Identifying Treatment Effects on Productivity: Theory with An Application to Production Digitalization*

Zhiyuan Chen† Moyu Liao‡ Karl Schurter§

September 3, 2023

Abstract

We study the identification and estimation of treatment effects on the productivity of firms. Our approach embeds standard methods of production function estimation into a dynamic potential outcome framework. This new framework clarifies the necessary assumptions and potential pitfalls when quantifying causal effects on productivity. Our methods can be applied under weaker assumptions than those have been previously employed in the literature and do not require a solution to the firm's dynamic optimization problem. We apply our method to study the effect of production digitalization on productivity growth. Our results robustly show that the average treatment effect of production digitalization is not significant in a window of five years after production digitalization. However, we find substantial heterogeneity in the impact of production digitalization on productivity across time and industries. Importantly, firms with lower productivity before production digitalization tend to receive less productivity gains as time evolves.

Keywords: dynamic treatment effects; productivity; potential outcomes

JEL: C10, C51, L11, L20

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

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

‡School of Economics, University of Sydney. Email: moyu.liao@sydney.edu.au.

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

1 Introduction

Researchers have long been interested in quantifying the effect of an investment or intervention on a firm's productivity.[1] A natural two-step approach would be to first estimate the firm's productivity and then compare this with an estimate of what its productivity would have been in the counterfactual world absent the change. Problematically, however, issues can quickly arise if one simply borrows one of the typical methods of estimating production functions and feeds the estimated productivities into a standard policy evaluation method for estimating treatment effects. In general, the issue is that both of these procedures rely on distinct sets of assumptions that may be incompatible with each other, leading to incorrect inferences about the causal effects (De Loecker and Syverson, 2021). In this paper, we propose a method of estimating causal effects on productivity that fits the general two-step description but adapt existing methods to ensure that the assumptions invoked to estimate the realized and counterfactual productivities are consistent and sufficient to identify the treatment effect.

Specifically, when estimating productivities from firm- or plant-level data, researchers typically assume that productivity follows a Markov process (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg et al., 2015; Gandhi et al., 2020). Meanwhile, when estimating treatment effects on productivity, the firm's productivity in the treated and untreated states would typically be modeled as potential outcomes. If the firm's potential productivities in the treated and untreated states follow Markov processes, then the realized productivity may not be Markovian. For example, if the intervention of interest is the adoption of a new production technology, the plant's potential productivities with and without the new technology might be modeled as independent Markov processes. In the period in which the firm first adopts the technology, the firm's realized productivity will be its treated productivity, whose distribution depends on the previous treated productivity as opposed to the previous untreated productivity that was realized in the data. As a result, the estimated productivity will be biased if the researcher employs standard methods which assume the sequence of realized productivities is Markovian. We show that one can simply restrict attention to periods in which the plant remained treated or untreated in sequential periods in order to estimate the production function and the realized productivity in each period. In fact, one can separately estimate production functions in

[1] See excellent literature reviews by Bartelsman and Doms (2000) and Syverson (2011). Empirical studies come from 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), and public economics (Liu and Mao, 2019).

the treated and untreated states in order to allow for some factor-bias in the intervention.

On the other hand, given consistent estimates of the plant's realized productivity in each period under observation, the researcher's remaining task is to estimate the "missing counterfactual," the potential productivity that was not realized in the data. If the intervention is purely exogenous, then a standard difference-in-differences approach can be used to estimate the average treatment effect. More typically, however, the decision of when to begin exporting or when to adopt a new technology is likely to be endogenous, and the standard parallel trends assumption is likely to fail. But the structural assumptions invoked to estimate the productivities can be redeployed to easily solve this selection issue. Namely, the fact that untreated productivity follows a Markov process implies that an untreated firm or plant can be matched to a treated one with the same realized productivity in the period before it was treated in order to fill in the missing counterfactual.

In comparison with earlier work, our approach is both more generally applicable and more narrowly focused on estimating the treatment effect. It is more general in the sense that earlier work estimated returns to research and development or to exporting by assuming realized productivity follows a controlled Markov process (e.g., De Loecker, 2013; Doraszelski and Jaumandreu, 2013; Chen et al., 2021). This assumption is more restrictive than the assumption we adopt and may not be satisfied if, for example, potential producitivities follow independent Markov processes. At the same time, our work is more narrowly focused because we do not attempt to estimate and identify all the features of the model. We are only interested in estimating a treatment effect, e.g., the average treatment on the treated where the treatment would be defined as actively investing in R&D or exporting. As a result, we do not have to solve the firm's dynamic optimization problem or identify the entire productivity process in order to measure, for example, the returns to R&D; we only need to identify the mean productivity in the next period conditional on current productivity when the treatment status does not change.

Because our goals and requirements are more focused, we opt not to solve the firm's dynamic optimization problem as others in the empirical industrial organization literature have done when faced with the same challenge. Our approach is more typical of the dynamic potential outcomes literature, in which the treatment selection rule is left unspecified except for some timing assumptions. Namely, we assume that the firm chooses its treatment status for the current period before the realization its productivity shock, but we allow the firm to select into or out of treatment based on its previous treated and untreated potential productivities. Thus, the firm's choice of treatment status is modeled in the same way that capital is in most of the production function estimation literature.

Our work also relates to earlier work that uses regression-based methods to estimate causal effects on productivity (e.g., Pavcnik, 2003; Amiti and Konings, 2007; Yu, 2015; He et al., 2020). In this approach, the firms' productivities are estimated in a first step that ignores the variation in the treatment status. In the second step, regression methods are used to estimate the average effect of a policy on productivity. We argue that because this procedure only yields consistent estimates of an average treatment effect under the relatively strong assumptions that realized productivity is Markovian and treatment is exogenously assigned, our more robust methodology provides a useful alternative.

The rest of the paper is organized as follows. In Section 2, we formally introduce a model of a firm that uses capital, labor, and intermediate inputs to produce. Output is further affected by the treatment status in the current period and a Hicks-neutral productivity factor. The firm's realized productivity is equal to one of the treated or untreated potential productivities depending on the firm's treatment status. The potential productivities are assumed to follow a Markov process that generalizes the assumptions typically used in the literature. The key restriction is that the counterfactual productivity in the previous period is independent of the current productivity if the treatment status does not change, but we do not restrict the evolution of the potential productivities in the period in which the treatment status changes. This allows us to accommodate a wide range of plausible scenarios. As previously mentioned, the potential productivities might evolve independently of one another. Alternatively, the potential productivities might follow parallel paths or the treated productivity path might branch from the untreated productivity path in the period in which the firm selects into treatment. The researcher does not need to take a stand about the nature of the treatment in order to estimate the treatment effect.

In Section 3, we review the identification of the production function using Gandhi et al. (2020) and Ackerberg et al. (2015) in the setting in which treatment status does not change. We then show how the moment conditions must be modified to allow for variation in treatment status. Here, the key assumption that enables the proposed method is that treatment is selected prior to the realization of the productivity shock. Otherwise, the assumptions and data requirements are analogous to those of Gandhi et al. (2020) and Ackerberg et al. (2015): we require panel data with at least two periods and many firms. In addition, we must observe many firms that remain untreated in consecutive periods, others that remain treated in consecutive periods, and a third group that switches from the untreated to the treated state in order to identify the production function and treatment effect. In an appendix, we discuss an alternative assumption and moment conditions that might be used if this assumption on firm types does not appear to be satisfied in the data.

In Section 4, we discuss the identification of the average treatment effect on the treated. The average treatment effect is generally not identified without stronger structural assumptions, and we thus discuss its identification in the appendix. In contrast to the literature on dynamic treatment effects (Heckman and Navarro, 2007; Abbring and Heckman, 2007; Vikström et al., 2018; Sun and Abraham, 2021), the outcome of interest is not directly observed, and additional structural assumptions are needed to infer it from data. Apart from this distinction, our approach builds on the dynamic potential outcome framework. As has been observed in the literature, one must be careful when defining and identifying treatment effects in a dynamic potential outcome framework because firms who are treated in one period may return to the untreated state but continue on an altered trajectory as a result of their temporary treatment assignment. We do not add to this discussion, but acknowledge the complexities involved. For the sake of simplicity, we focus on the case of an absorbing treatment state in which firms remain forever after they first select into treatment. Accordingly, our identification and estimation results target the ℓ-period ahead average treatment effect on the treated, which answers the question of how much more or less productive a firm is ℓ periods after it is initially treated compared to what its productivity would have been if it had remained untreated the entire time.

In Section 6, we use our methodology to estimate the productivity effects of production digitalization in the manufacturing sector of China. Recent work has struggled to find evidence in aggregate production statistics of any productivity gains associated with the rise of new production technologies (Brynjolfsson et al., 2017). In contrast with our approach, existing research relies on reduced-form regressions on the productivity estimates to detect the impact of the adoption of new technologies (Draca et al., 2009; Gal et al., 2019, among others). Our structural analysis of firm-level data shows that, in a window of 5 years, the average productivity gains from production digitalization are positive in the first three periods after production digitalization but are negative in later periods. However, the productivity effects are not statistically significant. Moreover, the productivity effects of digitalization varies substantially across industries, with industries using more complex production processes, like equipment, electronics and healthcare, tend to receive more productivity gains. The results on the firm-specific productivity effects reveal that the effects of digitalization on productivity become more dispersed as time evolves. Importantly, we find that firms with lower ex-ante productivity receive less productivity gains as time goes by. Notably, using the same dataset, we show that the simple regression of the productivity process leads to the finding of significantly negative impact of production digitalization on productivity. In general, our empirical results support the view that new digital technologies have unequal effects on firms' productivity, depending on the firm's characteristics such as managerial practices (Bloom et al., 2012) and complementary intangible assets (Bresnahan et al., 2002; Brynjolfsson et al., 2021).

Finally, Section 7 concludes.

2 The Econometric Framework

2.1 A Firm Model with Treatment and Potential Productivity

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

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

where Q_{it} is the output quantity, ω_{it} is the realized productivity, η_{it} is some ex-post shock of productivity that is not known when a firm make current period input choices, K_{it} is the capital, L_{it} is the labor, M_{it} is the material, D_{it} is the treatment, and $\boldsymbol{\beta}$ is the parameter vector. The dimension of $\boldsymbol{\beta}$ can be infinite when the production function is nonparametric. Moreover, $\boldsymbol{\beta}$ can also include a set of time dummies to account for a secular trend in the production function (e.g., Doraszelski and Jaumandreu (2013)). Note that we allow the treatment D_{it} as an input factor,[2] which captures possible impacts on managerial efficiencies (Chen et al., 2021).

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

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

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

Definition 1. *(Treatment switching indicator) We define a treatment regime change indicator*

[2] 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.

$G_{it} \in \{-1, 0, 1\}$: (1) Positive regime change: $G_{it} = 1$ if $D_{it} - D_{it-1} = 1$; (2) Unchanged regime: $G_{it} = 0$ if $D_{it} - D_{it-1} = 0$; (3) Negative regime change: $G_{it} = -1$ if $D_{it} - D_{it-1} = -1$.

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

$$\begin{aligned} \omega_{it}^1 &= \mathbb{1}(G_{it} = 0)\bar{h}_1(\omega_{it-1}^0, \omega_{it-1}^1) + \mathbb{1}(G_{it} = 1)h^+(\omega_{it-1}^0, \omega_{it-1}^1) + \epsilon_{it}^1, \\ \omega_{it}^0 &= \bar{h}_0(\omega_{it-1}^0, \omega_{it-1}^1), \end{aligned} \tag{3}$$

where h^+ is the transition function of ω_{it}^1 when the regime switch is positive . We focus on a baseline model where the treatment is absorbing and firms can only switch from not-treated to treated status. Later, we will consider a non-absorbing treatment case where we can also define a function h^- to govern the transition out of treatment. We compare our Markov process of potential productivity (3) with a traditional Markov process of realized productivity approach in Section 5, and show that (3) can accommodate more empirical contexts and has a different causal interpretation. We can even allow the evolution at the transition process h^+ to possibly depend on i but impose the same evolution process when the treatment variable is constant, but we abstract away from this case to avoid complicated notation.[3] Furthermore, the Markovian productivity process (3) is diagonal whenever there is no treatment status change:

Assumption 2.1. *(Diagonal Markov Process) The function $\bar{h}_d$ depends only on ω_{it}^d, so we may abuse notation to rewrite*

$$\bar{h}_d(\omega_{it}^0, \omega_{it}^1) = \bar{h}_d(\omega_{it}^d),$$

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

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

[3]For example, different firms may have different regime switch time within a year, which may lead to the difference in h_i^+.

