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

Zhiyuan Chen† Moyu Liao‡ Karl Schurter§

July 26, 2024

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 method can be applied under weaker assumptions than those that have been previously employed in the literature and does not require solving 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 experience less productivity growth over time.

Keywords: dynamic treatment effects; productivity; potential outcomes

JEL: C10, C51, L11, L20

*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 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 we 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 usually 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 usually 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 that 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 factor-bias in the intervention.

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," i.e., 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. 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. We show that 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 features of the model. We are only interested in estimating a binary treatment effect, for example, the average treatment on the treated where the treatment would be defined as actively investing in R&D or exporting.[2] 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 effect of adopting a new technology. We only need to identify the mean productivity in the next period conditional on the current productivity in the event that 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, which is similar to the approaches in existing empirical productivity literature (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Aw et al., 2011; Doraszelski and Jaumandreu, 2013). 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

[2]The methodology in this paper does not extend to the case of continuous treatment variables. Accordingly, we would not use this framework to estimate the marginal effect of R&D expenditures or export sales. We leave this extension for future work.

for the current period before the realization of its productivity shock, but we allow the firm to select its treatment status based on all previous treated and untreated potential productivities. Thus, the timing of a firm's choice of treatment status is modeled in the same way that capital is in most of the production function estimation literature.

Instead of precisely modeling how firms select into treatment, we make a high-level assumption about the fraction of observations in the treated and untreated states in order to consistently estimate the production function. Specifically, Assumption 3.3 requires that the data include positive fractions of observations in which firms remain untreated in consecutive periods, are untreated then become treated, and remain treated in consecutive periods. In an ideal setting for policy evaluation, we would observe two periods before and after treatment was randomly assigned among a group of firms. More generally, however, our framework allows treatment status to be endogenously determined outside of the model. For example, firms might select into treatment on the basis of an unmodeled idiosyncratic switching cost. Moreover, firms do not need to choose treatment optimally.

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. Because this procedure generally yields inconsistent estimates of an average treatment effect, even under the strong assumptions that realized productivity follows a Markov process and treatment is exogenously assigned, our more robust methodology provides a useful alternative. We discuss these issues in Section 3.3.

Finally, De Loecker (2007) and De Loecker (2013) represent early examples of the hybrid approach we pursue in this paper, which combines traditional production function estimation with policy evaluation tools. De Loecker (2007) estimates a production function under the assumption that productivity follows a controlled Markov process and subsequently estimates a treatment effect using one of the three methods: propensity score matching, the estimated productivity transition function, or a difference-in-differences analysis. Compared to this work, we relax the assumptions on the productivity process and emphasize the conditions under which the estimand can be interpreted as an average treatment effect or an average treatment effect on the treated.

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 takes 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 first 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 variations in treatment statuses. Here, the key assumption that enables the proposed method is that treatments are 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 some firms that remain untreated for two consecutive periods, and another group of firms that switch from the untreated to the treated state, and remain treated for two consecutive periods. This data feature allows us 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 5 we discuss the ramifications of placing the Markov assumption on the potential productivities as opposed to the realized productivities.

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 AI-related 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 five 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 vary substantially across industries. The greatest productivity gains are realized in industries that use more complex production processes like equipment, electronics and healthcare. The results on the firm-specific productivity effects reveal that the effects of digitalization on productivity become more dispersed as time evolves. Importantly, we also find that firms with lower pre-treatment productivity receive smaller productivity gains over time. In general, our results support the view that new digital technologies have unequal effects on firms' productivities depending on the firm's characteristics (Bloom et al., 2012; Bresnahan et al., 2002; Brynjolfsson et al., 2021).

Using the same dataset, we also show that regression-based methods may lead to qualitatively different conclusions. In fact, when we pattern our analysis off of existing methods, we find a significantly negative effect on production digitalization on productivity. This conflicting result is not surprising in light of the specification issues with regression-based approaches and the substantial heterogeneity revealed by our analysis, but it demonstrates the potential for quantitatively significant discrepancies between existing approaches and ours.

Finally, Section 7 concludes.

All proofs are collected in Appendix B.

2 The Econometric Framework

2.1 A Firm Model with Treatment and Potential Productivity

Firms produce with a Hicks-neutral production technology. Both the firms' production technology and the evolution of their productivity are affected by a binary treatment. Let D_{it} denote the treatment indicator $D_{it} \in \{0, 1\}$, with $D_{it} = 1$ indicating that firm i receives the treatment in period t. The treatment can be imposed externally (e.g., trade liberalization, environmental regulations, etc.) or chosen by the firm (e.g., R&D investment, importing, 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} denotes the output quantity, ω_{it} denotes the realized productivity, η_{it} denotes an ex-post shock to productivity that is not known when a firm choose inputs for time t, K_{it} denotes capital inputs, L_{it} denotes labor inputs, M_{it} denotes material inputs, D_{it} is the treatment status indicator, and $\boldsymbol{\beta}$ is a parameter vector. The dimension of $\boldsymbol{\beta}$ can be infinite when the production function is non-parametric. Moreover, $\boldsymbol{\beta}$ can also include time-related variables to account for secular trends in the production function (e.g., Doraszelski and Jaumandreu (2013)). Note that the production function may depend on the treatment D_{it}, which captures possible impacts on the organization of production or managerial efficiency (Chen et al., 2021).

The two potential outcomes of productivity are 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 may know its potential productivities when making decisions, but the econometrician does not directly observe either potential outcome. To facilitate our exposition, we define a ternary variable to classify possible changes in the treatment status:

$$G_{it} = D_{it} - D_{it-1} \in \{-1, 0, 1\},$$

where $G_{it} = 1$ indicates that firm i became newly treated in period t, $G_{it} = 0$ indicates no change in treatment status, and $G_{it} = -1$ indicates that a firm went from the treated to untreated state. We focus on the absorbing treatment case ($G_{it} \in \{0, 1\}$ for all i and t), and leave the general case as an extension. 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 assumption with the following Markov process for $(\omega_{it}^1, \omega_{it}^0)$:

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

where h^+ is the transition function of ω_{it}^1 when the firm becomes newly treated, and $\bar{h}_1$ and $\bar{h}_0$ are the transition functions when the treatment status is unchanged.[3] We further impose the following assumption to make sure that our potential productivity process (3) is consistent with the specification in the existing literature that assumes no variation in treatment:

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 ω_{it}^1 if the firm's treatment status does not change. This assumption generalizes the productivity processes previously considered in the literature because, when $G_{it} = 0$ for all i and t, each firm's productivity satisfies $\omega_{it}^d = \bar{h}_d(\omega_{it-1}) + \epsilon_{it}^d$, for $d \in \{0, 1\}$. Therefore, we can think of the conventional Markov productivity process as the special case in which there is no treatment, i.e., $D_{it} = 0$ for all i and t.

A wide range of models can be viewed as special cases of the generalized productivity evolution process (3). Below, we list three 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.

Example 1. *(Parallel Shifted Productivity) A policy might simply shift the productivity process by a constant amount. This can be accommodated by assuming potential productivities almost surely satisfy $\omega_{it}^1 - \omega_{it}^0 = C$ for some constant C. More precisely, we would assume that potential productivities (i) have an initial difference of C, i.e., $\omega_{i1}^1 = \omega_{i1}^0 + C$ almost surely for some constant C, (ii) are affected by the same shocks $\epsilon_{it}^1 = \epsilon_{it}^0$ almost surely for all t, and (iii) evolve according to transition functions that satisfy $\bar{h}_1 = h^+$, and $\bar{h}_1(\omega) = \bar{h}_0(\omega - C) + C$.*

[3] We will introduce another transition function h^- to govern the transition out of treatment when we later consider the case of a non-absorbing treatment.

Example 2. *(Diverging Productivity Paths) Consider a case where the binary treatment represents whether a firm invests in R&D. If a firm first begins investing in R&D at* $t+1$*, then we might assume that only* ω_{it}^0 *affects* ω_{it+1}^1*. This assumption can be modeled by* $h^+(\omega_{it}^0, \omega_{it}^1) = h^+(\omega_{it}^0)$*. Because* ω_{is}^1 *for* $s < t$ *is irrelevant to the firm's future outcomes, we are essentially imposing* $\omega_{is}^1 \equiv \omega_{is}^0$ *for all pre-treatment periods* $s \leq t$*.*

Example 3. *(Independent Productivity Evolution) In some cases, a firm chooses between two types of production technologies. We might assume that the firm's total factor productivity using either technology evolves without influence from the other and that the firm simply jumps from one productivity path to the other when it switches technologies, we would impose the additional restriction that* $\bar{h}_1 = h^+$*.*

We note that the second example could be equivalently modeled in terms of realized productivities as in $\omega_{it} = \tilde{h}(\omega_{it-1}, D_{it}, D_{it-1}) + \epsilon_{it}$ without refering to the counterfactual potential productivities. The only difference with earlier work that treats realized productivity as a controlled Markov process would then be that we include a lag of the treatment status indicator in the transition function to allow initial treatment effects and subsequent effects to be different. In Section 5, we more formally compare the Markov process for potential productivity in equation (3) to this alternative. We observe that (3) is more general because this particular extension to the previous literature maintains the assumption that realized productivity is Markovian. Consequently, our more general potential outcome framework imposes fewer restrictions on the causal effect of treatment. [4]

Finally, we observe that alternative generalizations of (3) would be difficult to accommodate without strengthening the assumptions commonly used to estimate production functions. For example, one might wish to relax Assumption 2.1 and allow the potential productivity ω_{it}^1 to depend on both the potential productivities. In general, however, we would not be able to identify the production function from the observed inputs and outputs because the evolution of productivity would be influenced by the unobserved counterfactual productivity. Unlike the exogenous productivity shock, this latent productivity would be correlated with the firm's input choices if the firm anticipates that their treatment status might change in the future, which would invalidate the moment conditions we use to identify the production function. Thus, in addition to highlighting the important distinction between the evolution of realized productivity and potential productivities, the dynamic potential outcome framework in this paper elucidates the assumptions required to identify causal effects on productivity.

