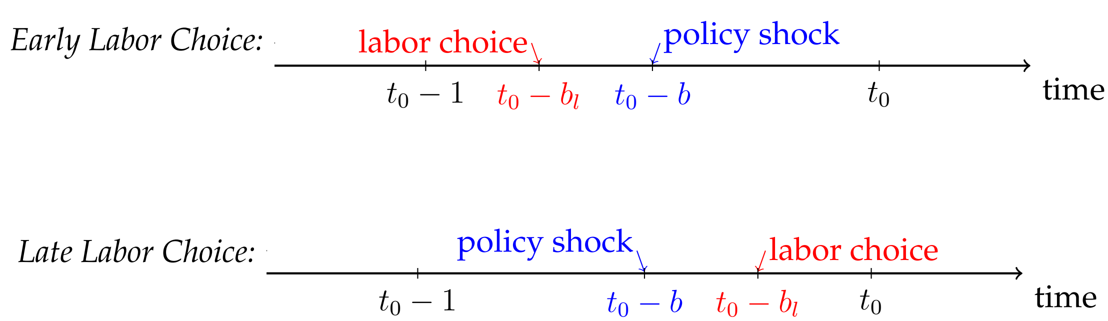
$$\omega^d_{it} = \rho_d^{b_l}\omega_{it-b_l}(d) + b_l \times \gamma \times d + \underbrace{\sqrt{1-\rho_d^{2b_l}}\epsilon^d_{it}}_{\epsilon^d_{it}(A)}, \text{ for } d \in \{0,1\},\ t \neq t_0 \quad \text{(A.8)}$$

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

$$\omega^d_{it} = \rho_d\omega_{it-1}(d) + \gamma \times d + \rho_d^{b_l}\sqrt{1-\rho_d^{2-2b_l}}\epsilon^d_{it-b_l} + \sqrt{1-\rho_d^{2b_l}}\epsilon^d_{it},\ t \neq t_0 \quad \text{(A.9)}$$

Note that Equation (A.9) is consistent with the AR(1) coefficient in (A.2) and (A.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 $Var(\epsilon^d_{it-b_l}) = Var(\epsilon^d_{it})$ so that $Var(\rho_d^{b_l}\sqrt{1-\rho_d^{2-2b_l}}\epsilon^d_{it-b_l} + \sqrt{1-\rho_d^{2b_l}}\epsilon^d_{it}) = Var(\epsilon^d_{it})$. 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 A.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 A.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:

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

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

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

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

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

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:

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

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

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

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

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

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

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:

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

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 $\hat{\omega}_{it-b_l} \equiv \omega_{it-b_l} - \frac{\gamma b_l}{1-\rho_1^{b_l}}$.[17]

A.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:

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

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

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 (A.3) as:

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

We can also re-write the production function as:

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

As a result, we can obtain the investment rule for $t \geq t_0$ from (A.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:

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

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 A.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.

B 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:

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

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}:

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

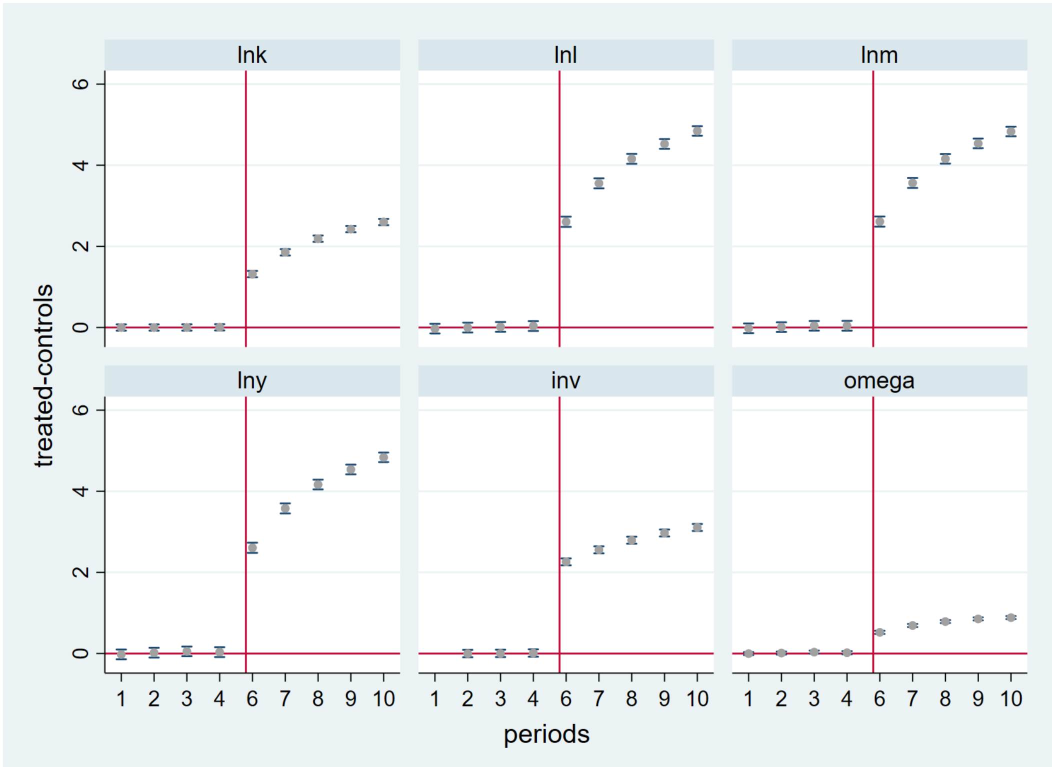
To estimate the treatment effects of productivity, the productivity process is incorporated into the standard production function estimation methods including Levinsohn and Petrin (2003), Wooldridge (2009), and Ackerberg et al. (2015). For the choice of instruments and algorithms, we refer to Kim et 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:

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

C Auxiliary Simulation Results

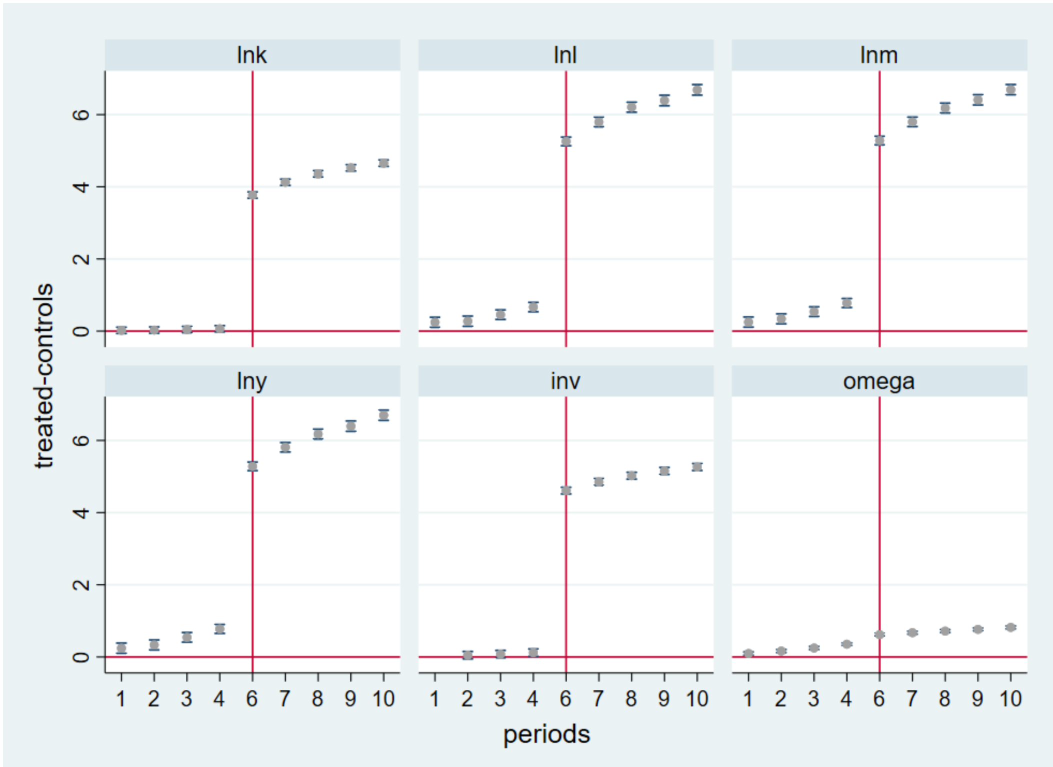
Figure C.1 and Figure C.2 plot the pattern of simulated data for strictly exogenous policy and conditional exogenous policy, respectively.

Figure C.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 C.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.