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* $\bar{h_1} = h^+$*, and* $\bar{h}_1(\omega) = \bar{h}_0(\omega - C) + C$*. These conditions lead to* $\omega_{it}^1 = \omega_{it}^0 + C$ *almost surely for all* i, t*.*

Example 2. *(Divergence of Productivity when Treatment Diverges.) Consider a case where the binary treatment represents whether a firm invests in R&D. If a firm chooses to switch from not investing in R&D at* t *to investing in R&D at* $t+1$*, then only* ω_{it}^0 *matters for the determination of* ω_{it+1}^1*. In this case, only the observed potential outcome before the regime switching matters for the productivity process. This model can be captured by imposing* $h^+(\omega_{it}^0, \omega_{it}^1) = h^+(\omega_{it}^0)$*. Essentially, we are imposing* $\omega_{it}^1 \equiv \omega_{it}^0$ *for all pre-treatment periods* t*.*

Example 3. *(Independent Productivity Evolution Process) In some cases, a firm needs to choose between two types of technologies. Each technology evolves without being influenced by the other technology. Firms can choose which technology to use. In this case,* $\bar{h}_1 = h^+$*.*

Firms' Behavior and Timing of Firms' Decisions

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

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

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

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

Firms make two types of choices at the time t. First, given the realized productivity ω_{it} and pre-determined inputs, firms choose the static inputs to maximize its short-run revenue. Then, firms choose the next period pre-determined inputs and possibly the treatment status D_{t+1} in period $t+1$ given a vector of state variables $S_{it} \equiv (K_{it}, L_{it}, D_{it}, \omega_{it}^1, \omega_{it}^0, \zeta_{it})$, where ζ_{it} is an idiosyncratic cost shock relevant the dynamic decision[4]. We summarize the

[4] For example, ζ_{it} can be an idiosyncratic cost of taking treatment.

firm-decision timeline in the following graph and define the firms' information set correspondingly.

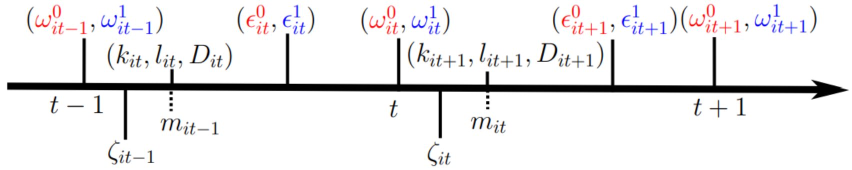


Figure 1: Timeline for firm's decision.

Definition 2. *When deciding the* $(K_{it+1}, L_{it+1}, D_{it+1}, M_{it})$*, firm* i*'s time-*t *information set is given by*

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

In the case of externally assigned treatment, our firm behavioral 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, firms can choose treatment status based on the potential productivity values and this leads to selection of D_{it+1} on firms' information set $\mathcal{I}^F_{it}$. Our firm model allows for the existence of unobserved dynamic cost shock ζ_{it}. This additional unobserved heterogeneity can bring additional difficulty of identifying the treatment effect on productivity. This is different from the endogenous productivity literature (Aw et al., 2011; De Loecker, 2013; Doraszelski and Jaumandreu, 2013; Peters et al., 2017) who are interested in the productivity differences between treatment takers and non-treatment takers.

2.2 Treatment-Effect Objects

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

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

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

When a firm switch its treatment status at time t, it materializes two effects: The trend effects $\bar{h}_1(\cdot) - \bar{h}_0(\cdot)$, which accumulates over time, and the one-time effect $\omega_{it}^1 - \omega_{it}^0$. 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.

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 and has to identify it first.

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

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

3.1 Recovering the Productivity in the Absence of Treatment

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

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

$$\mathbb{E}\left[s_{it} - \log\left(\frac{\partial f_0(k_{it}, l_{it}, m_{it}; \boldsymbol{\beta})}{\partial m_{it}}\right) \Big| k_{it}, l_{it}, m_{it}\right] = 0 \quad \forall t = 1, ..., T, \tag{4}$$

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:

$$\mathbb{E}[\omega_{it}(\boldsymbol{\beta}) - h(\omega_{it-1}(\boldsymbol{\beta}))|\{k_{it}, l_{it}, k_{it-1}, l_{it-1}, m_{it-1}\}] = 0 \quad \forall t = 1, ..., T, \tag{5}$$

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

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

$$\Phi_{it-1}(k_{it-1}, l_{it-1}, m_{it-1}) \equiv \mathbb{E}[q_{it-1}|k_{it-1}, l_{it-1}, m_{it-1}], \tag{6}$$

and use the moment condition

$$E\left[\omega_{it}(\boldsymbol{\beta}) - h\left[\Phi_{it-1}(k_{it-1}, l_{it-1}, m_{it-1}) - f_0(k_{it-1}, l_{it-1}; \beta)\right] \mid \{k_{it}, l_{it}, k_{it-1}, l_{it-1}, m_{it-1}\}\right] = 0. \tag{7}$$

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 (4) and (5) identify the gross production function $\boldsymbol{\beta}$ nonparametrically up to a constant difference; (2) The moment conditions (6) and (7) identify the value-added production function $\boldsymbol{\beta}$ nonparametrically up to a constant difference. Moreover, the h is identified nonparametrically in both the GNR and ACF cases.*

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

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

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

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

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

3.2 Recovering the Productivity with Variations in Treatment Status

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

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

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

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

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

Theorem 3.1. *Suppose Assumptions 2.1- 3.3 hold. The moment condition (4) (and respectively (6)) and*

$$\mathbb{E}[\omega_{it}(\boldsymbol{\beta}) - \bar{h}_0(\omega_{it-1}(\boldsymbol{\beta}))|\mathcal{Z}_{it}, D_{it} = D_{it-1} = 0] = 0, \quad (8)$$

$$\mathbb{E}[\omega_{it}(\boldsymbol{\beta}) - \bar{h}_1(\omega_{it-1}(\boldsymbol{\beta}))|\mathcal{Z}_{it}, D_{it} = D_{it-1} = 1] = 0, \quad (9)$$

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

Proof. We first look at equation (8), and the proof of expression (9) follows similarly. We can write

$$\begin{aligned} &\mathbb{E}[\omega_{it}(\boldsymbol{\beta}) - \bar{h}_0(\omega_{it-1}(\boldsymbol{\beta}))|\mathcal{Z}_{it}, D_{it} = D_{it-1} = 0] \\ &= \mathbb{E}[\omega_{it}^0(\boldsymbol{\beta}) - \bar{h}_0(\omega_{it-1}^0(\boldsymbol{\beta}))|\mathcal{Z}_{it}, D_{it} = D_{it-1} = 0] \\ &= \mathbb{E}[\epsilon_{it}^0|\mathcal{Z}_{it}, D_{it} = D_{it-1} = 0] = 0 \end{aligned} \tag{10}$$

where $\omega_{it}^0(\boldsymbol{\beta})$ denotes the potential productivity without treatment, recovered under parameter value $\boldsymbol{\beta}$ and $D_{it} = 0$. The first equality of (10) holds by the potential outcome equation and the last equality holds by Assumption 3.2. The moment condition (8) is well defined by Assumption 3.3. By Lemma 3.1, the result follows. □

The non-identification of the scale of the production function can contaminate the identification of treatment effects. Namely, 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 the ω_{it}^1 recovered under $(F(\cdot, 1; \boldsymbol{\beta}), \bar{h}_1)$ from the $\tilde{\omega}_{it}^1$ recovered under $(e^{c_1}F(\cdot, 1; \boldsymbol{\beta}), \tilde{h}_1)$. In particular, we have $\tilde{\omega}_{it}^d = \omega_{it}^d - c_d$.[5] As a result, we must normalize the scale of production functions before and after the treatment to interpret the treatment effect on productivity.

When treatment status does not vary, this normalization is innocuous because it merely shifts the location of the log-productivity distribution. Thus, one might choose to normalize the scale of the production function such that the mean log-productivity is zero. When treatment status varies, however, centering the log-productivity distribution in both the treated and untreated states complicates the interpretation of the estimated treatment effect. For instance, the estimated average treatment effect would be zero if firms were randomly selected into a treatment that increases productivity by a fixed percentage because the mean productivity in both states was "normalized" to zero. Instead, we suggest normalizing the scale of the production function.[6]

To implement the moment conditions in Theorem 3.1, we must discard the transition periods, which can be inefficient if the transition periods consist of large part of the data.

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

[6]One particular choice is to compute the industry mean of inputs across all periods, $(\bar{K}, \bar{L}, \bar{M})$, and impose $F(\bar{K}, \bar{L}, \bar{M}, 1, \boldsymbol{\beta}) = F(\bar{K}, \bar{L}, \bar{M}, 0, \boldsymbol{\beta})$. If we impose the production functions before and after the treatment to be the same, then there is no need to additionally normalize the scale.

We therefore propose some additional moment conditions that use the transition periods under special empirical contexts, see details in Appendix C.1.

Other Structural Objects In this firm model, there are many other interesting structural objects, such as the transition evolution function h^+ . This structural object is generally not identified under the assumptions in Theorem 3.1: The transition evolution function h^+ are not identified because we cannot observe ω_{it}^1 and ω_{it}^0 simultaneously.

3.3 A Revisit to Existing Methods

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

We consider a simple "difference-in-difference" policy context: An exogenous policy shock happens at $t = T_0 + \Delta$ for $\Delta \in (0, 1)$. A random subset of firms is influenced by the policy while others are not, and firms are separated into treated and control groups. For the firms in the controlled group, $D_{it} = 0$ for all t. In this context, the policy variable D_{it} is fully exogenous to the productivity process.[7] For the firms in the treated group, $D_{it} = \mathbb{1}(t \geq T_0 + 1)$.

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

The Ex-post Regression

The ex-post regression method consists of two steps: First, it estimates the firm model ignoring the existence of policy effect. To do so, it estimates the production function parameter $\boldsymbol{\beta}$ and the evolution process h using (4) and (5). Second, given the estimated parameter $\hat{\boldsymbol{\beta}}$ and $\hat{h}$, recover the pseudo firm-level productivity $\hat{\omega}_{it} = q_{it} - f(k_{it}, l_{it}, m_{it}; \hat{\boldsymbol{\beta}})$. They analyze the individual treatment effect based on $\hat{\omega}_{it}$. For example, ex-post regression method runs a two-way fixed effect regression by treating $\hat{\omega}_{it}$ as the outcome variable.

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

The ex-post regression method bears the following problem: The trend difference $\bar{h}_1 \neq \bar{h}_0$ is ignored in this method when estimating the productivity, and consequently the moment equality (5) fails. To see this, we first note that for all pre-treatment period $t \leq T_0$, the moment equation (5) becomes

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

By Proposition 3.1, this moment condition identifies $\boldsymbol{\beta}$ and $\bar{h}_0$. We now derive the inconsistency of (5). For $t \geq T_0 + 2$, the moment condition (5) becomes

$$\begin{aligned}
(5) &= \mathbb{E}[\omega_{it}(\boldsymbol{\beta}) - \bar{h}_0(\omega_{it-1}(\boldsymbol{\beta}))|\mathcal{Z}'_{it}] \\
&=_{(1)} \mathbb{E}[\omega^0_{it}(\boldsymbol{\beta}) - \bar{h}_0(\omega^0_{it-1}(\boldsymbol{\beta}))|\mathcal{Z}'_{it}, D_{it} = 0]Pr(D_{it} = 0) \\
&+ \mathbb{E}[\omega^1_{it}(\boldsymbol{\beta}) - \bar{h}_0(\omega^1_{it-1}(\boldsymbol{\beta}))|\mathcal{Z}'_{it}, D_{it} = 1]Pr(D_{it} = 1) \\
&=_{(2)} \underbrace{\mathbb{E}[\omega^0_{it}(\boldsymbol{\beta}) - \bar{h}_0(\omega^0_{it-1}(\boldsymbol{\beta}))|\mathcal{Z}'_{it}]}_{\text{Part A}} Pr(D_{it} = 0) + \underbrace{\mathbb{E}[\omega^1_{it}(\boldsymbol{\beta}) - \bar{h}_0(\omega^1_{it-1}(\boldsymbol{\beta}))|\mathcal{Z}'_{it}]}_{\text{Part B}} Pr(D_{it} = 1)
\end{aligned} \tag{11}$$

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 (11) 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 (5) fails for all $t \geq T_0 + 2$.

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

The Endogenous Productivity Method

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

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

This method solves the misspecification of the productivity process for treated and controlled group. Indeed, by defining $\bar{h}_d(\cdot) = \tilde{h}(\cdot, d)$ for $d = 0, 1$, we can show that moment condition (5) can be transforms to

$$\begin{aligned} &\mathbb{E}[\omega_{it}^0(\boldsymbol{\beta}) - \bar{h}_0(\omega_{it-1}^0(\boldsymbol{\beta}))|\mathcal{Z}_{it}] = 0 \quad \forall t \leq T_0, \quad \text{and} \\ &\mathbb{E}[\omega_{it}^0(\boldsymbol{\beta}) - \bar{h}_0(\omega_{it-1}^0(\boldsymbol{\beta}))|\mathcal{Z}_{it}, D_{it} = D_{it-1} = 0]Pr(D_{it} = D_{it-1} = 0) \\ +&\mathbb{E}[\omega_{it}^1(\boldsymbol{\beta}) - \bar{h}_1(\omega_{it-1}^1(\boldsymbol{\beta}))|\mathcal{Z}_{it}, D_{it} = D_{it-1} = 1]Pr(D_{it} = D_{it-1} = 1) \quad \forall t \geq T_0 + 2, \end{aligned} \tag{12}$$

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

$$\begin{aligned} &\underbrace{\mathbb{E}[\omega_{iT_0+1}^0(\boldsymbol{\beta}) - \bar{h}_0(\omega_{iT_0}^0(\boldsymbol{\beta}))|\mathcal{Z}'_{iT_0+1}, D_{iT_0+1} = D_{iT_0} = 0]}_{\text{Part A}} Pr(D_{iT_0+1} = D_{iT_0} = 0) + \\ &\underbrace{\mathbb{E}[\omega_{iT_0+1}^1(\boldsymbol{\beta}) - \bar{h}_1(\omega_{iT_0}^0(\boldsymbol{\beta}))|\mathcal{Z}'_{iT_0+1}, D_{iT_0+1} = 1, D_{iT_0} = 0]}_{\text{Part B}} Pr(D_{iT_0+1} = 1, D_{iT_0} = 0) = 0. \end{aligned} \tag{13}$$

Moment conditions (12) are correctly specified. In particular, by Proposition 3.1, $\boldsymbol{\beta}$, $\bar{h}_0$ are identified from the $t \leq T_0$ moment equality (12), and $\bar{h}_1$ is identified from the $t \geq T_0 + 2$ moment equality (12).