[4] We can even allow the evolution at the transition process h^+ to depend on i, but we ignore this extension for ease of exposition. For example, different firms may select into treatment at different times within a year, which may lead to differences in h_i^+ when the model is estimated using annual data.

Firms' Behavior and Timing of Firms' Decisions

Following Ackerberg et al. (2015) and Gandhi et al. (2020), we distinguish the static inputs from 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 during period t. Intermediate input M_{it} is determined no sooner than other inputs after the realization of ω_{it}.*

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

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

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 their 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 the history of state variables $S_{it} \equiv (K_{it}, L_{it}, D_{it}, \omega^1_{it}, \omega^0_{it}, \zeta_{it})$, where ζ_{it} is an idiosyncratic cost shock relevant the dynamic decision[5]. We summarize the firm-decision timeline in the following graph when the labor is predetermined and define the firms' information set correspondingly.

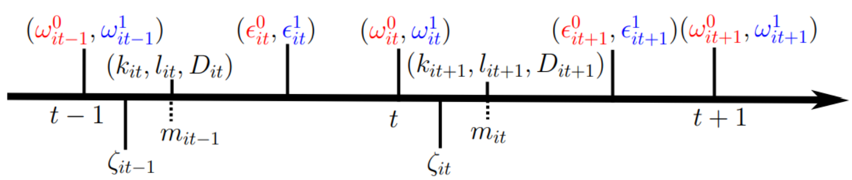


Figure 1: Timeline for a firm's decision.

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

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

When treatment is externally assigned, our model resembles a large class of models considered in the literature on productivity estimation (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg et al., 2015; Gandhi et al., 2020). However, when treatment

[5]For example, ζ_{it} can be idiosyncratic costs of taking treatment.

is endogenous, the firm can choose the treatment status D_{it+1} based on its expected productivity gains and the latent shock ζ_{it} given its information set $\mathcal{I}_{it}^F$. This selection on unobservables complicates the the identification of the treatment effect on productivity, and it prohibits the identification of the full model. Specifically, it prohibits the identification of the transition function h^+ that governs the period in which firms first select into treatment. In contrast, the endogenous productivity approach estimates the treatment effect as a function of all the primitives of the model (Aw et al., 2011; De Loecker, 2013; Doraszelski and Jaumandreu, 2013; Peters et al., 2017).

2.2 Treatment Effects

Before discussing the identification of the treatment effect, we first define the objects of interest that we attempt to recover from the data. In a typical potential outcome framework, the causal effect on an individual would be straightforwardly defined as the difference in the potential outcomes. In our setting, however, the realized potential outcome is not directly observed in the data; rather, a firm's total factor productivity is inferred from observed inputs and outputs. Moreover, the total factor productivity can only be identified up to scale. We must therefore carefully consider what we mean by the causal effect of a treatment and choose a normalizing factor that leads to sensible conclusions. Specifically, we need to normalize the scale of productivity so that we can estimate an average treatment effect of zero if exogenously changing a representative firm's treatment status does not increase physical output using the same vector of inputs.

To clarify the issue, we note that selection into treatment can affect firm output relative to the counterfactual in which the firm remains untreated in three ways. First, when $G_{it} = 1$, firm i's current productivity switches from ω_{it}^0 to ω_{it}^1. We will refer to $\omega_{it}^1 - \omega_{it}^0$ as the contemporaneous treatment effect. Second, the productivity evolves according to $\bar{h}_1$ instead of $\bar{h}_0$, which does not have a contemporaneous effect but changes the trajectory of the firm's future productivity. We will refer to $\bar{h}_1(\omega^1) - \omega^1 - \bar{h}_0(\omega^0) + \omega^0$ as the trend effect. Third, the production function can change in ways that we have not yet specified.

The potential change in production technologies poses a problem when defining the causal effect of an intervention on total factor productivity because, by definition, total factor productivities explain why some firms produce more using the same observable inputs and the same production technology. If the production technology can change arbitrarily when treatment changes, then it is not meaningful to compare total factor productivities between the firm and itself in the counterfactual world in which it did not receive treatment. For example, if $F(K_{it}, L_{it}, M_{it}, D_i = 0; \boldsymbol{\beta}) = A\, F(K_{it}, L_{it}, M_{it}, D_i = 1; \boldsymbol{\beta})$

for all (K_{it}, L_{it}, M_{it}), then we might reasonably wish to conclude that the treatment caused total factor productivity to increase by a factor of A even if the treatment had no effect on the evolution of the idiosyncratic, time-varying component of total factor productivity.

Consequently, we must either assume that the intervention has no effect on the production technology or normalize the relative scale of the production functions so that we can make comparisons across the treated and untreated states. We suggest choosing a benchmark vector of inputs (K_0, L_0, M_0), e.g. the mean input vector in an industry, and relating the scale of the production using $F(K_0, L_0, M_0, D_i = 0; \boldsymbol{\beta}) = F(K_0, L_0, M_0, D_i = 1; \boldsymbol{\beta})$. Although the choice of benchmark affects the relative scale of the production technologies in the treated and untreated states, and therefore influences the estimate of the contemporaneous treatment effect, the econometrician can draw policy-relevant conclusions by choosing (K_0, L_0, M_0) to be representative of the industry or a group of firms within the industry. Under this normalization, we can conclude the intervention has no contemporaneous effect if it has no expected contemporaneous effect on the idiosyncratic component of a representative firm using the benchmark input vector. If a policymaker then asks whether the representative firms would instantly become more productive if the firm were exogenously assigned the treatment, the answer would be no, on average.

On the other hand, this normalization does not affect the identification of the trend effect. Notice that if the log-productivity ω_{it}^1 is identified up to a constant, then we will be able to recover the transition function $\bar{h}_1$ up to a shift by that same constant. Namely, suppose $\tilde{\omega}^1 = \omega^1 + a$ for some scalar a. Then the data would be equally well explained by the transition function $\bar{h}_1(\omega)$ and $\tilde{h}_1(\tilde{\omega}^1) = \bar{h}_1(\tilde{\omega}^1 - a) + a$. Nonetheless, the trend effect would be identified because $\tilde{h}_1(\tilde{\omega}^1) - \tilde{\omega}^1 = \bar{h}_1(\omega^1) - \omega^1$ for all a.

More generally, we want to learn how the contemporaneous and trend effects combine to produce dynamic treatment effects of the form $\omega_{it+\ell}^1 - \omega_{it+\ell}^0$ for $\ell > 0$ for a firm i that is first treated at time t. In the next section, we first discuss the identification of the firm's realized total factor productivity before discussing the identification of average treatment effects in Section 4.

3 Recovering the Unobserved Productivity

To recover a firm's productivity, we assume that the econometrician observes a panel of inputs and outputs:

Assumption 3.1. *The econometrician has access to the instrument set* $\mathcal{Z}_{it} = \mathcal{I}_{it}^F / \{(\omega_{is}^1, \omega_{is}^0, \zeta_{is})_{s \le t}\}$. *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 is justified by the assumption on the timing of the firm's decisions.

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. There are two strands of literature that use different moments to identify the production functions. For the gross output production function, we follow GNR (Gandhi et al., 2020) and use a first-order condition based on the profit-maximizing choice of material inputs. For the value-added production function, we follow ACF (Ackerberg et al., 2015) method and proxy for productivity using materials and other inputs. In both cases, we impose a mean independence assumption on the productivity shocks. We use the lower and upper case letters to represent logs and levels of the corresponding variables, respectively.

GNR Approach to Gross Production Functions When firms are output price takers, the GNR approach uses the following equation for the expenditure share of materials derived from the firm's profit-maximizing problem:

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

where $f_0(k_{it}, l_{it}, m_{it}; \boldsymbol{\beta}) \equiv f(k_{it}, l_{it}, m_{it}, D_{it} = 0; \boldsymbol{\beta})$ is the log production function and s_{it} is the log material-cost-to-revenue-ratio. This moment condition identifies the elasticity of output with respect to materials. To recover the other parameters of the production function and the productivity evolution process, we then use Assumption 3.1 and specialize equation (2) to the case in which $D_{it} = 0$ for all i and t to obtain an additional moment condition:

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

where $\omega_{it}(\boldsymbol{\beta}) = q_{it} - f_0(k_{it}, l_{it}, m_{it}; \boldsymbol{\beta})$ denotes the log-productivity implied by $\boldsymbol{\beta}$.

ACF Approach to Value-added Production Functions 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. Ackerberg et al. (2015) 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

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

When treatment status does not vary, the econometrician can use either set of moment conditions to non-parametrically identify a gross or value-added production function and a sequence of total factor productivities for each firm.