However, the moment condition at the regime switching period (13) is misspecified. Let $\bar{h}_0$ be identified from (12). Part A in (13) equals zero. However, the Part B may not equal zero. Given the evolution process (3), the transition process at the positive regime switching period should be $h^+(\omega_{it-1}^1, \omega_{it-1}^0)$, where in the Part B of (13), 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 $h^+ = \bar{h}$. Therefore, the Part B of (13) becomes $\mathbb{E}[\omega_{iT_0+1}^1(\boldsymbol{\beta}) - \omega_{iT_0}^0(\boldsymbol{\beta})|\mathcal{Z}_{iT_0+1}, D_{iT_0+1} = D_{iT_0} = 1]$. The moment value of the Part B is C at the true production parameter rather than 0, so the model is misspecified.

In Example 2, the evolution at the transition period only depends on the observed outcome in the last period. If we impose $h^+ = \bar{h}_1$, then the Part B of (13) equals zero and the model is not misspecified. However, this is a strong assumption and may not be satisfied in some empirical contexts. Let's consider the regime switch happens at $T_0 + \Delta$ for some $\Delta < 1$. In this case, $\omega_{iT_0}^0$ first evolves to $\omega_{iT_0+\Delta}^0$ under the controlled process $\bar{h}_0$, and then the treatment status changes and the productivity evolves from $\omega_{iT_0+\Delta}^0$ to $\omega_{iT_0+1}^1$. In other words, the productivity only enjoys the benefit of the policy effects during the period $[T_0 + \Delta, T_0 + 1]$. If the policy variable D_{iT_0+1} affects the productivity process at the beginning of the period, then it is likely that $\bar{h}_1 \neq h^+$.

4 Evaluating the Treatment Effect on Productivity

Recall that the treatment effect of interest are given in Definition 3. Since we only observe a firm either in the treated or non-treated state, the individual treatment effect $\omega_{it}^1 - \omega_{it}^0$ is typically not identified. Therefore, we focus on the average treatment effect on the treated (ATT). Identifying average treatment effect (ATE) is generally difficult and requires more structural assumptions. We instead discuss the idenfitification of ATE in Appendix A.2.

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

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

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

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

For ATT, we find it instructive to discuss the identification for absorbing treatment and non-absorbing treatment, separately.

4.1 ATT: Absorbing Treatment

The absorbing treatment is at the core of literature on estimating dynamic treatment effects (Sun and Abraham, 2021; Athey and Imbens, 2022). As a benchmark for analyzing ATT, we consider the absorbing policy for which the treatment indicator is non-decreasing $D_{it-1} \leq D_{it}$. For any treatment that is not absorbing, we can replace the treatment status D_{it} with an indicator for ever having received the treatment to obtain a new treatment being absorbing.[8]

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

Let $e_i > 1$ be the first period that firm i starts to receive treatment.Since the treatment is absorbing, when the 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 subgroup of firms whose treatment effects of interest to us, and $\ell \geq 0$ be the time relative to the first treatment period. It is helpful to think of group g as a cohort of treatment, and we may be interested in the treatment effect for different cohort. The ℓ-period-ahead ATT at time t for group g is given by

$$ATT_{g,\ell} = \mathbb{E}[\omega_{it}^1 - \omega_{it}^0 | t = e_i + \ell, i \in g]. \tag{14}$$

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. We first look at a simple parallel trend assumption that is needed in the Diff-in-Diff analysis:

Assumption 4.1. *(Simple Parallel Trend) The following condition is the simple parallel trend condition:*

$$\mathbb{E}[\omega_{it}^0 - \omega_{it-1}^0 | e_i = t] = \mathbb{E}[\omega_{it}^0 - \omega_{it-1}^0 | e_i > t]. \tag{15}$$

If condition (15) holds, then the $ATT_{g,0}$ is identified as $\mathbb{E}[\omega_{it} | e_i = t] - \mathbb{E}[\omega_{it-1} | e_i = t] - (\mathbb{E}[\omega_{it} | e_i > t] - \mathbb{E}[\omega_{it-1} | e_i > t])$. However, Assumption 4.1 is a high-level condition because it is imposed on the potential productivity before and after the treatment and can be hard to justify. To see it, note that from the productivity process (3), we can derive that:

$$\begin{aligned} \text{positive switchers: } \mathbb{E}[\omega_{it}^0 - \omega_{it-1}^0 | e_i = t] &= \mathbb{E}[\bar{h}_0(\omega_{it-1}^0) - \omega_{it-1}^0 | e_i = t], \\ \text{non switchers: } \mathbb{E}[\omega_{it}^0 - \omega_{it-1}^0 | e_i > t] &= \mathbb{E}[\bar{h}_0(\omega_{it-1}^0) - \omega_{it-1}^0 | e_i > t], \end{aligned} \tag{16}$$

where we use the condition (3.2) to derive (16). From (16) we see that the parallel trend condition can fail due to the selection into treatment: The treatment D_{it} can depend on the value of ω_{it-1}^0 and can correlate with the initial treatment time e_i. Consider Example 2 with a R&D decision, the firm chooses to invest in R&D only when ω_{it-1}^0 exceeds a certain level. In this case, D_{ie_i} is a function of ω_{it-1}^0, and (15) 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^0_{it-1} in equation (16), we have

$$\begin{aligned}\mathbb{E}[\omega^0_{it} - \omega^0_{it-1}|e_i = t, \omega^0_{it-1}] &= \bar{h}_0(\omega^0_{it-1}) - \omega^0_{it-1},\\ \mathbb{E}[\omega^0_{it} - \omega^0_{it-1}|e_i > t, \omega^0_{it-1}] &= \bar{h}_0(\omega^0_{it-1}) - \omega^0_{it-1}.\end{aligned} \tag{17}$$

The two equations in (17) coincide as a result of the assumptions used to estimate the production function and productivity process. We call this the conditional parallel trend assumption.

To clarify the meaning of the conditional parallel trend assumption, we introduce a more general evolution process for the ω^0_{it}:

$$\omega^0_{it} = \mathbb{1}(G_{it} = 0)\bar{h}_0(\omega^0_{it-1}) + \mathbb{1}(G_{it} = 1)h^+_0(\omega^0_{it-1}, \omega^1_{it-1}) + \epsilon^0_{it}. \tag{18}$$

This general framework allows the potential untreated productivity to have a different evolution process when firms switch the treatment status. Equation (17) is a direct consequence of imposing $h^+_0 = \bar{h}$.

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

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

In the baseline absorbing treatment case, the evolution process (3) imposes Assumption 4.2 directly because the conditional parallel trend assumption is quite natural: if the firm had not entered treatment then its counterfactual untreated productivity would of course have evolved according to $\bar{h}_0$.

When the treated state is not absorbing, however, the conditional parallel trend assumption is restrictive because the untreated productivity process may be realized again in the future. For example, if temporarily adopting a new technology destroys some intangible capital associated with the old technology, then the conditional parallel trend assumption would fail. In this case, it may be worthwhile to consider the more general process (18).

Identifying Treatment Effect Object Based on the implication of (17), we can study the following l-period-ahead conditional average treatment effect on the treated (CATT):

$$CATT_{g,l}(\omega) = E[\omega^1_{it} - \omega^0_{it}|t = e_i + l, \omega_{ie_i-1} = \omega, i \in g],$$

which further conditions on the pre-treatment realized productivity. By law of iterative expectation, $ATT_{g,l} = E[CATT_{g,l}(\omega_{ie_i-1})]$, and we can identify $ATT_{g,l}$ if $CATT_{g,l}(\omega)$ is identified.

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

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

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

where (1) by replacing ω^0_{it} with the evolution process and using Assumptions 3.2 and 4.2, (2) follows by the potential outcome (2) and $\omega_{ie_i-1} = \omega^0_{ie_i-1}$. Further take the expectation with respect to the treatment time e_i to get the result. □

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

For notation purpose, let $\bar{h}^\ell_0$ be the ℓ-period productivity transition process, we can write $\omega^0_{ie_i+\ell} = \bar{h}^{(\ell)}_0(\omega^0_{ie_i}, (\epsilon^0_{is})^{e_i+l}_{s=e_i})$. We now consider a strong constraint on the productivity shocks but relax the constraint on the shape of $\bar{h}_0$.

Assumption 4.3. *There is a group-time pair (g', s) such that all firms i' such that $i' \in g'$ are untreated by l-periods since time s, i.e. $e_{i'} > s + l$. Moreover, the conditional distribution of* $(\epsilon^0_{ie_i}, ..., \epsilon^0_{ie_i+l})|(i \in g, \omega^0_{is-1})$ *is the same as the conditional distribution of* $(\epsilon^0_{i's}, ..., \epsilon^0_{i's+l})|(i' \in g', \omega^0_{is-1})$.

We will discuss an example for the group g and g' later. The group g' firms serve as the controlled match for the treated firms in g. We require more than the conditional mean-independence of the future productivity shocks with respect to the treatment time. Assumption 4.3 allows for nonlinearity in $\bar{h}^{(s)}_0(\cdot)$.

Proposition 4.2. *Suppose Assumption 4.2 and 4.3 hold. Then the ℓ-period-ahead CATT is identified as* $CATT_{g,\ell}(\omega) = \mathbb{E}[\omega_{ie_i+\ell}|i \in g, \omega_{ie_i-1} = \omega] - \mathbb{E}[\omega_{is+\ell}|i \in g', \omega_{is-1} = \omega]$. *The corresponding ATT is identified as* $ATT_{g,\ell} = \mathbb{E}[CATT_{g,\ell}(\omega_{ie_i-1})|i \in g]$, *where the expectation is taken over the conditional distribution of ω_{ie_i-1} given $i \in g$.*

Proof. By the definition of CATT:

$$\begin{aligned} CATT_{g,\ell}(\omega) &=_{(a)} \mathbb{E}[\omega_{ie_i+\ell}|i \in g, \omega_{ie_i-1} = \omega] - \mathbb{E}[\bar{h}_0^{(\ell)}(\omega_{ie_i-1}, \epsilon^0_{ie_i}, ..., \epsilon^0_{ie_i+\ell})|i \in g, \omega_{ie_i-1} = \omega] \\ &=_{(b)} \mathbb{E}[\omega_{ie_i+\ell}|i \in g, \omega_{ie_i-1} = \omega] - \mathbb{E}[\bar{h}_0^{(\ell)}(\omega_{is-1}, \epsilon^0_{is}, ..., \epsilon^0_{is+\ell})|i \in g', \omega_{is-1} = \omega] \\ &=_{(c)} \mathbb{E}[\omega_{ie_i+\ell}|i \in g, \omega_{ie_i-1} = \omega] - \mathbb{E}[\omega_{is+\ell}|i \in g', \omega_{is-1} = \omega], \end{aligned} \tag{19}$$

where (a) follows by the productivity evolution process and the potential outcome equation, (b) follows by Assumptions 4.3, and (c) follows by the productivity evolution procedure for untreated firms. □

Proposition 4.2 requires us to match over the lagged productivity for each group g-firms with g'-firms since time s. This is because we cannot observe the untreated shocks ϵ^0_{it} for treated firms and the higher order moments of ϵ^0_{it} matters for the ℓ-period evolution process $\bar{h}_0^{(\ell)}$. We present an empirical context where the matching group g' can be found.

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

In this case, Assumption 4.3 hold under the following empirical context: Before g_0, no firms are treated. At time g_0, firms can decide whether to take the absorbing treatment. Between g_0 and $g_0 + l + 1$, firms cannot change their treatment status due to regulations or contracts.

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

However, if all firms can choose the initial treatment time freely after g_0, then Assumption 4.3 typically fails for the $g_0 + l + 1$-not-yet-treated firms: A firm chooses not to be treated until $g_0 + l + 1$ are likely the firms whose potential productivity $(\epsilon^0_{ie_i}, ..., \epsilon^0_{ie_i+l})$ are high and they are reluctant to switch to the treated status. In this case, we can choose $g' = \{all\ firms\}$ and $s = 2$ in Assumption 4.3 : We use all firms and periods before the initial treatment period g_0, and use the firms' initial productivity ω_{i1} as the match.

To identify $ATT_{g,l}$ from Proposition 4.2, we typically need to match a treated firm with a controlled firm with the same lagged productivity. This matching procedure can be hard to implement due to two reasons: (1) We may not be able to find a (g', s) pair that satisfies the independence restriction; (2) Even when (g', s) is found, we may not have enough observation in group-time pair (g', s). Moreover, if all firms are treated at $g_0 + l$, we cannot identify the ATT_{g_0+l+s} for all $s > 0$. We thus propose a stronger condition:

Assumption 4.4. *(i). The shocks satisfy $\epsilon^0_{is} \sim_{i.i.d.} G^0_\epsilon(\cdot)$, where the $i.i.d$ is over both firm index i and time index s. (ii). We can find a group-time pair (g', s) such that all firms in g' are untreated in period-s. (iii). There is no selection in shocks: $\epsilon^0_{i's}|i' \in g' \sim G^0_\epsilon(\cdot)$ and $\epsilon^0_{it}|i \in g \sim G^0_\epsilon(\cdot)$ for all $t \geq e_i$.*

There are two things in Assumption 4.4: First, we assume that the productivity shocks are $i.i.d$ across both firms and time. This assumption allows us to impute the unobserved productivity shocks for group-g firms using the distribution G^0_ϵ; Second, we can find a controlled-matching group g' and time s such that the marginal distribution of ϵ^0_{is} is identified.

Proposition 4.3. *Under Assumption 4.2, 4.3, 4.4, G^0_ϵ is identified, and the ℓ-period-ahead CATT for group g is identified as*

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

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

Proof. With the identified $\bar{h}_0$ from Proposition 3.1, for any group-g' firm i at time s, we can recover its $\epsilon^0_{is} \equiv \omega_{is} - \bar{h}_0(\omega_{is-1})$, so the distribution G^0_ϵ is identified. By condition (iii) in Assumption 4.4, the joint distribution of $(\epsilon_{ie_i-1}, ..., \epsilon_{ie_i+l})$ is identified as the product distribution $(G^0_\epsilon)^\ell$. The identification result follows by the evolution process (10). □

Proposition 4.3 implies a simulation-based method to estimate ATT. We illustrate the method using Example 4: Suppose all firms are not treated before $t = 3$, and firms can choose to select whether to take the treatment from $t = 3$. In this case, using the period $t = 2$ data, we can recover the pre-treatment shocks ϵ^0_{i2} for all firms and identify the distribution G^0_ϵ. If the conditions in Proposition 4.3 is satisfied, for each treated firm, we can find its lagged productivity ω_{ie_i-1}, draw ϵ^0_{it} from G^0_ϵ, and simulate its counterfactual untreated future productivity by $\bar{h}^{(\ell)}_0$. The ATT is simply the average of difference between the realized productivity and the simulated untreated productivity for treated firms.

4.2 ATT: Non-absorbing Treatment

In some scenarios, the treatment is non-absorbing by nature. In reality, firms participate in import, export, or R&D activities occasionally.[9] 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[10].

We specify a general Markov process for the potential productivity process that considers the switching back behavior:

$$\begin{aligned}\omega_{it}^1 &= \mathbb{1}(G_{it}=0)\bar{h}_1(\omega_{it-1}^1) + \mathbb{1}(G_{it}=1)h_1^+(\omega_{it-1}^0, \omega_{it-1}^1) + \mathbb{1}(G_{it}=-1)h_1^-(\omega_{it-1}^0, \omega_{it-1}^1) + \epsilon_{it}^1,\\ \omega_{it}^0 &= \mathbb{1}(G_{it}=0)\bar{h}_0(\omega_{it-1}^0) + \mathbb{1}(G_{it}=1)h_0^+(\omega_{it-1}^0, \omega_{it-1}^1) + \mathbb{1}(G_{it}=-1)h_0^-(\omega_{it-1}^0, \omega_{it-1}^1) + \epsilon_{it}^0.\end{aligned} \tag{20}$$

Compared to (3), (20) allows the firms to turn on and off the treatment across time and allows much more flexible transition dynamics when firms change their treatment status. Since the identification of production function and realized productivity does not rely on the $G_{it} \neq 0$ period, Theorem 3.1 still holds. The key difference is the definition of treatment effect and its identification.

Many dynamic treatment effects are not identified under the volatile treatment context. Instead, we focus on some treatment effect of firms that switch its treatment status at time g and maintain the status for ℓ-period. Here we abuse the notation to use g to both denote the treatment cohort group and the group's initial treatment time. Formally, we consider two types of ATTs for the ℓ-period persistent treatment for a time g-positive/negative treatment switcher:

$$\begin{aligned}ATT_{g,\ell}^+ &= \mathbb{E}[\omega_{ig+\ell}^1 - \omega_{ig+\ell}^0 | D_{ig-1}=0, D_{ig}=...=D_{ig+\ell}=1]\\ ATT_{g,\ell}^- &= \mathbb{E}[\omega_{ig+\ell}^1 - \omega_{ig+\ell}^0 | D_{ig-1}=1, D_{ig}=...=D_{ig+\ell}=0]\end{aligned} \tag{21}$$

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

Proposition 4.4. *Under Assumption 4.2, the 0-period-ahead positive/negative switching ATT*

[9] 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.

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

effects at time g are identified as $ATT^+_{g,0} = \mathbb{E}[\omega_{ig} - \bar{h}_0(\omega_{ig-1})|D_{ig-1} = 0, D_{ig} = 1]$, and $ATT^-_{g,0} = \mathbb{E}[\omega_{ig} - \bar{h}_1(\omega_{ig-1})|D_{ig-1} = 1, D_{ig} = 0]$.

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

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

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

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

Assumption 4.5. *There is a cohort group g' such that all firms i' such that $i' \in g'$ are untreated by l-periods since time g', i.e. $D_{ig'-1} = D_{ig'} = ... = D_{ig'+l} = 0$. Moreover, the conditional distribution of $(\epsilon^0_{ig}, ..., \epsilon^0_{ig+l})|(i \in g, \omega^0_{ig-1})$ is the same as the conditional distribution of $(\epsilon^0_{i'g'}, ..., \epsilon^0_{i'g'+l})|(i' \in g', \omega^0_{ig'-1})$.*

Assumption 4.5 generalizes Assumption 4.3 to the non-absorbing treatment case using firms that are not treated between g' and $g'+l$. Since treatment is not absorbing, we further need to condition on the lagged treatment D_{it-1}.

Proposition 4.5. *Suppose Assumption 4.2 and 4.5 hold. The $ATT^+_{g,l}$ is identified by the same expression as the $ATT_{g,l}$ in Proposition 4.2.*

The proof of Proposition 4.5 is the same as the proof of Proposition 4.2 and is hence omitted here. Assumption 4.5 has a similar restriction as Assumption 4.3. However, if firms are allowed to change the treatment status every period, then the g'-matching cohort is very hard to find: The l-period untreated firms are likely to face a very high $\epsilon^0_{ig'}$, and hence these firms are not good match for the g-cohort firms.

However, Assumption 4.5 is likely to hold for treatment that must be maintained for several periods: For example, the treatment decision is whether to use a new technology, and the new technology is not available before time g. At time g, firms can decide whether to take the new technology, and the contract for adopting the new technology must last for at least l period. In this case, we can use all firms at $g' = 0$ as the match group.

5 Discussion of the Potential Productivity Process (3)

Our embedding of the potential outcome into the productivity Markov process is new to the literature, and a structural modeling of the realized productivity ω_{it} cannot achieve this goal. In this section, we use the absorbing case to illustrate the difference between potential and realized productivity process.

If we were to model the evolution process of the realized productivity, and wanted to incorporate the transition period, we could consider:

$$\omega_{it} = h_0(\omega_{it-1})\mathbb{1}(D_{it-1} = D_{it} = 0) + h_1(\omega_{it-1})\mathbb{1}(D_{it-1} = D_{it} = 1) + h^+(\omega_{it-1})\mathbb{1}(D_{it-1} = 0, D_{it} = 1) + \epsilon_{it} \tag{22}$$

This modeling of the realized productivity is appealing since it results in the same moment conditions (8), (9) for estimating the production functions. Indeed, when we consider the potential productivity process, we only use the policy-consistent period to estimate the production, leaving the transition period out. However, the evolution process for the realized productivity (22) differs from the potential productivity process (3) in several ways: The Markovian Assumption and the causal effect interpretation.

Markovian Assumption Difference The realized productivity under (22) is a controlled Markov process. The transition probability only depends on the values of ω_{it-1}, D_{it-1} and D_{it}, but not the time t. However, the realized productivity under the potential productivity process can be non-Markovian. To illustrate this, we consider Example 3. Suppose there are two firms i and i' that take transition at period t and t' respectively, and $\omega_{it} = \omega_{i't'}$. Since treatment are absorbing, we also know $D_{it-1} = D_{i't'-1} = 0$ and $D_{it} = D_{i't'} = 1$. If we model the realized productivity with the Markov process (22), the distribution of ω_{it} and $\omega_{i't'}$ should be the same. In comparison, if we consider the potential productivity process, we have

$$E[\omega_{it}|\omega_{it-1}, D_{it}, D_{it-1}] = E[h^+(\omega_{it-1}, \omega^0_{it-1})|\omega_{it-1}, D_{it}, D_{it-1}].$$
$$E[\omega_{i't'}|\omega_{i't'-1}, D_{i't'}, D_{i't'-1}] = E[h^+(\omega_{i't'-1}, \omega^0_{i't'-1})|\omega_{i't'-1}, D_{i't'}, D_{i't'-1}].$$

The above two quantities are generally different. The non-Markovian property of the realized productivity is substantial in some empirical context. Consider a country that decides between giving a treatment to firms at period t and t' respective. The treatment time can have significant influence on the potential productivity: In general ω^1_{it} and $\omega^1_{it'}$ have different distribution.

Causal Interpretation Difference The control Markov process (22) does not directly give an interpretation of the treatment effect in terms of the potential outcome framework. However, we can put (22) under the divergence of productivity, i.e. Example 2. Note that under the controlled Markov process (22), we can identify h^+ by the following moment condition:

$$E[\omega_{it}(\beta) - h^+(\omega_{it-1}(\beta))|\mathcal{Z}_{it}, D_{it} = 1, D_{it-1} = 0] = 0.$$

Along with the conditional parallel trend Assumption 4.2, we can derive the instantaneous conditional average treatment effect as:

$$E[\omega_{it}^1 - \omega_{it}^0|\omega_{it-1}] = h^+(\omega_{it-1}) - \bar{h}_0(\omega_{it-1}). \tag{23}$$

This is an appealing result since the treatment effect is identified even if the treatment decision D_{it} is dependent on ω_{it-1}. Recall that with the potential productivity process (3), we are only able to identify the average treatment effect on the treated. Such a difference comes from the causal interpretation of the treatment decision: The potential productivity process (3) allows firms' decision to depend on some additional unobserved potential productivity, while (22) essentially assumes that the two potential productivity coincides and there is no unobserved variables in the decision. We discuss the identification of average treatment effect under the specification of Example 2 in Appendix A.2.

6 Empirical Study

6.1 Background

The rise of technologies such as artificial intelligence, robotics, cloud computing, and big data analytics has ushered in a new era of digitalization in firms' production activities. This transformation has sparked significant interests among researchers and policymakers due to its strong implications for productivity growth. In this section, using a firm-level dataset from China, we employ our proposed method to investigate the impact of manufacturing firms' production digitalization on productivity growth.

Production digitalization refers to the integration of utilization of advanced digital technologies and tools throughout the entire production process. For a long time, China's manufacturing is concentrated on making low-ended goods, with an intensive usage of labor. In such background, in addition to industrial policy, Chinese government has been intensively investing in infrastructure in information and communication technologies.

In 2015, China issued the *Made in China 2025* as a national development plan and a comprehensive set of industrial policy to further develop China's manufacturing sector. With all these efforts, Chinese manufacturing firms have been actively investing in digital technologies to upgrade their production processes. Our data covers the period in which many Chinese firms started to adopt digital transformation as an important development strategy. The popularity and policy legitimacy of the digitalization strategy ensures that firms would record their digitalization strategy in their annual reports in explaining their operations to shareholders.

6.2 Data

The empirical study combines two datasets. The first dataset is on publicly traded manufacturing firms in China stock market between 2005 and 2019. This dataset is collected by CSMAR (equivalent to Compustat in the US) and contains rich information on firms' production activities. The second dataset is the annual reports for China's A-shares manufacturing firms downloaded from websites of Shanghai Stock Exchange, Shenzhen Stock Exchange, and CNINF[11] between 2005 and 2019. We use the texts in the annual reports to construct the variable of production digitalization. Our data covers the period in which many Chinese firms started to adopt production digitalization as an important development strategy.