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; and (2) The moment conditions (6) and (7) identify the value-added production function $\boldsymbol{\beta}$ nonparametrically up to a constant difference. Moreover, h is identified nonparametrically using either method.*

As discussed in Section 2.2, 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 $c \in \mathbb{R}$.

3.2 Recovering the Productivity with Variation 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 that treatment is exogenous with respect to the unanticipated productivity shocks $(\epsilon^1_{it}, \epsilon^0_{it})$.

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

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

Notably, Assumption 3.2 allows treatment assignment to depend on past potential outcomes ω^0_{is} and ω^1_{is} for $s < t$.

Because we do not model the treatment decision rule, we have to make the high-level assumption that, as the number of firms in the data grows, we will observe infinitely many firms remain untreated, as well as infinitely many firms remain treated[6]. In addition, to identify treatment effects, we will need to assume that infinitely many firms transition into the treated group.

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

[6] For the absorbing treatment case, it requires a positive fraction of firms receives treatment.

Note that the periods t_0 and t_1 may be equal to each other.

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

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

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

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

To implement the moment conditions in Theorem 3.1, we must discard the transition periods, which might amount to a substantial fraction of the observations in a short panel. We therefore propose some additional moment conditions that use the transition periods under special empirical contexts, see details in Appendix D.1.

Other Structural Objects In some applications, the researcher may be interested in other structural objects of the model, such as the transition function h^+. This object is generally not identified under the assumptions in Theorem 3.1 because we cannot simultaneously observe ω_{it}^1 and ω_{it}^0, but would be identified under stronger assumptions, e.g. the assumption in Example 2.

By recommending our approach, we do not suggest that these objects are always less important than the treatment effects we consider in this paper. Researchers might also be interested in simulating a counterfactual policy intervention, which would require h^+ as well as a model for how firms choose their treatment status. If, however, a treatment effect of an observed intervention is one of the researcher's primary interests, we suggest performing this component of the analysis using our methodology as a more robust alternative to existing methods.

3.3 Revisiting Existing Methods

A simple example illustrates the limitations of two common methods of recovering productivity in the context of a time-varying treatment. Suppose that a policy is implemented between measurement times T_0 and $T_0 + 1$ and exogenously affects a random subset of firms. Using the ex-post regression method, the firms' productivities would be estimated in the first step, and a regression would be used to estimate the average effect of the policy on productivity. Alternatively, using a model of the endogenous productivity method, productivity could be assumed to follow a controlled Markov process in which the transition function depends on the treatment status. Under this assumption, the firm's productivities could be estimated using a set of moment conditions that account for the varying treatment status. The average treatment effect would then be computed from the estimated transition function.

In general, both methods yield inconsistent estimates of the productivities under the general framework of this paper. The ex-post regression method will be inconsistent whenever the policy affects the evolution of productivity, while the endogenous productivity approach will be consistent only if all the primitives of the model are correctly specified. In particular, the expected contemporaneous effect must be zero, and the firm's productivity at the time of treatment must be a function of its previous productivity plus an exogenous shock.

Some of the misspecification issues can be partially resolved within the respective frameworks. Using ex-post regression, the production function could be estimated using only pre-treatment data, while additional lags of the treatment indicator could be included in the endogenous productivity approach to more flexibly model the evolution of productivity. However, these solutions do not make either approach as robust as our proposed method. These issues and alternative remedies are discussed in detail below.

The Ex-post Regression The ex-post regression method proceeds in two steps. First, it estimates the production function and productivity evolution process ignoring the existence of the policy via (4) and (5). Second, given the estimated parameters $\hat{\boldsymbol{\beta}}$ and $\hat{h}$, it might then regress the estimated firm-level productivity $\hat{\omega}_{it} = q_{it} - f(k_{it}, l_{it}, m_{it}; \hat{\boldsymbol{\beta}})$ on the treatment indicator with time- and group-effects to obtain the standard difference-in-differences estimate.

The problem with this approach is that the realized productivity only satisfies the moment condition (5) when $\omega^1_{it} - \bar{h}_1(\omega^1_{it}) = \omega^0_{it} - \bar{h}_0(\omega^0_{it})$ and $h^+(\omega^0_{it}, \omega^1_{it}) = \bar{h}_0(\omega^0_{it})$. If the treatment has a trend effect or a contemporaneous effect, then the estimated production function will generally be biased.

To demonstrate the source of the specification bias, we first specialize the moment condition to this example in the pre-treatment periods $t \leq T_0$. Equation (5) becomes

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

where $\mathcal{Z}'_{it} = \{k_{it}, l_{it}, k_{it-1}, l_{it-1}, m_{it-1}\}$. By Proposition 3.1, this moment condition identifies $\boldsymbol{\beta}$ and $\bar{h}_0$. Substituting $\boldsymbol{\beta}$ and $\bar{h}_0$ for h into (5) for $t = T_0 + 1$ and $t > T_0 + 1$, we obtain

$$0 = \mathbb{E}[\omega^1_{iT_0+1} - \bar{h}_0(\omega^0_{iT_0})|\mathcal{Z}'_{iT_0+1}]\, Pr(D_{iT_0+1} = 1) + \mathbb{E}[\omega^0_{iT_0+1} - \bar{h}_0(\omega^0_{iT_0})|\mathcal{Z}'_{iT_{0+1}}]\, Pr(D_{iT_0+1} = 0), \tag{10}$$

$$0 = \mathbb{E}[\omega^1_{it} - \bar{h}_0(\omega^1_{it-1})|\mathcal{Z}'_{it}]\, Pr(D_{it} = 1) + \mathbb{E}[\omega^0_{it} - \bar{h}_0(\omega^0_{it-1})|\mathcal{Z}'_{it}]\, Pr(D_{it} = 0), \quad \forall t > T_0 + 1\,. \tag{11}$$

In deriving these equations, we have used the fact that $\omega_{it}(\boldsymbol{\beta}) = \omega^{D_{it}}_{it}$ because $\boldsymbol{\beta}$ is the true parameter vector and the assumption that treatment status is exogenously assigned. The second additive terms in equations (10) and (11) are zero by construction. However, if the policy has any effect on the expected evolution of productivity, the first terms will generally not be zero. In particular, (10) will be violated unless $h^+(\omega^0_{it}, \omega^1_{it}) = \bar{h}_0(\omega^0_{it})$, while (11) fails whenever $\bar{h}_1 \neq \bar{h}_0$.

As a result, the production function simultaneously estimated from all of the moments will be biased, with the effect that $\hat{\omega}_{it}$ is not consistent for ω_{it}. This inconsistency then biases any subsequent policy evaluation.

As a simple solution to the problem in this example, we could estimate the production function using only observations from the periods before the policy took effect. If we assume that the treatment does not affect the production technology, then the post-policy productivities can be computed as out-of-sample residuals from the production equation, e.g. $\hat{\omega}_{it} = q_{it} - f(k_{it}, l_{it}, m_{it}; \hat{\boldsymbol{\beta}}_{T_0})$ for $t > T_0$, where $\hat{\boldsymbol{\beta}}_{T_0}$ denotes the parameter estimates using pre-treatment observations. Effectively, this solution uses only the moment conditions in (4) and (8) to estimate the production function. By comparison, our proposed estimator uses more observations and allows the production technology to change.

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

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

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

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

and the moment condition in the first period of treatment becomes

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

The moment conditions in equation (12) are correctly specified. In particular, by Proposition 3.1, $\boldsymbol{\beta}$ and $\bar{h}_0$ 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 for the initial treatment period (13) is misspecified. Substituting the value of $\bar{h}_0$ identified from (12), Part A in (13) equals zero. However, Part B is generally nonzero under the general evolution process (3). The transition function in the first treatment period should be $h^+(\omega^0_{it-1}, \omega^1_{it-1})$, whereas in Part B of (13), the transition function is $\bar{h}_1(\omega^0_{it-1})$. If either the transition function has a contemporaneous effect so that $\omega^0_{it-1} \neq \omega^1_{it-1}$ or the expected evolution in the initial period of treatment is different so that $h^+(\omega^0_{it-1}, \omega^1_{it-1}) \neq \bar{h}_1(\omega^1_{it-1})$, then the moment condition will not be satisfied.

To provide more specific scenarios in which the added flexibility of our framework would be desirable, we return to the previous examples. In Example 1, potential productivities follow parallel paths. For simplicity, we may further specialize this example to the case in which productivity is perfectly persistent: (1) $\omega^1_{it} = \omega^1_{it-1}$; (2) $\omega^0_{it} = \omega^0_{it-1}$; and (3) $\omega^1_{it} = \omega^0_{it} + C$. At time $T_0 + 1$, the treated firm's observed last period productivity is the untreated potential outcome $\omega^0_{iT_0}$. Part B of (13) then becomes $\mathbb{E}[\omega^1_{iT_0+1}(\boldsymbol{\beta}) - \omega^0_{iT_0}(\boldsymbol{\beta})|\mathcal{Z}_{iT_0+1}, D_{iT_0+1} = D_{iT_0} = 1]$. The value of Part B is C at the true production parameter rather than 0, so the model is misspecified.