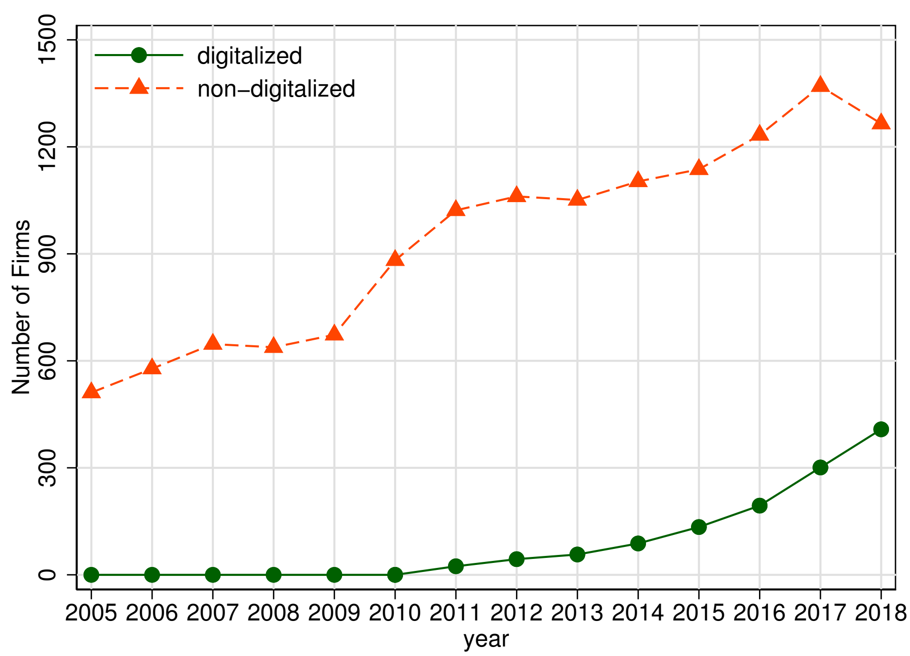
We construct the measure of production digitalization by combining text analysis tools with manual reading of the annual reports of listed manufacturing firms. Based on a set of digitalization related technologies (e.g., big data analytics, artificial intelligence, internet of things (IoT), cloud computing, and robotics), we first extract the digitalization-related keywords and manually read the texts around the keywords in each annual report. Following Zhai et al. (2022), we hired two research assistants independently to manually read the extracted texts to determine whether the firm undertakes production digitalization in each year. In particular, as an improvement of existing method, we have excluded scenarios in which the firm only describes the development of digitalization as a trend in its own industry or as an introduction of the national development strategy. The detailed procedures and several concrete examples of texts on the identified production digitalization are presented in Appendix B. For the empirical purposed, we define a dummy variable $Digit_{it}$ to capture the status of production digitalization of firm i in year t. The variable $Digit_{it}$ takes the value of one for years after year t if firm i's initial year of pro-

[11] CNINF (http://www.cninfo.com.cn/new/index) is a large-scale professional website of securities in China to fully disclose the announcement information and market data of more than 2500 listed companies in Shenzhen and Shanghai.

duction digitalization is identified as year t. Otherwise, $Digit_{it}$ is equal to zero. Note that by the construction of $Digit_{it}$, the treatment is absorbing, fitting the context of our econometric framework.

Figure 2 shows the strong growing trend of production digitalization within the sample period. The number of firms that have adopted production digitalization was zero before 2011, but increased rapidly to be 408 in 2018. This is consistent with the rapid development of digital economy and the building of infrastructure for information technologies in China during this period. Note that there has been an increase in the number of non-digitalized firms. This is because more and more manufacturing firms had become listed firms during the sample period, while the firm exits is relatively rare. The growing trend of production digitalization in our sample provides us a suitable empirical background to employ our proposed method to investigate the productivity effects of production digitalization.

Figure 2: Trend of Production Digitalization



Note: Based on the authors' calculation.

To account for industrial heterogeneity and keep an enough number of observations in each industry, we classify firms according to the first digit of the industry code. The estimation sample has 14,438 observations (13,171 untreated and 1,267 treated), covering seven main manufacturing industries. We provide the summary statistics in Appendix B.

6.3 Estimation Procedure

6.3.1 Production Function Estimation

We employ the Ackerberg et al. (2015) method to estimate a value-added production function, with the extension of potential productivity process (3). We use the translog specification as the benchmark model:

$$y_{it} = \beta_t t + \beta_l l_{it} + \beta_k k_{it} + \beta_{ll} l_{it}^2 + \beta_{kk} k_{it}^2 + \beta_{lk} k_{it} l_{it} + \omega_{it} + \varepsilon_{it} \tag{24}$$

where y_{it}, l_{it}, k_{it} are the logged value added, logged number of employees, and logged capital, respectively. $\beta_t t$ captures the exogenous trend in the production function, and ε_{it} is the exogenous idiosyncratic output shocks. In light of our econometric framework, the realized productivity ω_{it} can be expressed as $\omega_{it} = Digit_{it} \times \omega_{it}^1 + (1 - Digit_{it})\omega_{it}^0$, where $Digit_{it} \in \{0, 1\}$ is the defined indicator for production digitalization. We specify the following dynamic equation for the productivity process for non-switching periods[12]:

$$\omega_{it}^d = \rho_0^d + \rho_1^d \omega_{it-1}^d + \rho_2^d (\omega_{it-1}^d)^2 + \rho_3^d (\omega_{it-1}^d)^3 + \xi_{it}^d, \quad d \in \{0,\, 1\}, \tag{25}$$

where $d = 1$ indicates the treated firms that have adopted production digitalization in the sample period, and $d = 0$ represents the control firms that have never started production digitalization.

Since the productivity process in the switching period is not well characterized by the above specification, we drop the switching period when estimating the production function.[13] As guided by Theorem 3.1, we construct moment conditions by using instruments following Ackerberg et al. (2015). To account for the industrial heterogeneity in production technologies, we estimate the production functions and productivity evolution processes separately for each industry. After the estimation of production function, we compute the productivity and recover the productivity evolution process.

6.3.2 Estimation of the Effects of Production Digitalization on Productivity

Based on the productivity estimates and the recovered productivity process, we use the proposed simulation-based approach to estimate the firm-specific treatment effects by constructing multiple counterfactual productivity paths for each firm. To simulate counterfactual productivity paths for treated units, we draw productivity shocks ξ_{it}^0 from the

[12] The switching period is defined as the initial period in which the firm starts production digitalization

[13] An alternative way is to add a dummy variable indicating the transition period. Our results stay stable if we use this alternative specification.

untreated observations before the year of 2011 when almost no firms was digitalized.

Inspired by the identification of CATT and the simulation based method in Proposition 4.3, we propose to study the following firm-specific ℓ-period treatment effect $TT_{i\ell} \equiv \omega_{ie_i+l} - E_{G_\epsilon^0}[\bar{h}_0^{(l)}(\omega_{ie_i-1}, \epsilon_{e_i}^0, ..., \epsilon_{e_i+l}^0)]$. This firm-specific object allow us to study the heterogeneity of treatment effect across firms that is missed in the ATT statistics. Moreover, the $TT_{i\ell}$ is easier to calculate than the CATT, which requires taking averages of $TT_{i\ell}$ across firms that has the same lagged productivity. Specifically, for firm i that started production digitalization in year e_i, we estimate the firm-specific ℓ-period treatment effects of production digitalization as:

$$\widehat{TT}_{i\ell} = \hat{\omega}_{ie_i+\ell} - \frac{1}{M}\sum_{m=1}^{M} \hat{\omega}_{ie_i+\ell}^0(m), \tag{26}$$

where $\hat{\omega}_{ie_i+\ell}^0(m)$ is the unrealized potential productivity obtained through the simulated productivity path m, and M is the total number of counterfactual productivity paths. In our estimation, we set M to be 100. After experimentation, We notice that the TT estimate is sensitive to the outliers in the distribution of potential productivity shocks ξ_{it}^0. To deal with this problem, instead of drawing from the non-parametric distribution of productivity shocks, we exclude the outliers of productivity shocks by discarding values smaller than 1^{st} percentile or greater than 99^{th} percentile and parameterize the distribution of productivity shocks to be a normal distribution $\xi_{it}^0 \sim \mathcal{N}(0, \sigma_\xi^2)$. The stand deviation σ_ξ is estimated as the sample analog.

Based on the estimated firm-specific treatment effects, we then compute group-specific treatment effects. We consider two types of group-specific treatment effects: the first is the dynamic treatment effects, which are obtained by averaging $\widehat{TT}_{i\ell}$ by period ℓ; the second is the industrial treatment effects, which are computed by averaging $\widehat{TT}_{i\ell}$ by industries.

Due to the concern for overly small sample size, we set ℓ to be 0 to 4. We use the block-bootstrap to construct the confidence interval for the group-level ATT estimates. In particular, we resample observations for each industry by firm-level clusters and repeat the two-step estimation procedure. Considering that different industries have distinct level of production digitalization, we bootstrap the sample by industry-level strata of treated firms and untreated firms.

6.4 Empirical Results

6.4.1 Group-specific Average Treatment Effects

Dynamic Average Treatment Effects We first report the estimation results of dynamic treatment effects in Table 1.[14] We find positive effects of production digitalization on productivity in from period 0 to period 2, but slightly negative treatment effects on productivity in periods 3 and 4. In aggregation, the average effects of production digitalization on productivity is around 0.035. Notably, none of the estimates are statistically significant at the 10% significance level, which means that on average production digitalization has not caused significant productivity growth among these Chinese manufacturing firms. The large standard error suggests that there is a substantial variation in the treatment effects of production digitalization on productivity. This motivates to dig into firm-level treatment effects of production digitalization on the productivity.

Table 1: Treatment Effects of on Productivity

Periods After Digitalization	ATT	SE	Treated Obs.
0	0.069	0.490	330
1	0.028	0.653	219
2	0.036	0.723	140
3	-0.040	0.706	94
4	-0.007	0.817	59
Total	0.035	0.628	842

Note: The production function is specified as translogged production functions. For each firm, 100 counterfactual productivity paths are simulated. Standard errors are obtained by bootstrapping 500 times.

Industrial Average Treatment Effects Table 2 reports the industry-level treatment effect and its contribution to the overall treatment effect in the sample. We obtain the overall treatment effect of production digitalization by averaging over all observations. Note that we do not report the dynamic treatment effect for each industry due to the small sample size. The industry of equipment manufacturing ($\widehat{ATT}$=0.062) and electronics manufacturing ($\widehat{ATT}$=0.193) have the highest ATT of productivity, contributing around 36.9% and 106.2% to the sample's overall ATT, respectively. In contrast, the chemical synthesis industry show the lowest ATT of productivity ($\widehat{ATT}$=-0.157), accounting for -43.4% of the sample's overall ATT. The industrial heterogeneity reflects that firms obtain different productivity gains from adopting production digitalization. The finding that production dig-

[14] The estimated parameters for the translog production function are reported in Appendix B.3.

italization tends to have larger positive productivity effects on manufacturing industries like equipment, electronics, and healthcare may be due to their intricate processes and high technological intensity. The integration of digital technologies into these processes can lead to substantial efficiency gains, precision improvements, and customization opportunities. In contrast, industries like print & paper and food & beverage might have comparatively simpler operations that may not benefit as significantly from digitalization.

Table 2: Industry-level Treatment Effects on Productivity

Industries	Mean	SE	Contribution	Treated Obs.
Equipment Manufacturing	0.062	0.677	106.2%	508
Electronics Manufacturing	0.193	0.422	36.9%	57
Healthcare Manufacturing	0.063	0.363	10.2%	48
Print & Paper	0.023	0.429	2.7%	35
Food & Beverage	-0.004	0.531	-0.6%	53
Metal Processing	-0.061	0.520	-12.1%	59
Chemical Synthesis	-0.157	0.710	-43.4%	82
Total	0.035	0.628	100%	842

Note: The contribution of each industry is calculated as the ratio of sample-share-weighted treatment effects to the average treatment effects in the whole sample.

Comparison with Ex-post Regressions To emphasize the difference between our method and the existing method, we also estimate the treatment effects on productivity using the ex-post regression method. We estimate the following two-way fixed effects model:

$$\hat{\omega}_{it} = \delta Digit_{it} + \rho_1 \hat{\omega}_{it-1} + \rho_2 \hat{\omega}_{it-1}^2 + \rho_3 \hat{\omega}_{it-1}^3 + \lambda_i + \lambda_t + u_{it} \tag{27}$$

where $\hat{\omega}_{it}$ is the productivity estimate for firm i in year t, and $Digit_{it}$ is the dummy variable indicating production digitalization.λ_i and λ_t represent the firm and year fixed effects, respectively. The error term is u_{it}. The parameter δ is usually interpreted as the treatment effects of production digitalization on productivity (e.g., Liu and Mao, 2019).

We follow the estimation strategy of ex-post method to estimate the productivity and run the regression as specified in equation (27). The results are presented in Table 3. We experiment with three ways of estimating the production function and productivity. The first productivity process we specify is an exogenous productivity process without including the information on digitalization. In the other two productivity processes, we specify an endogenous productivity process and estimate the production function either using the entire sample or excluding the transition period. The estimated coefficient of $Digit_{it}$

in equation (27) is robustly negative and significant in various specifications (see Table 3). In our empirical context, if the researcher interpret the estimated coefficient to be the productivity impacts of production digitalization, she would conclude that productivity digitalization has led to productivity declines in the sample period. As we have illustrated, it is not a surprise that the logical inconsistency underlying the ex-post method can lead to misleading empirical results.

Table 3: Productivity Effects Estimation Results Using Alternative Methods

	Dependent var.: $\hat{\omega}_{it}^a$			Dependent var.: $\hat{\omega}_{it}^b$			Dependent var.: $\hat{\omega}_{it}^c$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$Digit_{it}$	-0.146***	-0.102***	-0.103***	-0.150***	-0.104***	-0.104***	-0.164***	-0.130***	-0.134***
	(0.034)	(0.031)	(0.031)	(0.034)	(0.031)	(0.031)	(0.037)	(0.041)	(0.041)
$\hat{\omega}_{it-1}$		0.434***	1.286***		0.437***	1.286***		0.437***	1.112***
		(0.009)	(0.130)		(0.009)	(0.128)		(0.010)	(0.114)
$\hat{\omega}_{it-1}^2$			-0.042***			-0.042***			-0.034***
			(0.006)			(0.006)			(0.006)
$\hat{\omega}_{it-1}^3$			0.001***			0.001***			0.001***
			(0.000)			(0.000)			(0.000)
N	11584	11584	11584	11584	11584	11584	11252	10974	10974
R^2	0.996	0.997	0.997	0.996	0.997	0.997	0.996	0.997	0.997

Note: $\hat{\omega}_{it}^a$ is estimated using an exogenous productivity process, $\hat{\omega}_{it}^b$ is estimated using an endogenous productivity process incorporating the digitalization variable for the whole sample, and $\hat{\omega}_{it}^c$ is obtained through estimating the endogenous process but dropping the switching period. All regressions include firm and year fixed effects. Standard errors are in parentheses. *** $p < 0.01$.

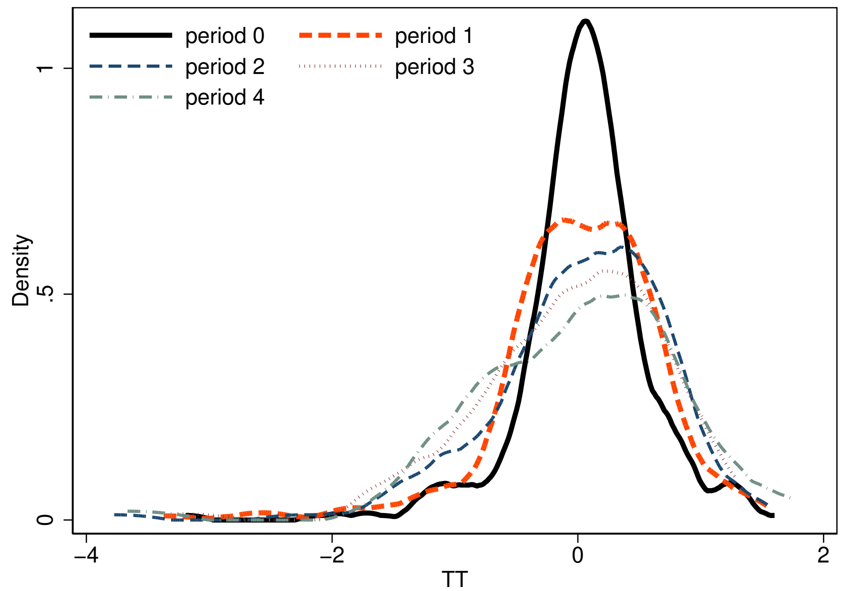
6.4.2 Firm-specific Treatment Effects

In Figure 3, we display the kernel density of the firm-specific treatment effects in different periods after production digitalization. The large variation in productivity gains may reflect the differences in firms' organization efficiency to build the new digital production technology and the learning ability to harness the new digital technology in the production process.

It is clear that the density of firm-specific treatment effects ($\widehat{TT}_{i\ell}$) shifts towards left from period 0 ($\ell = 0$) to period 4 after digitalization. This indicates that production digitalization tends to generate lower productivity gains as time goes by. We also observe a larger dispersion in the firm-specific treatment effects in later periods than earlier periods, indicating that the productivity effects of digitalization become more distinct across firms. The is consistent with the observation that the success rate of digital transformation is low (Bughin et al., 2019), and also largely supports the theory that firms may encounter organizational or technological barriers in the process of upgrading their business practices and the skills of the workforce in order to fully harness the new technology (Taylor and

Helfat, 2009; Feigenbaum and Gross, 2021). However, as production digitalization has large negative productivity effects for a non-negligible portion of firms, the arithmetic mean of digitalization on productivity remains to be negative in later periods.

Figure 3: Firm-specific Treatment Effects of Production Digitalization on Productivity



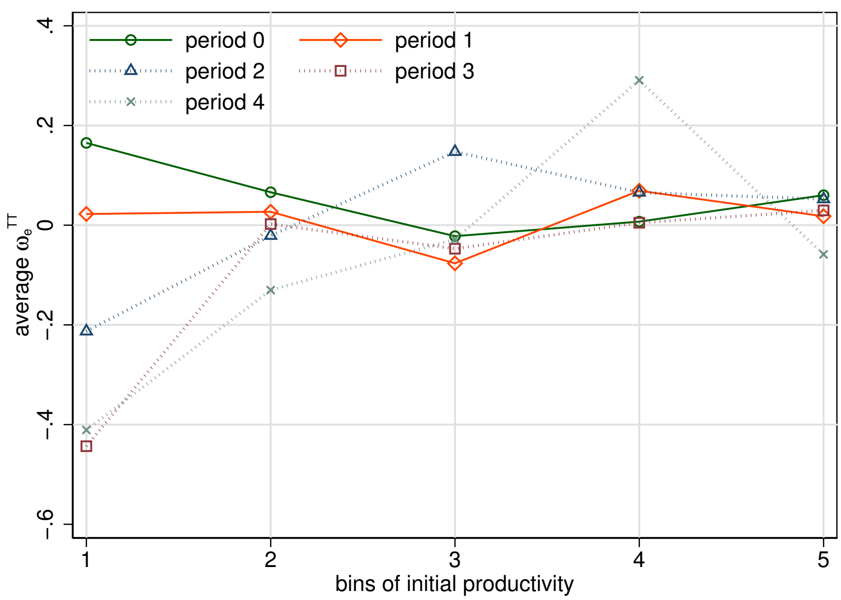
Note: This figure shows the probability density of firm-specific treatment effects of digitalization on productivity. Firm-specific treatment effects on productivity is obtained by simulating 100 counterfactual productivity paths for each treated observation.

Next we examine the average treatment effects for different levels of productivity prior to production digitalization. To this end, we construct 5 productivity bins by splitting the productivity evenly into 5 groups based on the percentiles of the productivity. As time evolves, the productivity growth caused by production digitalization of the low-productivity firms changes from positive to negative. In contrast, in all periods, the productivity gains for high-productivity firms tend to be larger than low-productivity firms.

We further examine the statistical significance for the positive correlation between initial productivity $\hat{\omega}^0_{ie_i-1}$ and firm-specific treatment effects on productivity $\widehat{TT}_{i\ell}$. In the regression of $\widehat{TT}_{i\ell}$ against $\hat{\omega}^0_{ie_i-1}$, we control industry and year fixed effects to account for the industry- and year-specific factors that may affect the impact of production digitalization on productivity (see Table 4). Except for period 0, the regression coefficients of $\hat{\omega}^0_{ie_i-1}$ are positive. In particular, the regression coefficient are statistically significant for periods 2, 3, and 4.

The fact that more productive firms are likely to receive more productivity gains than less productive firms implies that the application of digital technologies in the production process leads to a higher dispersion of the productivity. As productivity is essentially a residual in the production function, it may represent many factors that may affect the output conditional on quantities of capital and labor input. For example, productivity may be positively correlated with a larger stock of intangible assets including human capital (Bowlus and Robinson, 2012) and/or innovation capital (Hall et al., 2010), as well as managerial practices (Bloom et al., 2016). From this perspective, our results echo with a series of findings that the productivity of firms with better management practices grow more rapidly during the episode of information technology (IT) investment in the US (Bloom et al., 2012), so does firms with intangible assets that are complementary to the IT (Bresnahan et al., 2002).

Figure 4: Bins of Initial Productivity and the Impact of Digitalization on Productivity



Note: the horizontal axis indicates the bins of initial productivity, a smaller number of bins indicate lower productivity. The initial productivity is normalized by subtracting the industry-level average productivity to facilitate the cross-industry comparison.

Table 4: Initial Productivity and Firm-specific Treatment Effects on Productivity

	(1) $\widehat{TT}_{i0}$	(2) $\widehat{TT}_{i1}$	(3) $\widehat{TT}_{i2}$	(4) $\widehat{TT}_{i3}$	(5) $\widehat{TT}_{i4}$
$\hat{\omega}^0_{ie_i-1}$	-0.012	0.021	0.191**	0.192***	0.203*
	(0.035)	(0.060)	(0.086)	(0.070)	(0.114)
N	330	219	140	94	59
R^2	0.111	0.114	0.185	0.284	0.283

Note: All regressions include industry- and year- fixed effects. Standard errors are in parentheses. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

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

As an empirical application, we have investigated the impact of production digitalization on productivity by using a sample of Chinese manufacturing firms. We find positive but not significant average treatment effects of production digitalization on productivity. The effects of digitalization on productivity differs substantially across firms, periods, and industries. Specifically, firms with higher ex-ante productivity tend to receive relatively higher productivity gains as time moves on. In general, our findings support the view that new digital technologies have unequal impact on firms' productivity. In sharp contrast, using the traditional two-step method leads to the finding of negative productivity effects of production digitalization.

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.

Bartelsman, E. J. and M. Doms (2000, September). Understanding productivity: Lessons from longitudinal microdata. *Journal of Economic Literature 38*(3), 569–594.

Bloom, N., R. Sadun, and J. Van Reenen (2012, February). Americans do it better: Us multinationals and the productivity miracle. *American Economic Review 102*(1), 167–201.

Bloom, N., R. Sadun, and J. Van Reenen (2016, June). Management as a technology? Working Paper 22327, National Bureau of Economic Research.

Bowlus, A. J. and C. Robinson (2012, December). Human capital prices, productivity, and growth. *American Economic Review 102*(7), 3483–3515.

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.

Bresnahan, T. F., E. Brynjolfsson, and L. M. Hitt (2002). Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. *The quarterly journal of economics 117*(1), 339–376.

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

Brynjolfsson, E., D. Rock, and C. Syverson (2021). The productivity j-curve: How intangibles complement general purpose technologies. *American Economic Journal: Macroeconomics 13*(1), 333–72.

Bughin, J., J. Deakin, and B. OâĂŹBeirne (2019). Digital transformation: Improving the odds of success. *McKinsey Quarterly 22*, 1–5.

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

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

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

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

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

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

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

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

Draca, M., R. Sadun, and J. Van Reenen (2009, 02). 100 Productivity and ICTs: A review of the evidence. In *The Oxford Handbook of Information and Communication Technologies*. Oxford University Press.

Feigenbaum, J. J. and D. P. Gross (2021). Organizational frictions and increasing returns to automation: Lessons from at&t in the twentieth century. Technical report, National Bureau of Economic Research.

Gal, P., G. Nicoletti, T. Renault, S. Sorbe, and C. Timiliotis (2019). Digitalisation and productivity: In search of the holy grail–firm-level empirical evidence from eu countries.

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.

Hall, B. H., J. Mairesse, and P. Mohnen (2010). Measuring the returns to r&d. In *Handbook of the Economics of Innovation*, Volume 2, pp. 1033–1082. Elsevier.

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.

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

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

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

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

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

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

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

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

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

Syverson, C. (2011). What determines productivity? *Journal of Economic literature 49*(2), 326–365.

Taylor, A. and C. E. Helfat (2009). Organizational linkages for surviving technological change: Complementary assets, middle management, and ambidexterity. *Organization Science 20*(4), 718–739.

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

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

Zhai, H., M. Yang, and K. C. Chan (2022). Does digital transformation enhance a firm's performance? evidence from china. *Technology in Society 68*, 101841.

Appendices

A Connection to the Dynamic Treatment Effect

A.1 The 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. Following the notation in Abbring and Heckman (2007), we let $\boldsymbol{D}_i^t = (D_{i1}, ..., D_{it})$, and $\boldsymbol{\omega}_i^{dt} = (\omega_{i1}^d, ..., \omega_{it}^d)$ for $d = 0, 1$. We state the NA condition in our framework.

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

The no-anticipation condition says that if two sequences of treatment coincides up to time t, then the potential productivity up to time t should also coincide. No-anticipation is the crucial assumption for analysis of dynamic treatment effect (Sun and Abraham, 2021).

Given the Markovian evolution process (3), NA Assumption A.1 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 $\boldsymbol{D}_i^t$ coincides with the shock sequence $(\tilde{\epsilon}_{is}^0, \tilde{\epsilon}_{is}^1)_{s \leq t}$ under $\tilde{\boldsymbol{D}}_i^t$. We view Assumption A.1 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 average treatment effects. We state the firm's SR condition in our framework.

Assumption A.2. *(SR-F) $D_{it+1} \perp (\omega_{is}^1, \omega_{is}^0)_{s \geq t} | \mathcal{I}_{it}^F$ holds for all t.*

We call Assumption A.2 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 A.2 holds when D_{it+1} given the firm's information set $\mathcal{I}_{it}^F$. Indeed, from the firm's dynamic optimization behavior, we know D_{it+1} is a function of $\mathcal{I}_{it}^F$, denoted by $D_{it+1} = g(\mathcal{I}_{it}^F)$. Then given the information set $\mathcal{I}_{it}^F$, D_{it+1} is a degenerative variable and thus Assumption A.2 holds. When the treatment variable is externally imposed, and the assigner randomizes the treatment up to the firm's knowledge, i.e. $D_{it+1} = \tilde{g}(\mathcal{I}_{it}^F, \eta_{it})$ for some η_{it} independent of $(\omega_{is}^1, \omega_{is}^0)_{s\geq t}$, then SR-F also holds.

Now, suppose the treatment is absorbing, and the firms can only choose the treatment status D_{ie} at time e. Under the Assumption A.2, define the propensity score as $\kappa(\mathcal{I}_{it-1}^F) \equiv E[D_{it}|\mathcal{I}_{it-1}^F]$. Then we can rewrite the average treatment effect as:

$$E[\omega_{ie}^1 - \omega_{ie}^0] = \mathbb{E}\left[\frac{\omega_{it}D_{ie}}{\kappa(\mathcal{I}_{ie-1}^F)} - \frac{\omega_{ie}(1-D_{ie})}{1-\kappa(\mathcal{I}_{ie-1}^F)}\right]. \tag{A.1}$$

In general, when the sequential randomization fails, the average treatment effect is not identified without further restrictions, see Abbring and Heckman (2007) for discussion.