In Example 2, the evolution at the transition period only depends on the observed outcome in the previous period. If we impose $h^+ = \bar{h}_1$, then Part B of (13) equals zero and the model is not misspecified. However, this assumption may not be satisfied in some empirical contexts. For instance, suppose that the policy takes effect between measurement times T_0 and $T_0 + 1$ and that productivity evolves over shorter time intervals. In this case, $\omega^0_{iT_0}$ first evolves according to $\bar{h}_0$ and before the treated transition function $\bar{h}_1$ takes effect, with the results that the expected difference between $\omega^0_{iT_0}$ and $\omega^1_{iT_0+1}$ is some combination of $\bar{h}_0$ and $\bar{h}_0$. In particular, unless $\bar{h}_0 = \bar{h}_1$, we should expect that $\bar{h}_1 \neq h^+$.

A straightforward solution to this problem with Example 2 would be to include D_{it-1} in $\tilde{h}$, which would allow the transition function during the first period of treatment to differ from both $\bar{h}_0$ and $\bar{h}_1$. Although this is in fact a complete solution to the problem in this example, it does not achieve the same level of generality as our approach because h^+ is assumed to depend only on the previous realized productivity. This solution would generally not yield correctly specified moment conditions under scenarios such as Example 3 in which realized productivity is not Markov (See Section 5).

4 Evaluating the Treatment Effect on Productivity

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 the average treatment effect (ATE) is generally difficult and requires more structural assumptions. We discuss the identification of the ATE in Appendix A.2.

The first step in our analysis is to recognize that the realized productivities are identified as a corollary to Theorem 3.1:

Corollary 4.1. *Under Assumption 2.1-3.3, we can recover the realized ex-post productivity* $\omega_{it}^d + \eta_{it}$ *for firms such that* $D_{it} = d$.

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 define the econometrician's information set below.

Definition 2. *The econometrician's information set is* $\mathcal{I}_{it}^E = \mathcal{Z}_{it} \cup \{\omega_{is}\}_{s \le t-1} \subseteq \mathcal{I}_{it}^F$.

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

4.1 ATT: Absorbing Treatment

The absorbing treatment is at the core of the literature on estimating dynamic treatment effects (Callaway and Sant'Anna, 2021; 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 $D_{it-1} \le D_{it}$. 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 that is absorbing.[7]

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 $t = e_i$

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

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 are of interest, 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 effects for different cohorts. 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 when the treatment is not randomly assigned, the difference-in-differences (DID) method allows us to identify the ATT if a parallel trend assumption is satisfied. We first discuss the fallacy of the conventional DID assumption, and then provide a remedy.

Assumption 4.1. *(Simple Parallel Trend) The following 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 it can be hard to justify. In particular, note that from the productivity process (3), we can derive

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

under Assumption 3.2. From (16) we see that the parallel trend condition will fail if firms select intro treatment based on $\omega_{ie_i-1}^0$.

The Conditional Parallel Trend Assumption We note that, by further conditioning on the value of ω_{it-1}^0 in equation (16), we have

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

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 condition.

This condition relies on the implicit assumption that the firm's untreated potential productivity continues to evolve as it would have if the firm had not been treated. This assumes, for example, that the firm does not lose any intangible capital specific to the untreated production technology. This assumption does not have any empirical content when treatment is absorbing, because the untreated potential productivity is never again relevant to any of the firm's decisions after it takes treatment. In the more general case, however, we must make an additional assumption.

To be concrete, we introduce a more general evolution process for ω_{it}^0:

$$\omega_{it}^0 = \mathbb{1}(G_{it} = 0)\bar{h}_0(\omega_{it-1}^0) + \mathbb{1}(G_{it} = 1)h_0^+(\omega_{it-1}^0, \omega_{it-1}^1) + \epsilon_{it}^0. \tag{18}$$

This general framework allows the potential untreated productivity to have a different evolution process when firms first take treatment. The conditional parallel trends condition (17) is then a direct consequence of the assumption that $h_0^+ = \bar{h}_0$. We refer to this as the conditional parallel trend assumption.

Assumption 4.2. *(Conditional Parallel Trend)* $h_0^+(\omega_{it}^0, \omega_{it}^1) = \bar{h}_0(\omega_{it}^0)$.

Identifying Treatment Effects Based on the implication of (17), we study ℓ-period-ahead conditional average treatment effect on the treated (CATT)

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

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

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

In general, the ℓ-period-ahead CATT and ATT is not identified for $\ell \geq 1$, because we cannot recover the untreated potential outcome $\omega_{ie_i+\ell-1}^0$. Therefore, we need further restrictions to carry out the same derivation in Proposition 4.1. We now formalize several assumptions that help identify the ℓ-period-ahead ATT.

For notation purpose, let $\bar{h}_0^{(\ell)}$ be the ℓ-period productivity transition process, we can write $\omega_{ie_i+\ell}^0 = \bar{h}_0^{(\ell)}(\omega_{ie_i}^0, (\epsilon_{is}^0)_{s=e_i}^{e_i+\ell})$. 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 (i) for all firms* $i' \in g'$ *that are untreated by* ℓ*-periods since time* s*, i.e.,* $e_{i'} > s + \ell$*; and (ii) the conditional distribution of* $(\epsilon^0_{ie_i}, ..., \epsilon^0_{ie_i+\ell})|(i \in g, \omega^0_{ie_i-1})$ *is the same as the conditional distribution of* $(\epsilon^0_{i's}, ..., \epsilon^0_{i's+\ell})|(i' \in g', \omega^0_{is-1})$.

For treatment group g, the group-time pair $(g',\ s)$ is chosen by the econometrician to serve as a comparison group. If Assumption 4.3 holds, firms in group g' can serve as the control group for the firms in the treatment group g and help identify the interested treatment effect on productivity. To accommodate the possible nonlinearity in $\bar{h}^{(\ell)}_0(\cdot)$, we must strengthen the assumption that future productivity shocks are mean-independent of the time of treatment.

We now present an identification result for the ℓ-period-ahead ATT in the following proposition.

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* $i \in g$*.*

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 $\bar{h}^{(\ell)}_0$. As an example of the type of empirical application for which the matching group-time pair $(g',\ s)$ can be determined, consider the following scenario:

Example 4. *In many empirical settings, we are interested in a cohort of firms that start their treatment in period* g_0*:* $g = \{i : e_i = g_0\}$*. In this case, the researcher may use firms that are not treated until period* $g_0 + \ell + 1$ *as the control:* $g' = \{i' : e_{i'} > g_0 + \ell\}$ *and set the time* $s = g_0$*. This means that we use not-yet-treated observations in the same time window as the control group. This choice reflects the belief that the productivity shocks of the not-yet-treated firms during the same period are similar to the treated firms after the treatment takes place.*

Intuitively, Assumption 4.3 holds if the following conditions are met. Before period g_0*, no firms were treated. At time* g_0*, firms decide whether to take the absorbing treatment. Between* g_0 *and* $g_0+\ell+1$*, firms cannot change their treatment status due to governmental regulations or contracts. In this empirical setting, at the time* g_0*, the firm makes the decision on whether the initial treatment time* e_i *is* $e_i = g_0$ *or* $e_i > g_0 + \ell$*, and firms can only make treatment choices 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 gradually roll out in several phases.*

For example, the privatization of Chinese State-Owned enterprises started with an experiment phase in 18 cities, then a phase of an additional 32 cities, and gradually rolled 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. Firms that choose not to be treated until $g_0 + \ell + 1$ are likely those with higher potential productivity $(\epsilon^0_{ie_i}, ..., \epsilon^0_{ie_i+\ell})$ and, therefore, are reluctant to switch to be treated. In this case, we can choose $g' = \{all\ firms\}$ and $s = 2$ in Assumption 4.3. In other words, we use all firms and periods before the initial treatment period g_0, and match firms on the basis of their initial productivity ω_{i1}.

To identify $ATT_{g,\ell}$ from Proposition 4.2, we need to match a treated firm with an untreated firm with the same ex-ante productivity. This matching procedure can be difficult to implement due to two reasons. First, we may not be able to find a (g', s) pair that satisfies the independence restriction. Second, even if (g', s) is found, we may not have enough observations in group-time pair (g', s). Moreover, if all firms are treated at period $g_0 + \ell$, we cannot identify the $ATT_{g_0+\ell+s}$ for all $s > 0$. In such cases, we can rely on the stronger assumption that productivity shocks are independent and identically distributed:

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

This assumption allows us to impute the unobserved productivity shocks for group-g firms using the distribution $G^0_\epsilon(\cdot)$. In order to implement this assumption, however, we require an untreated group-time pair $(g',\ s)$ such that the marginal distribution of ϵ^0_{is} is identified, e.g. a random group of firms that is ineligible for the treatment prior to time s.

Proposition 4.3. *Under Assumption 4.3 and 4.4, the distribution of ϵ^0_{it}, $G^0_\epsilon(\cdot)$, is identified, and the ℓ-period-ahead CATT for group g can be identified as*

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

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

Proposition 4.3 motivates a simulation-based procedure to estimate ATT. This procedure includes two steps. In the first step, the researcher can identify the distribution of ϵ^0_{it}, i.e., $G^0_\epsilon(\cdot)$, for all firms. Then for each treated firm, starting from the ex-ante productivity ω_{ie_i-1}, we can draw ϵ^0_{it} from G^0_ϵ, and simulate its counterfactual productivity path in ℓ

periods ahead using $\bar{h}_0^{(\ell)}(\omega_{ie_i-1}, \epsilon^0_{ie_i}, \cdots, \epsilon^0_{ie_i+\ell})$. The desired ATT is simply the average of difference between the realized productivity and the simulated counterfactual 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.[8] We now discuss the identification of the effects of non-absorbing treatment. Since treatment can be volatile, the individual treatment effect can be influenced by a sequence of past treatment status[9].

We specify a general Markov process for the potential productivity that admits returning to the untreated state:

$$\begin{aligned}\omega^1_{it} &= \mathbb{1}(G_{it}=0)\bar{h}_1(\omega^1_{it-1}) + \mathbb{1}(G_{it}=1)h_1^+(\omega^0_{it-1},\omega^1_{it-1}) + \mathbb{1}(G_{it}=-1)h_1^-(\omega^0_{it-1},\omega^1_{it-1}) + \epsilon^1_{it},\\ \omega^0_{it} &= \mathbb{1}(G_{it}=0)\bar{h}_0(\omega^0_{it-1}) + \mathbb{1}(G_{it}=1)h_0^+(\omega^0_{it-1},\omega^1_{it-1}) + \mathbb{1}(G_{it}=-1)h_0^-(\omega^0_{it-1},\omega^1_{it-1}) + \epsilon^0_{it}.\end{aligned} \tag{19}$$

Compared to (3), equation (19) 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 $G_{it} \neq 0$ periods, Theorem 3.1 still holds. The key difference is the definition of treatment effect and its identification.

Dynamic treatment effects are usually not identified under the context of volatile treatment. Instead, we focus on a particular category of treatment effects for firms that switch their treatment statuses at time g and maintain their new statuses for ℓ periods. Here we abuse the notation to use g to both denote the treatment cohort group and the group's initial treatment time. Formally, we define $ATT^+_{g,\ell}$ and $ATT^-_{g,\ell}$ as the ℓ-period persistent treatment for firms that, at time g, switch into or out of treatment, respectively:

$$\begin{aligned}ATT^+_{g,\ell} &= \mathbb{E}[\omega^1_{ig+\ell} - \omega^0_{ig+\ell} | D_{ig-1}=0, D_{ig}=...=D_{ig+\ell}=1],\\ ATT^-_{g,\ell} &= \mathbb{E}[\omega^1_{ig+\ell} - \omega^0_{ig+\ell} | D_{ig-1}=1, D_{ig}=...=D_{ig+\ell}=0].\end{aligned} \tag{20}$$

We first show that the 0-period ahead treatment effect is identified under the condi-

[8] 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.

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

tional parallel trend assumption for both negative and positive switchers.

Proposition 4.4. *Under Assumption 4.2, the 0-period-ahead positive/negative switching ATT effects at time g are identified as* $ATT^{+}_{g,0} = \mathbb{E}[\omega_{ig} - \bar{h}_0(\omega_{ig-1})|D_{ig-1} = 0, D_{ig} = 1]$, *and* $ATT^{-}_{g,0} = \mathbb{E}[\omega_{ig} - \bar{h}_1(\omega_{ig-1})|D_{ig-1} = 1, D_{ig} = 0]$.

Similar to the absorbing-treatment case, evaluating the ℓ-period-ahead ATT requires an 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 within ℓ periods after period g', i.e., $D_{ig'-1} = D_{ig'} = ... = D_{ig'+\ell} = 0$. Moreover, the conditional distribution of $(\epsilon^0_{ig}, ..., \epsilon^0_{ig+\ell})|(i \in g, \omega^0_{ig-1})$ is the same as the conditional distribution of $(\epsilon^0_{i'g'}, ..., \epsilon^0_{i'g'+\ell})|(i' \in g', \omega^0_{ig'-1})$.*

Assumption 4.5 generalizes Assumption 4.3 to the non-absorbing treatment case using firms that are not treated between g' and $g'+\ell$. 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 $ATT^{+}_{g,\ell}$ is identified by the same expression as the $ATT_{g,\ell}$ 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 a 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, suppose that the treatment decision is whether to use a new technology and that the new technology is not available before the time g. If contractual obligations or the fixed cost of switching technology effectively guarantee that the firm's treatment status will persist for at least ℓ periods, we can use all firms at $g' = 0$ as the match group.

5 Discussion of the Potential Productivity Process (3)

Our approach can be succinctly described as embedding production function estimation into a dynamic potential outcome framework. This approach is new to the literature. Moreover, we argue that our approach is not merely a relabeling of familiar terms. The key distinction is that the previous literature has progressed by making structural assumptions on the realized productivity. In contrast, we have shown that estimates of the ATTs will be more robust to misspecification when analogous assumptions are instead placed on the potential productivities.

By way of comparison, if we were to restrict ourselves to a model of realized productivity, and wanted to allow for more flexible dynamics in the transition period, we could simply replace D_{it} with (D_{it}, D_{it-1}) in h:

$$\begin{aligned}\omega_{it} &= h(\omega_{it-1}, D_{it}, D_{it-1}) + \epsilon_{it} \\ &= \bar{h}_0(\omega_{it-1})\mathbb{1}(D_{it-1} = D_{it} = 0) + \bar{h}_1(\omega_{it-1})\mathbb{1}(D_{it-1} = D_{it} = 1) \\ &\quad + h^{+}(\omega_{it-1})\mathbb{1}(D_{it-1} = 0, D_{it} = 1) + \epsilon_{it}\,.\end{aligned} \tag{21}$$

This model for realized productivity is appealing since it results in the same moment conditions (8), (9) for estimating the production functions. In fact, it implies an additional moment condition for the initial treatment period because (21) asserts that realized productivity is Markovian, while the more general potential productivity process (3) does not. We first argue that the Markov assumption on realized productivity may not be justified and, second, that it is not necessary to identify causal effects.

Markov Assumption The realized productivity under (21) 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 ω_{is} for $s < t-1$. However, the realized productivity under the potential productivity process can be non-Markovian. To illustrate this, we consider the independent productivity paths from Example 3. Suppose firm i first becomes treated at time t and its productivity in period t is therefore equal to $h^{+}(\omega^0_{it-1}, \omega^1_{it-1}) + \epsilon^1_{it}$. By assumption, ϵ_{it} is independent of past potential productivities, so this term cannot be the source of the non-Markov behavior. In addition, ω^1_{it-1} is assumed independent of ω^0_{it-1} in this example. On the other hand, the treatment indicator may be correlated with additional lags of the realized productivity ω^0_{is} for $s < t-1$ conditional on ω^0_{it-1}. For instance, the treatment indicator would be correlated with additional lags of realized productivity if firms anticipate transitioning to the treated state more than one period into the future due to a series of negative untreated productivity shocks. More generally, the counterfactual productivity can be correlated with ω^0_{is} for $s < t-1$ either because the past productivity shocks $(\epsilon^0_{is}, \epsilon^1_{is})$ are correlated or because initial productivities ω^0_{i0} and ω^1_{i0} are correlated.

Indeed, it is perhaps easier to enumerate the situations in which the realized productivity is Markov rather than to describe the ways in which realized productivity might not be Markov. Notably, if ω_{is}^1 is unknown to the firm prior to treatment or has a degenerate distribution conditional on ω_{it-1}^0, then the realized productivity will exhibit the Markov property.[10] When these conditions are not satisfied, the Markov assumption is likely to be violated. In turn, because inflexible inputs can be correlated with past realizations of productivity conditional on ω_{it-1}, the moment condition based on the Markov assumption will be violated in the transition period. Moreover, without additional assumptions, the moment condition cannot be corrected by including additional observables in the transition function or in the set of instruments.[11]

Causal Effects Using the controlled Markov process (21) an ATT or ATE can be computed after identifying the production function and all of the productivity transition functions. Using our more general framework, additional assumptions would be required to identify an ATE via this approach because h^+ is generally not identified in our more general framework. For instance, we can identify the unknowns in (21) under the assumptions of Example 2. Then h^+ is identified by the following moment condition

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

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

$$E[\omega_{it}^1 - \omega_{it}^0|\omega_{it-1}] = h^+(\omega_{it-1}) - \bar{h}_0(\omega_{it-1}). \tag{22}$$

This is an appealing result since the average treatment effect is identified even if the treatment decision D_{it} is dependent on ω_{it-1}. In contrast, recall that, under our more general potential productivity process in (3), we are only able to identify the average treatment on the treated.

Of course, this additional identifying power comes at the cost of generality. The potential productivity process (3) allows firms' decisions to depend on additional unobserved potential productivity, while (21) essentially assumes that the two potential productivities coincide so that there are no unobservable variables in the decision process. Indeed, if one were to impose any type of assumption that permits identification of h^+, then it would

[10] The transition function may not be stationary if the distribution of the counterfactual productivity ω_{it}^1 depends on t, but the process will be Markov in this case as long as the conditional distribution of ω_{it-1}^1 given that the firm chooses treatment does not depend on ω_{is}^0 for $s < t - 1$. Nonstationarity can be accommodated by more flexibly estimating the transition function h.

[11] The underlying issue is a form of essential heterogeneity (see, for example, Heckman et al., 2006).

be possible to identify the ATE.[12] The remarkable fact, however, is that some treatment effects of interest can be identified without identifying all of the primitives of the model.