A.2 Identify the Average Treatment Effect

When we write down the average treatment effect equation (A.1), we use the firms' information set $\mathcal{I}_{ie-1}^F$. Many variables in $\mathcal{I}_{ie-1}^F$, such as $(\omega_{ie-1}^1, \omega_{ie-1}^0)$ and ζ_{ie-1} are not available to the econometrician, and (A.1) does not identify the ATE. Instead, the econometrician has access to the information set $\mathcal{I}_{ie-1}^E$, see Definition 4. To identify the ATE under the absorbing treatment context, we require a sequential randomization condition for the econometrician:

Assumption A.3. *(SR-E) $D_{it} \perp (\omega_{is}^1, \omega_{is}^0)_{s\geq t}|\mathcal{I}_{it}^E$ holds for $t = e$.*

In general, we have $\mathcal{I}_{ie-1}^F \backslash \mathcal{I}_{ie-1}^E = \{(\omega_{is}^1, \zeta_{is})_{s\leq e-1}\}$. If D_{ie} is dependent of ω_{ie-1}^1, then Assumption A.3 fails because ω_{ie-1}^1 is dependent of ω_{ie}^1. However, there are special cases where we can still use the econometrician's information set to identify the ATE.

Proposition A.1. *Suppose the potential productivity process satisfies Example 2, i.e. $\omega_{is}^1 = \omega_{is}^0$ for $s \leq e-1$. Moreover, the cost shock ζ_{ie-1} is independent of the evolution shocks $(\epsilon_{is}^1, \epsilon_{is}^0)_{s\geq e}$ conditional on the econometrician's information set $\mathcal{I}_{ie-1}^E$. Then, we can identify the ℓ-period-ahead average treatment effect as*

$$E[\omega_{ie+l}^1 - \omega_{ie+l}^0] = \mathbb{E}\left[\frac{\omega_{ie+l}D_{ie}}{\kappa(\mathcal{I}_{ie-1}^E)} - \frac{\omega_{ie+l}(1-D_{ie})}{1-\kappa(\mathcal{I}_{ie-1}^E)}\right]. \tag{A.2}$$

Proof. If $\omega_{is}^1 = \omega_{is}^0$ for $s \leq e-1$, then firms' treatment decisions satisfy $D_{ie} = g(\mathcal{I}_{ie-1}^E, \zeta_{ie-1})$. By the potential productivity process (3), $\omega_{ie+l}^1 = \bar{h}_1^{(l)}(h^+(\omega_{ie-1}) + \epsilon_{ie}^0, \epsilon_{ie+1}^1, ..., \epsilon_{ie+l}^1)$, and (3), $\omega_{ie+l}^0 = \bar{h}_0^{(l+1)}(\omega_{ie-1}, \epsilon_{ie}^0, \epsilon_{ie+1}^0, ..., \epsilon_{ie+l}^0)$, where the $\bar{h}_d^{(l)}$ is the l-times composited evolution process of $\bar{h}_d$. Then conditional on the econometrician's information set $\mathcal{I}_{ie-1}^E$, the variation of D_{ie} is caused by ζ_{ie-1}, the variation of $(\omega_{ie+l}^1, \omega_{ie+l}^0)$ is caused by $(\epsilon_{is}^1, \epsilon_{is}^0)_{s\geq e}$. By assumption, they are independent. Therefore, Assumption A.3 is satisfied, and equation (A.2) follows by the propensity score matching method. □

When $\ell = 0$, it can be shown that (A.2) is the same as (23). However, when $\ell > 0$, we cannot directly calculate the ATE from the identified $\bar{h}_1$ and h^+. This interpretation of the treatment effect is also different from the endogenous productivity method.

Even if the productivity process in Example 2 is the same as that in Doraszelski and Jaumandreu (2013), we note that the identified average treatment effect (A.2) is neither $E[h^+(\omega_{ie-1}) - \bar{h}_0(\omega_{ie-1})]$ nor $E[\bar{h}_1(\omega_{ie-1}) - \bar{h}_0(\omega_{ie-1})]$, which are usually interpreted as treatment-related effect in the fully structural models. As we note, the quantity $h^+(\cdot) - h_0(\cdot)$ reflects the trend difference, but it fails to account the selection bias when treatment D_{it} is not exogenous.

B Data Appendix

B.1 Construction of Other Variables

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

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

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

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

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

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

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

Table B.1: Summary Statistics of Main Production Variables

Variable	Mean	SD	P5	P25	P50	P75	P95
$ln(m)$	21.095	1.263	19.209	20.193	21.001	21.875	23.418
$ln(l)$	7.675	1.024	6.073	6.945	7.620	8.359	9.444
$ln(k)$	20.007	1.265	18.034	19.135	19.907	20.800	22.290
$ln(y)$	18.895	1.260	16.893	18.055	18.831	19.708	21.130
$ln(sale)$	21.141	1.222	19.327	20.272	21.041	21.898	23.405

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

Table B.2: Number of Treated and Untreated Observations for Different Industries

Industries	Untreated	Treated	Total
Print & Paper	438	50	488
Food & Beverage	890	74	964
Electronics Manufacturing	1,431	87	1,518
Healthcare Manufacturing	1,627	85	1,712
Metal Processing	1,925	79	2,004
Chemical Synthesis	2,525	130	2,655
Equipment Manufacturing	4,335	762	5,097
Total	13,171	1,267	14,438

Note: The Electronics Manufacturing industry encompasses the production of various electronic equipment, including the manufacturing of other electronic equipment, daily-use electronic appliances, and electronic components. The Equipment Manufacturing industry involves the production of specialized equipment, transportation equipment, instrumentation, cultural and office machinery, general machinery, and electrical machinery and equipment. The Healthcare Manufacturing industry specializes in the production of pharmaceuticals and biotechnology products. The Print & Paper industry covers activities such as printing, manufacturing of cultural, educational, and sports goods, as well as paper and paper product manufacturing. The Food & Beverage industry focuses on food manufacturing, food processing, and beverage manufacturing. The Metal Processing industry encompasses various activities, including nonferrous metal smelting and rolling, metal product manufacturing, nonmetallic mineral product manufacturing, and ferrous metal smelting and rolling. The Chemical Synthesis industry includes the manufacturing of chemical raw materials and chemical products, chemical fiber, plastics, petroleum processing and coking, and rubber products.

B.2 Defining Production Digitalization

The text analysis of annual reports contains two main steps: keyword searching and refining.

Step 1: Keywords Searching To capture state-of-art digital technologies involved in production digitalization, we choose the following keywords (Chinese bopomofo in brackets): *digitalization (Shu Zi Hua), smartness (Zhi Neng), intelligence (Zhi Hui), Internet of Things (Wu Lian Wang or IoT), industrial internet (Gong Ye Wu Lian Wang), big data (Da Shu Ju), cloud computing (Yun Ji Suan), industrial cloud (Gong Ye Yun), platform (Ping Tai), SaaS, C2M and various management information systems (such as PDM, ERP, SRM, CRM, MES, SCADA, PLM and their Corresponding Chinese names)*. Among these words, *"Smart" (Zhi Neng), "Intelligent" (Zhi Hui), and "Platform" (Ping Tai). These keywords appear in annual reports in various forms, such as "Smart Manufacturing" (Zhi Neng Zhi Zao), "Smart Factory" (Zhi Hui Gong Chang), "Smart Production" (Zhi Neng Zhi Zao), "Smart Firms" (Zhi Hui Xing Qi Ye), "Cloud Platform" (Yun Ping Tai), and "Digital Purchasing Platform" (Dian Zi Cai Gou Ping Tai), etc.* To avoid missing useful information on digitalization, we only use the stem words *"Zhi Neng"*, *"Zhi Hui"*, and *"Platform"* to identify digitalization-related texts.

Step 2: Manual Reading and Refining By manual reading of the annual reports, any paragraphs on digitalization that are related to production, manufacturing and equipment or workshop upgrade are classified as production digitalization. However, we notice that in some annual reports, firms may describe the developing of digitalization in its own sector or China's national strategy, which is not related to the firm's own implementation of digitalization. In our construction of the digitalization indicator, we exclude such scenarios by manually identify them and exclude them from the firm's own digitalization strategy. We list three examples below:[15]

- **Example 1** (Stock ID: 000008, Year: 2018) "Driven by the trend of technological progress, *rail transit operation and maintenance equipment has been upgraded from informatization to digitalization characterized by intelligence, data, internet and deep learning*. Traditional equipment is upgrading to intelligent equipment; operation and maintenance system is upgrading from single-product intelligence to unmanned maintenance factory. It is the right time for data-oriented development of equipment in rail transit industry."

[15] The English texts are translated from Chinese texts in firms' annual reports.

- **Example 2** (Stock ID: 300161, Year: 2017): "*Made in China 2025* puts forward the strategic goal of achieving manufacturing power through 'three steps'. Centering on the top-level design of *Made in China 2025*, relevant supporting policies have been issued successively, and *intelligent manufacturing pilot demonstration projects have been accelerated, with obvious demonstration effect*. With the further deepening of transformation, China's manufacturing industry will be further enhanced in digitalization, networking and intelligence."

- **Example 3** (Stock ID: 000020, Year: 2012) "...domestic and international economic environment is complex, with difficult concerns and positive factors co-existing. On the one hand, the ability of technological innovation is insufficient. *In the new wave of industrial revolution which centers on global digital and intelligent manufacturing, the gap between domestic enterprises and developed countries in Europe and the United States in the field of high-end technology is facing the risk of being widened again*, and enterprises will bear the pain of structural adjustment in the process of industrial upgrading ..."

The quoted paragraph in *Example 1* talks about the digitalization development in its own industry, but not the firm's own digitalization. In *Example 2*, the paragraph is a description of China's national strategy on digitalization. *Example 3* mentions the global environment of digitalization, but not the firm's own strategy.

Examples of Identified Production Digitalization To be concrete, we provide some examples of texts that are identified as production digitalization after performing the text analysis:

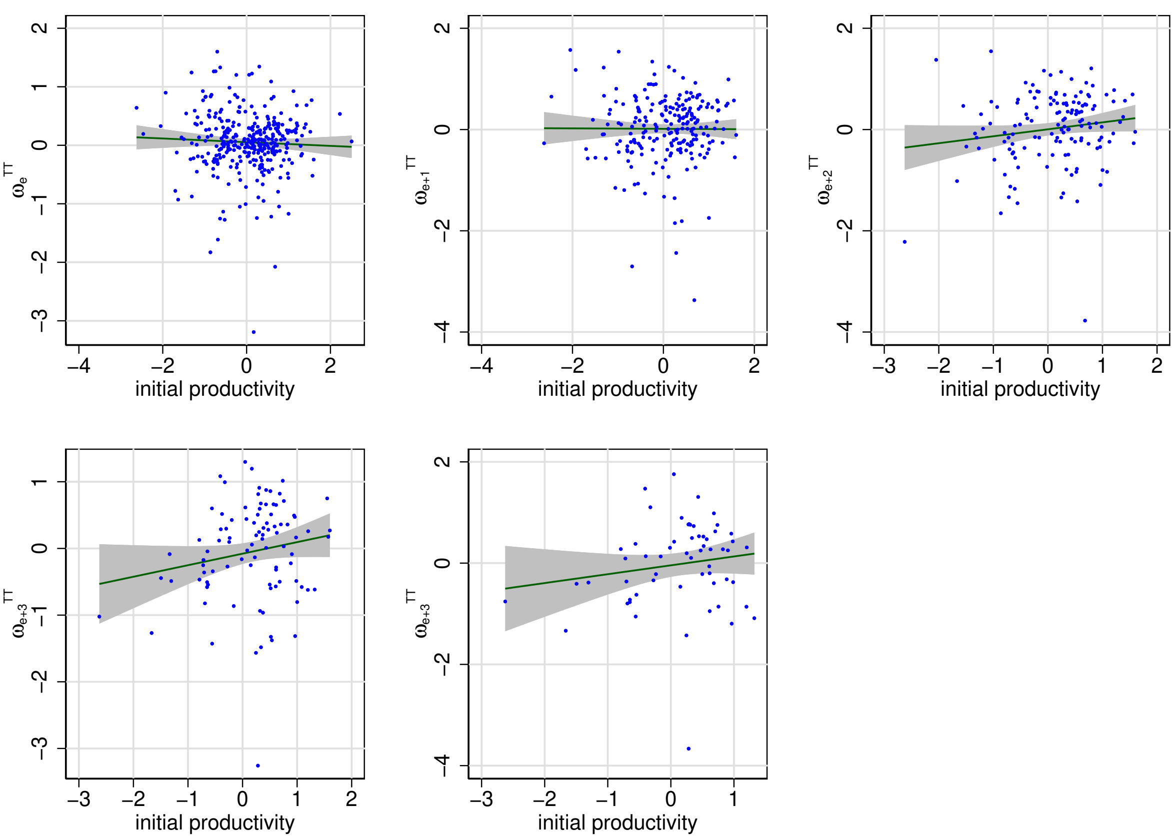
- **Example 1** (Stock ID: 002085, Year: 2018): "Our company has intensified the transformation and upgrading efforts, established intelligent factories with robots as the core, improved the automation level of manufacturing industry, and improved the core competitiveness. By building a digital platform in the whole field of digital research and development, digital technology and digital manufacturing of Wanfeng, our company optimized and standardized the operating system, realized the product life cycle management, and provided data support for company information construction and intelligent manufacturing of intelligent factory."