6 Empirical Study

6.1 Background

The rise of technologies such as artificial intelligence, robotics, cloud computing and big data analytics have ushered in a new era of digitalization in firms' production activities. This transformation has sparked significant interest among researchers and policymakers due to its potential for productivity growth. In this section, we apply our proposed method using a firm-level dataset from China 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 has concentrated on making low-end, labor-intensive goods. Against this backdrop, the 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 policies to further develop China's manufacturing sector. In response to these efforts, Chinese manufacturing firms have been actively investing in digital technologies to upgrade their production processes.

6.2 Data

The empirical study combines two datasets. The first dataset is publicly traded manufacturing firms in the Chinese stock market between 2005 and 2019. This dataset is collected by CSMAR (similar 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 the websites of the Shanghai Stock Exchange, Shenzhen Stock Exchange, and CNINF[13] 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

[12] We further discuss the identification of average treatment effect in Appendix A.2.

[13] CNINF (`http://www.cninfo.com.cn/new/index`) is a large-scale professional website that discloses the announcement information and market data of more than 2500 listed companies in Shenzhen and Shanghai.

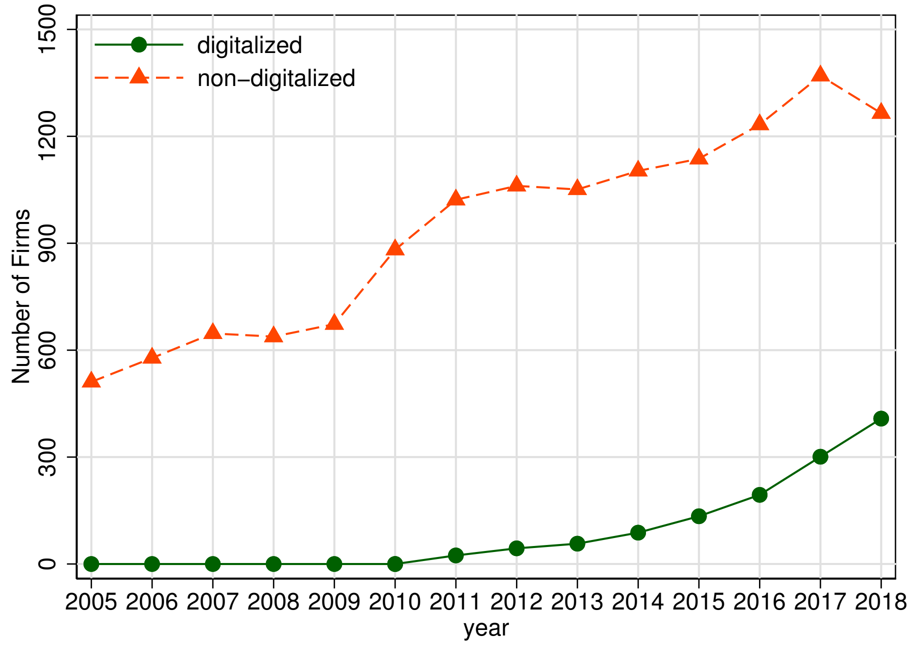
development strategy. Due to the popularity of digitalization among stakeholders and policymakers, we believe that firms have adequate incentive to record any digitalization of their production process in their annual reports to shareholders.

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 the 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 C. For the empirical purpose, 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 production 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.[14]

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 408 in 2018. This is consistent with the rapid development of the 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 manufacturing firms were listed during the sample period, while firm exit is relatively rare. The growing trend of production digitalization in our sample provides a suitable empirical setting to employ our proposed method to investigate the productivity effects of production digitalization.

[14] One may argue that the level of digitalization may change over time. We acknowledge this limitation and leave the investigation of the dynamic treatment effect for continuous treatment variables as an important future research direction.

Figure 2: Trend of Production Digitalization



Note: Based on the authors' calculation.

To account for industrial heterogeneity and keep an sufficient 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 C.

6.3 Identification Assumptions and Estimation Procedures

6.3.1 Identification Assumptions

Our identification strategy relies on the conditional parallel trends (Assumption 4.2), which asserts that, if the firm had not started production digitalization, its counterfactual productivity would have continued to evolve according to $\bar{h}_0$. However, we emphasize that, because we assume the adoption of digital technologies is an absorbing treatment state, the conditional parallel trend assumption is untestable. We cannot observe the evolution of counterfactual productivity once a firm starts production digitalization, and the counterfactual productivity is not relevant to any future actions or outcomes of the firm.

In the present application, we also invoke Assumption 4.3 to identify the treatment effect of production digitalization on productivity. Essentially, this assumption requires that for each treated firm, we can find a comparison group that faced the same distribution of untreated productivity shocks ϵ_{it}^0 conditional on the firm's pre-adoption productivity. Because a negligible fraction of firms reported digitalization prior to 2011, we use all firms in the same industry during 2005-2010 as the comparison group for treated firms.

To investigate this assumption, we compare the (unconditional) distribution of productivity shocks in 2005-2010 with the distribution of productivity shocks for untreated firms in 2011-2018. The suggestive evidence is shown in Appendix C.3. Figure C.1 displays the empirical cumulative distributions for these two groups by industry. Visual inspection strongly suggests that the difference between them is small, but we also conduct Kolmogorov-Smirnov (K-S) tests to test the equality of the two distributions in each industry. The results are reported in Table C.3. The K-S test statistics are small, and the p-values are generally large, indicating support for Assumption 4.3. The exception is the metal processing industry, which appears to have experienced a weaker distribution of untreated productivity shocks after 2010.[15]

6.3.2 Production Function Estimation

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

$$y_{it} = \beta_t t + \beta_l l_{it} + \beta_k k_{it} + \beta_{ll} l_{it}^2 + \beta_{kk} k_{it}^2 + \beta_{lk} k_{it} l_{it} + \omega_{it} + \eta_{it}, \tag{23}$$

where y_{it}, l_{it}, k_{it} are the logged value-added, logged number of employees, and logged capital, respectively. Any exogenous trends in the productivity are captured by $\beta_t t$, and η_{it} is an exogenous idiosyncratic ex-post output shock. As we have mentioned, China has been actively investing in infrastructure in information and communication technologies with the goal of developing the manufacturing sector. The exogenous time trend in the production function may capture the common trend in productivity growth. 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 assume $\bar{h}_0$ and $\bar{h}_1$ are cubic polynomials of the potential productivity with and without digitalization:

$$\omega_{it}^d = \rho_0^d + \rho_1^d \omega_{it-1}^d + \rho_2^d (\omega_{it-1}^d)^2 + \rho_3^d (\omega_{it-1}^d)^3 + \epsilon_{it}^d, \quad d \in \{0, 1\}, \tag{24}$$

[15]Given that the number of treated observations in the metal processing industry is relatively small, our main results stay stable if we remove the metal processing industry.

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 period of adoption is not modeled by (24), we drop this period when estimating the production function.[16] Guided by Theorem 3.1, we construct moment conditions using the instruments of 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. The estimation results for the production functions are presented in Table C.4. Notably, the estimation result shows a significant positive time trend in the production function, indicating rapid technological progress in the Chinese manufacturing sector. This is also consistent with China's intensive investment in infrastructure and information and communication technologies.

In the current specification, we do not allow the production function to vary with treatment status in addition to productivity. As discussed in previous sections, we can in principle allow production digitalization to affect the production technologies, but we do not implement this feature in our empirical application because we have relatively few observations of firms treated in consecutive periods in most industries (see Table C.2).

6.3.3 Estimation of the Effects of Production Digitalization on Productivity

After the estimation of the production function, we compute the productivity and recover the productivity evolution process. 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 ϵ^0_{it} from the untreated observations before the year 2011 when almost no firms were digitalized.

Following the identification argument for the CATT and the simulation-based method in Proposition 4.3, we propose to study the following firm-specific ℓ-period treatment effect $TT_{i\ell} \equiv \omega_{ie_i+\ell} - E_{G^0_\epsilon}[\bar{h}^{(\ell)}_0(\omega_{ie_i-1}, \epsilon^0_{e_i}, ..., \epsilon^0_{e_i+\ell})]$. This firm-specific object allows us to study the heterogeneity of treatment effect across firms that would be missed in the aggregated ATT estimates. Moreover, the $TT_{i\ell}$ is easier to calculate than the CATT, which requires taking averages of $TT_{i\ell}$ across firms that have the same lagged productivity. Specifically, for firm i that started production digitalization in year e_i, we estimate the firm-specific

[16] An alternative way is to add a dummy variable indicating the transition period. Our results are robust to this alternative specification.

ℓ-period treatment effects of production digitalization as

$$\widehat{TT}_{i\ell} = \hat{\omega}_{ie_i+\ell} - \frac{1}{M}\sum_{m=1}^{M} \hat{\omega}^0_{ie_i+\ell}(m), \tag{25}$$

where $\hat{\omega}^0_{ie_i+\ell}(m)$ 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 noticed that the TT estimate is sensitive to the outliers in the distribution of potential productivity shocks ϵ^0_{it}. 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 assume a Gaussian distribution for the productivity shocks: $\epsilon^0_{it} \sim \mathcal{N}(0,\, \sigma^2_\epsilon)$. The standard deviation σ_ϵ 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 an 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 levels 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.[17] 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 aggregate, the average effect 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

[17] The estimated parameters for the translog production function are reported in Appendix C.4.

large standard errors suggest that there is substantial variation in the treatment effects of production digitalization on productivity. This motivates us to further explore firm-level treatment effects of production digitalization on productivity in Section 6.4.2.

Table 1: Treatment Effects 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 shows 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 digitalization 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:

$$\hat{\omega}_{it} = \delta Digit_{it} + \rho_1 \hat{\omega}_{it-1} + \rho_2 \hat{\omega}_{it-1}^2 + \rho_3 \hat{\omega}_{it-1}^3 + \lambda_i + \lambda_t + u_{it}, \tag{26}$$

where $\hat{\omega}_{it}$ is the productivity estimate for firm i in year t, and $Digit_{it}$ is the dummy variable indicating production digitalization. The parameters 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 an ex-post method to estimate the productivity and run the regression as specified in equation (26). 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. Specifically, the productivity process is as follows:

$$\omega_{it} = \varrho_0 + \varrho_1 \omega_{it-1} + \varrho_2 \omega_{it-1}^2 + \varrho_3 \omega_{it-1}^3 + \varepsilon_{it}. \tag{27}$$

In the other two productivity processes, we specify an endogenous productivity process by including the variable of production digitalization in the productivity process:

$$\omega_{it} = \tilde{\varrho}_0 + \tilde{\varrho}_1 \omega_{it-1} + \tilde{\varrho}_2 \omega_{it-1}^2 + \tilde{\varrho}_3 \omega_{it-1}^3 + \tilde{\delta} Digit_{it} + \varepsilon_{it}, \tag{28}$$

and estimate the production function either using the entire sample or excluding the transition period. After obtaining the productivity estimates, we then estimate equation (26), experimenting with controlling for different lagged productivity terms.

The results are reported in Table 3. The estimated coefficient of $Digit_{it}$ is robustly negative and significant in various specifications. In our empirical context, if the researcher interpreted the estimated coefficient to be the productivity impacts of production digitalization, she would conclude that productivity digitalization has led to a significant productivity decline in the sample period. As we have illustrated, it is not a surprise that the logical inconsistency underlying the ex-post regression method can lead to misleading empirical results.

Table 3: Productivity Effects Estimation Results Ex-post Regression Methods

	Dependent var.: $\hat{\omega}_{it}^{a}$			Dependent var.: $\hat{\omega}_{it}^{b}$			Dependent var.: $\hat{\omega}_{it}^{c}$		
Variables	(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. *** $p < 0.01$.

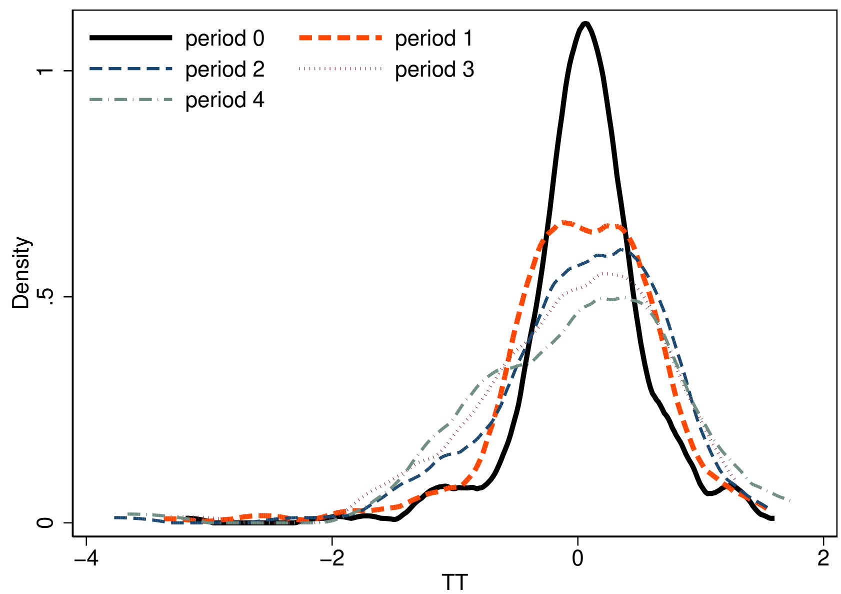
6.4.2 Firm-specific Treatment Effects

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' organizational efficiency in building the new digital production technology and the learning ability to harness the new digital technology in the production process.

The density of firm-specific treatment effects ($\widehat{TT}_{i\ell}$) is more dispersed and appears to shift to the left from period zero ($\ell = 0$) to four periods after digitalization. This indicates that production digitalization tends to have smaller or more negative productivity effects over time for many firms. However, the increased dispersion indicates that firms' experiences with production digitalization also become more strongly differentiated. This 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 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 are obtained by simulating 100 counterfactual productivity paths for each treated observation.

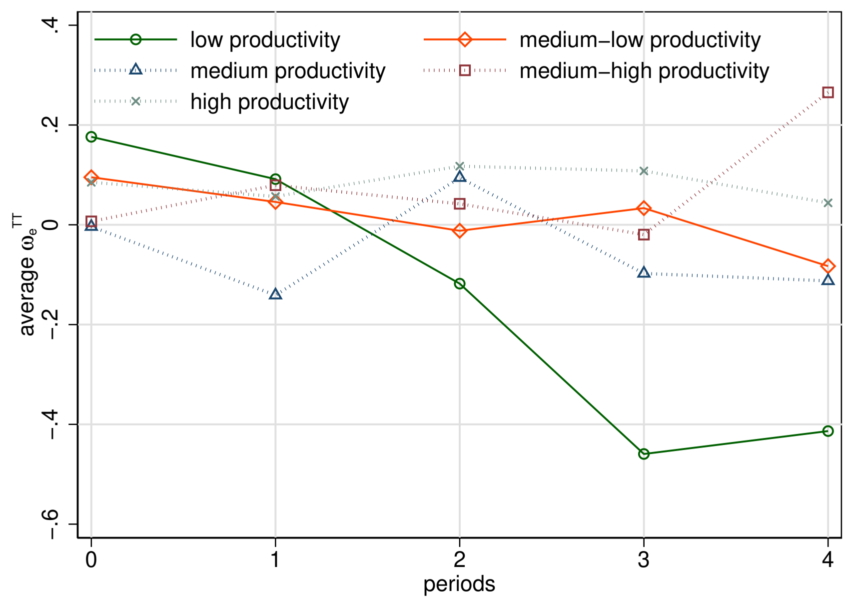
Next, we examine the average treatment effects for different levels of productivity prior to production digitalization. To this end, we construct 5 productivity bins by splitting the productivity evenly into 5 groups based on the percentiles of the productivity. As time evolves, the productivity growth caused by production digitalization of the low-productivity firms changes from positive to negative. In contrast, the productivity gains for high-productivity firms tend to be larger in all periods 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 is 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 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 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 do firms with intangible assets that are complementary to the IT (Bresnahan et al., 2002).

Figure 4: Initial Productivity and the Dynamic Effects of Digitalization on Productivity



Note: the horizontal axis indicates the periods after production digitalization. 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

This paper generalizes standard production function estimation methods by incorporating a binary treatment that affects the evolution of firm-level productivity and production technology. As a theoretical framework, this generalization illuminates potential issues when we identify the parameters of a model with both structural and reduced-form features. As a practical tool, our framework suggests a robust method of estimating the effect of treatment on firm-level productivity that only makes the necessary assumptions to identify the treatment effect and avoids common specification issues with regression-based approaches.

Equivalently, our framework can be viewed as a generalization of the dynamic potential outcome framework to a setting in which the outcome is not directly observed, but must be inferred through the lens of a model. In our case, the outcome of interest is a total factor productivity, which must be inferred under additional assumptions about firm behavior. From either perspective, the goal of the framework is to reconcile the sets of assumptions required to infer the outcome and identify a treatment effect.

The discussion in this paper exclusively pertains to treatment effects on total factor productivity because this setting describes much of the recent literature, but our approach generalizes to other settings. We leave these alternatives and extensions for future work. Some natural alternatives would involve estimating total factor productivity using other means, for instance, using cost shares to estimate input elasticities of output and extracting the residual from the logged output regression as a measure of log-productivity. Compared with our baseline model, this alternative makes additional assumptions about the firm's optimizing behavior and relaxes assumptions about the evolution of productivity. It could be implemented in the same spirit as our approach by separately estimating cost shares in the treated and untreated state, although some assumptions would be required to estimate the missing counterfactual when treatment is not exogenous. In this paper, the Markov assumption on the potential productivity process provides additional moments to estimate the production function as well as permits matching treated and untreated firms on the basis of their lagged productivities. Without assuming a Markov productivity process, some other assumptions would have to fulfill this second role.

Other alternatives might consider factor-augmenting productivity instead of, or in addition to, total factor productivity. Unless the alternative productivity measure is directly measured, such as value-added per hour of labor, these alternatives would most likely entail adaptations of the moment conditions used to recover the outcome. Again, if the new productivity measure is not assumed to be Markov, another assumption would have to be invoked to estimate the missing counterfactual when treatment is endogenously assigned.

Extensions such as allowing for multiple levels of treatment could also be routinely incorporated by applying our binary treatment approach to pairwise differences in the levels of treatment. Other extensions are less obvious but no less desirable in some circumstances, such as accommodating a continuous treatment variable.