- **Example 2** (Stock ID: 002920, Year: 2018): "......The company has built digital intelligent factory in an all-round way and established industry-leading highly automated and information-based production lines. Now digital intelligent factory and intelligent storage system have been put into use successively. The construction project of integrated industrialization of automobile electronics and mobile Internet technology has officially laid the foundation and is under construction."

B.3 Supplementary Empirical Results

Scatter Plots of Treatment Effects and the Initial Productivity As a supplement to the analysis in the main text, Figure B.1 presents scatter plots for different periods after production digitalization. From period $\ell = 0$ to period $\ell = 4$, we see that the correlation between $\widehat{TT}_{i\ell}$ and $\hat{\omega}^0_{ie_i-1}$ becomes more and more positive. This means that firms with higher ex-ante productivity obtain higher productivity gains as time evolves.

Figure B.1: Initial Productivity and Productivity Effects of Production Digitalization



Note: All fitted lines are from a linear regression of firm-specific treatment effects $\widehat{TT}_{i\ell}$ on the firm's initial productivity $\hat{\omega}^0_{ie_i-1}$. The initial productivity is normalized using the industry average productivity for each industry for comparison across industries. Shaded areas indicate the 95% confidence intervals for the predicted mean value of firm-specific treatment effects.

Estimates of the Translog Production Functions Table B.3 displays the estimates of the translog production functions for 7 industries in the sample. Note that there are large differences in the production function coefficient estimates, indicating the necessity of estimating the production function separately for each industry. Moreover, almost all the coefficient estimates are statistically significant, confirming the plausibility of using the translog specification to allow the output-input elasticities to depend on the input levels.

Table B.3: Estimates of Translog Production Functions

Industry	β_l	β_k	β_{ll}	β_{lk}	β_{kk}	β_t
Food &Beverage	1.282	-2.916	-0.024	0.010	0.073	0.050
	(0.001)	(0.000)	(0.001)	(0.000)	(0.003)	(0.005)
Print & Paper	-1.333	-1.757	0.111	0.028	0.035	0.080
	(0.083)	(0.109)	(0.007)	(0.002)	(0.004)	(0.007)
Chemical Synthesis	1.769	-2.868	0.118	-0.135	0.100	0.079
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.005)
Electronics Manufacturing	1.282	-2.265	0.140	-0.120	0.079	0.094
	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)	(0.004)
Metal Processing	-1.499	0.848	0.125	0.026	-0.027	0.068
	(0.001)	(0.000)	(0.001)	(0.000)	(0.002)	(0.001)
Equipment Manufacturing	1.133	-2.514	0.077	-0.066	0.076	0.070
	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)	(0.004)
Healthcare Manufacturing	-0.432	-0.047	0.079	0.014	-0.003	0.081
	(0.001)	(0.000)	(0.001)	(0.000)	(0.002)	(0.004)

Note: The standard errors in the parenthesis are obtained by bootstrapping 500 times.

C Additional Results

C.1 Additional Moments for Restricted Productivity Processes

Our moment conditions in Proposition 3.1 impose no additional assumptions on the productivity evolution process (3). While implementing moment conditions in Proposition 3.1 requires minimal structural assumptions, we require a relatively large sample of two-year consecutive observations as in Assumption 3.3. Such data requirements can be satisfied when the panel satisfies a difference-in-difference type design. However, if the treatment variable is volatile over time, we may need to discard a substantial fraction of the firms to implement (8) and (9), which leads to inefficient use of data. We now consider several alternative assumptions on the evolution process that allow us to derive more flexible moment conditions and make use of firms with volatile treatment status.

C.1.1 Independent Evolution Process

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

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

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

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

Assumption C.1 is satisfied for the well-known AR(1) process. The linearity of $r(\cdot)$ ensures that we can generalize moment conditions (8) and (9) to an s-period lagged evolution process. Note that we have to rule out the evolution process h^+ and h^- for the transition periods.

Corollary C.1. *Suppose Assumption C.1 holds and the productivity process satisfies Example 3, then the following two moment conditions hold:*

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

$$\mathbb{E}[\omega_{it}(\boldsymbol{\beta}) - \bar{h}_1^{(s)}(\omega_{it-s}(\boldsymbol{\beta}))|\mathcal{Z}_{it-s+1}, D_{it} = D_{it-s} = 1] = 0. \tag{C.2}$$

Moment conditions (C.1) and (C.2) allow us to use a larger fraction of firms in the dataset. However, we recommend combining moment conditions (C.1) and (C.2) with (8) and (9) to estimate the production functions unless Assumption 3.3 fails. It's unfortunate that we cannot show non-parametric identification of production function with moments (C.1) and (C.2) alone: The error terms ϵ_{it-s} is correlated with k_{it} and l_{it} for all $s \geq 1$, and thus they are not in instrument set $\mathcal{Z}_{it-s+1}$. Therefore, we cannot apply the GNR trick to differentiate both sides of (C.1) to identify the production function.

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

C.1.2 Divergent Productivity Processes

Now we consider the productivity process in Example 2. Since only the observed productivity matters for the evolution process, we can further derive the moment conditions at the transition periods.

Corollary C.2. *Suppose Assumptions 2.1-3.3 hold and the productivity evolution process satisfies Example 2. Then the moment condition (4) (and respectively (6)), (8), (9) and*

$$\mathbb{E}[\omega_{it}(\boldsymbol{\beta}) - h^{+}(\omega_{it-1}(\boldsymbol{\beta}))|\mathcal{Z}_{it}, D_{it} = 1, D_{it-1} = 0] = 0, \tag{C.3}$$

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

The additional moment conditions in Corollary C.2 when the panel is short or when we only observe one period after the treatment status changes. Corollary C.2 requires the transition period to be treated separately from the consistent treatment status period. Moment condition (C.3) is imposed to identify the positive transition process h^+.

C.2 Identifying CATT with an Alternative Assumption

Assumption C.2. *The Markov process ω_{it}^0 satisfies*

$$\omega_{it}^0 = \bar{h}_0^{(s)}(\omega_{it-s}^0) + r(\epsilon_{it}^0, ..., \epsilon_{it-s+1}^0)$$

where $\bar{h}_0^{(s)}$ is an s-period transition function and $r(\cdot)$ is linear in all its arguments. Moreover, the $E[\epsilon^0_{it-s+\ell}|\omega^0_{it-s}] = 0$ for all $\ell \geq 0$.

Proposition C.1. *Under Assumption 3.2, 4.2, and C.2, the ℓ-period-ahead CATT is identified as $CATT_{g,\ell}(\omega) = \mathbb{E}[\omega_{ie_i+\ell} - \bar{h}_0^{(\ell)}(\omega_{ie_i-1})|i \in g, \omega_{ie_i-1} = \omega]$.*

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

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

where (i) by substituting the ω^0_{it} with the evolution process in Assumption C.2. Note that the treatment is absorbing, so

$$E[r(\epsilon^0_{it}, ..., \epsilon^0_{it-s+1})|e_i = t, i \in g, \omega_{ie_i-1} = \omega] = E[r(\epsilon^0_{it}, ..., \epsilon^0_{it-s+1})|D_{it} = 1, i \in g, \omega_{ie_i-1} = \omega].$$

As a result, (ii) follows by Assumptions 3.2 and linearity of $r(\cdot)$. The result in the proposition follows by further taking expectations with respect to e_i. □

Assumption C.2 is satisfied for an AR(1) productivity process, but generally fails when non-linearity appears in the transition function $\bar{h}_0$. Therefore, Assumption C.2 can be restrictive.

C.3 Counterfactual Treatment Effect

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

In this section, we consider a program that rolls out in several phases and the treatment status is absorbing. We start with an initial full set of firms (denoted by $\mathcal{S}$) that are not treated. At time t_0, a subset of firms become treated (denoted by $\mathcal{S}^{tr}$) while the rest of firms remain untreated (denoted by $\mathcal{S}^{ut}$). Untreated firms cannot change their treatment status unless a new phase of the program begins. A policymaker stands at time period $t_0 + s$ and has access to firm-level data up to time $t_0 + s - 1$ and needs to decide whether to start a new phase of the program. There are many empirical examples where the treatment program rolls out in several phases: For example, the State-Owned Enterprise reform in China[16] first took place in the Northeast provinces and rolled out to the

[16] This is known as the privatization process of the state-owned enterprises.

rest of the country in several phases.

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

We, therefore, seek to characterize the counterfactual treatment effect objects that allow the policymaker to evaluate the value of extending the program to the rest of the firms at time $t + s$. In general, without imposing further structural assumptions other than Assumptions 2.2-A.3, it is almost impossible to identify the counterfactual treatment effect objects: The target treatment effect is defined as the difference $\omega^1_{it_0+s} - \omega^0_{it_0+s}$, but for the $\mathcal{S}^{ut}$ firms, we have at best the knowledge of $\omega^0_{it_0+s-1}$ but not $\omega^1_{it_0+s-1}$. We therefore consider several additional structural assumptions that allow us to evaluate the counterfactual treatment effects defined in the following:

$$ATE^{count}_{s,l} \equiv E[\omega^1_{it_0+s+l} - \omega^0_{it_0+s+l}|i \in \mathcal{G}], \tag{C.4}$$

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

C.3.1 Divergent Productivity Process

Recall that the difficulty of characterizing the counterfactual treatment effect comes from the lack of knowledge of $\omega^1_{it_0+s-1}$. However, the divergent productivity process in Example 2 implies the coincidence of two potential outcomes before treatment status changes: $\omega^0_{it_0+s-1} = \omega^1_{it_0+s-1}$ for all untreated firms $i \in \mathcal{S}^{ut}$ and $s \leq 0$. Therefore, we can characterize the counterfactual treatment effect.

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

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

where h^+ is identified from Corollary C.2.

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

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

Assumption C.3. *(i). The shocks satisfy $\epsilon^d_{it} \sim_{i.i.d.} G^d_\epsilon(\cdot)$ for $d \in \{0,1\}$, where the $i.i.d$ is over both firm index i and time index t. (ii). No selection in pre-treatment shocks: $\epsilon^0_{it} \sim_{i.i.d} G^0_\epsilon(\cdot)$ for $t < t_0$. (iii) No selection in already-treated group shocks: $\epsilon^1_{it}|i \in \mathcal{S}^{tr} \sim_{i.i.d} G^1_\epsilon(\cdot)$ for $t_0 \leq t < t_0+s$.*

Assumption C.3 is similar to Assumption 4.4 except that we also require that the distribution $G^1_\epsilon(\cdot)$ is identified from the already treated firms $\mathcal{S}^{tr}$. This is because, for the factual treatment, we can observe the ω^1_{it+s} once the firms are treated. However, for the counterfactual treatment effect, we need to simulate both the treated and untreated future productivity.

Proposition C.3. *Under Assumption 4.2, 4.3, and C.3, $G^0_\epsilon, G^1_\epsilon$ are identified, and the ℓ-period-ahead counterfactual treatment effect at period t_0+s is identified as*

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

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

Remark C.1. *The characterization of the counterfactual treatment effect also highlights another reason in favor of the potential productivity process over the endogenous productivity method (Doraszelski and Jaumandreu, 2013). Recall that Doraszelski and Jaumandreu (2013) do not model the transition period and implicitly assume that $h^+ = \bar{h}_1$ in the identifying moment condition. While imposing $h^+ = \bar{h}_1$ may not lead to a large bias in the production function estimates when the panel is long, it does lead to a bias in the counterfactual treatment effect, especially the instantaneous treatment effect $ATE^{count}_{s,0}$.*

C.3.2 Stationary Conditional Potential Outcome Moments

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

Assumption C.4. *The distribution of* $\omega^1_{it_0-1}|(\omega^0_{it_0-1}=w,\ i\in\mathcal{S}^{tr})$ *is the same as the distribution of* $\omega^1_{it_0+s-1}|(\omega^0_{it_0+s-1}=w,\ i\in\mathcal{G})$*.*

Assumption C.4 is the high-level assumption on potential productivity distribution. There are two constraints embedded in C.4: 1. No selection in the potential outcome. This is reflected in the requirement that we condition on the different firm groups $\mathcal{S}^{tr}$ and $\mathcal{S}^{ut}$; 2. The conditional distribution of ω^1_{it} is stationary at time t_0-1 and t_0+s-1.

Proposition C.4. *Suppose Assumptions C.3 and C.4 hold. The counterfactual treatment effect is identified as*

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

Proof. We first note that

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

where we use Assumptions C.3 and C.4 in the $(*)$ step.

Then the counterfactual treatment effect is identified as

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

The result follows. □

The identified counterfactual treatment effect in Proposition C.4 uses two different approaches to impute the unrealized future potential productivities. For the treated future potential productivity, we use the stationary distribution assumption and use the already treated firms to impute the ω_{it_0+s+l}. In particular, we match each not-yet-treated firm at time $t_0 + s - 1$ with an already-treated firm at time $t_0 - 1$ with the same realized productivity. For the untreated potential future productivity, we simulate the productivity process into the future.