As an example of an application of our baseline methodology, we ask whether firm-level data can explain the apparent lack of a productivity revolution in aggregate productivity measures following the introduction of artificial intelligence technologies. Consistent with the macro-level data, we find positive but statistically insignificant effects of production digitalization on productivity. However, our analysis uncovers substantial heterogeneity across firms and industries, as well as over time. Specifically, ex-ante more productive firms tend to enjoy greater productivity gains three to four years after adopting new digital technologies. In sharp contrast, we find significantly negative effects of digitalization on productivity using regression-based methods that mimic existing approaches.

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. O'Beirne (2019). Digital transformation: Improving the odds of success. *McKinsey Quarterly 22*, 1–5.

Callaway, B. and P. H. Sant'Anna (2021). Difference-in-differences with multiple time periods. *Journal of econometrics 225*(2), 200–230.

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

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.

Heckman, J. J., S. Urzua, and E. Vytlacil (2006). Understanding instrumental variables in models with essential heterogeneity. *The Review of Economics and Statistics 88*(3), 389–432.

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 coincide 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 are conditional on the econometrician's information set.

Our structural model implies that Assumption A.2 holds. 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. Instead, the econometrician has access to the information set $\mathcal{I}_{ie-1}^E$, see Definition 2. 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*

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

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), $\omega_{ie+\ell}^1 = \bar{h}_1^{(\ell)}(h^+(\omega_{ie-1}) + \epsilon_{ie}^0, \epsilon_{ie+1}^1, ..., \epsilon_{ie+\ell}^1)$, and

(3), $\omega^0_{ie+\ell} = \bar{h}_0^{(\ell+1)}(\omega_{ie-1}, \epsilon^0_{ie}, \epsilon^0_{ie+1}, ..., \epsilon^0_{ie+\ell})$, where the $\bar{h}_d^{(\ell)}$ is the ℓ-times composited evolution process of $\bar{h}_d$. Then conditional on the econometrician's information set $\mathcal{I}^E_{ie-1}$, the variation of D_{ie} is caused by ζ_{ie-1}, the variation of $(\omega^1_{ie+\ell}, \omega^0_{ie+\ell})$ is caused by $(\epsilon^1_{is}, \epsilon^0_{is})_{s\geq e}$. By assumption, they are independent. Therefore, Assumption A.3 is satisfied, and equation (A.2) follows by the propensity score matching method. □

When $\ell = 0$ and Assumption A.3 holds, it can be shown that (A.2) is the same as (22). 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 effects in fully structural models. As we note, the quantity $h^+(\cdot) - h_0(\cdot)$ reflects the trend difference, but it fails to account for the selection bias when treatment D_{it} is not exogenous.

B Proofs

B.1 Proofs in Section 3

B.1.1 Proof of Lemma 3.1

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 the solution to partial differential equations. □

B.1.2 Proof of Theorem 3.1

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

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

where $\omega^0_{it}(\boldsymbol{\beta})$ denotes the potential productivity without treatment, recovered under parameter value $\boldsymbol{\beta}$ and $D_{it} = 0$. The first equality of (B.1) 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. □

B.2 Proofs in Section 4

B.2.1 Proof of Corollary 4.1

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 $\omega_{is} + \eta_{is} = (q_{is} - f(k_{is}, l_{is}, m_{is}, D_{is} = d; \boldsymbol{\beta})$, which is $\omega^d_{is} + \eta_{is}$ since $D_{is} = d$. □

B.2.2 Proof of Proposition 4.1

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

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

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. □

B.2.3 Proof of Proposition 4.2

Proof. By the definition of CATT:

$$\begin{aligned} CATT_{g,\ell}(\omega) &=_{(a)} \mathbb{E}[\omega_{ie_i+\ell}|i \in g, \omega_{ie_i-1} = \omega] - \mathbb{E}[\bar{h}_0^{(\ell)}(\omega_{ie_i-1}, \epsilon_{ie_i}^0, ..., \epsilon_{ie_i+\ell}^0)|i \in g, \omega_{ie_i-1} = \omega] \\ &=_{(b)} \mathbb{E}[\omega_{ie_i+\ell}|i \in g, \omega_{ie_i-1} = \omega] - \mathbb{E}[\bar{h}_0^{(\ell)}(\omega_{is-1}, \epsilon_{is}^0, ..., \epsilon_{is+\ell}^0)|i \in g', \omega_{is-1} = \omega] \\ &=_{(c)} \mathbb{E}[\omega_{ie_i+\ell}|i \in g, \omega_{ie_i-1} = \omega] - \mathbb{E}[\omega_{is+\ell}|i \in g', \omega_{is-1} = \omega], \end{aligned} \tag{B.2}$$

where (a) follows by the potential productivity evolution process and the potential outcome equation, (b) follows by Assumptions 4.3, and (c) follows by the potential productivity evolution procedure for untreated firms, and $\omega_{is+\ell}^0 = \omega_{is+\ell}$ for untreated firms. □

B.2.4 Proof of Proposition 4.3

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

B.2.5 Proof of Proposition 4.4

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 $G_{ig} = 1$,

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

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

C Data Appendix

C.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 C.1: Summary Statistics of Main Production Variables

Variable	Mean	SD	P5	P25	P50	P75	P95
m	21.095	1.263	19.209	20.193	21.001	21.875	23.418
l	7.675	1.024	6.073	6.945	7.620	8.359	9.444
k	20.007	1.265	18.034	19.135	19.907	20.800	22.290
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 C.2.

Table C.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.

C.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 development of digitalization in their 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 identifying them and excluding them from the firm's own digitalization strategy. We list three examples below:[18]

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

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

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 . . . ”

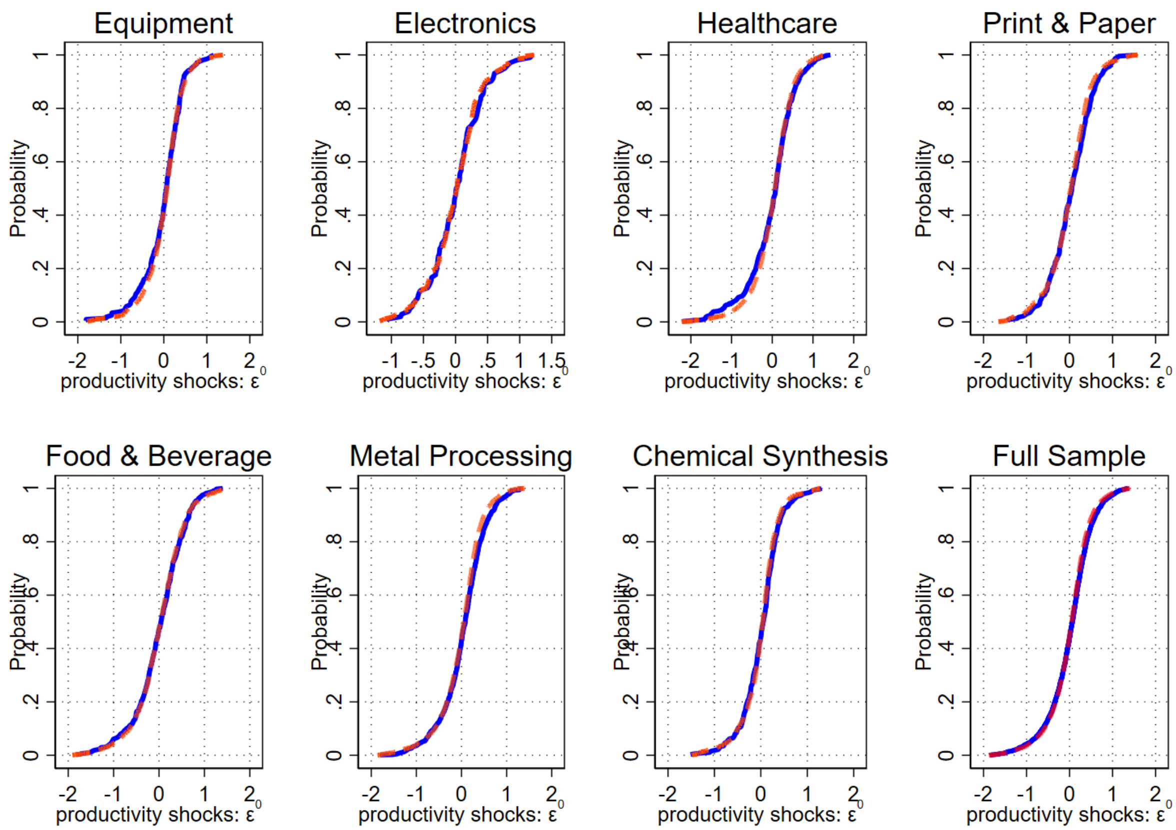
The quoted paragraph in *Example 1* talks about the digitalization development in its own industry, but not the firm’s own digitalization. In *Example 2*, the paragraph is a description of China’s national strategy for digitalization. *Example 3* mentions the global environment of digitalization, but 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 industries, 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 a digital intelligent factory in an all-round way and established industry-leading highly automated and information-based production lines. Now digital intelligent factories and intelligent storage systems 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.”

C.3 Suggestive Evidence of Empirical Assumptions

Figure C.1: Empirical Cumulative Distribution of ϵ_{it}^{0}



Note: The blue solid line represents the CDF of ϵ_{it}^{0} before 2010, and the red dashed line denotes the CDF of ϵ_{it}^{0} after 2010.

Table C.3: Results of Kolmogorov–Smirnov equality-of-distributions test

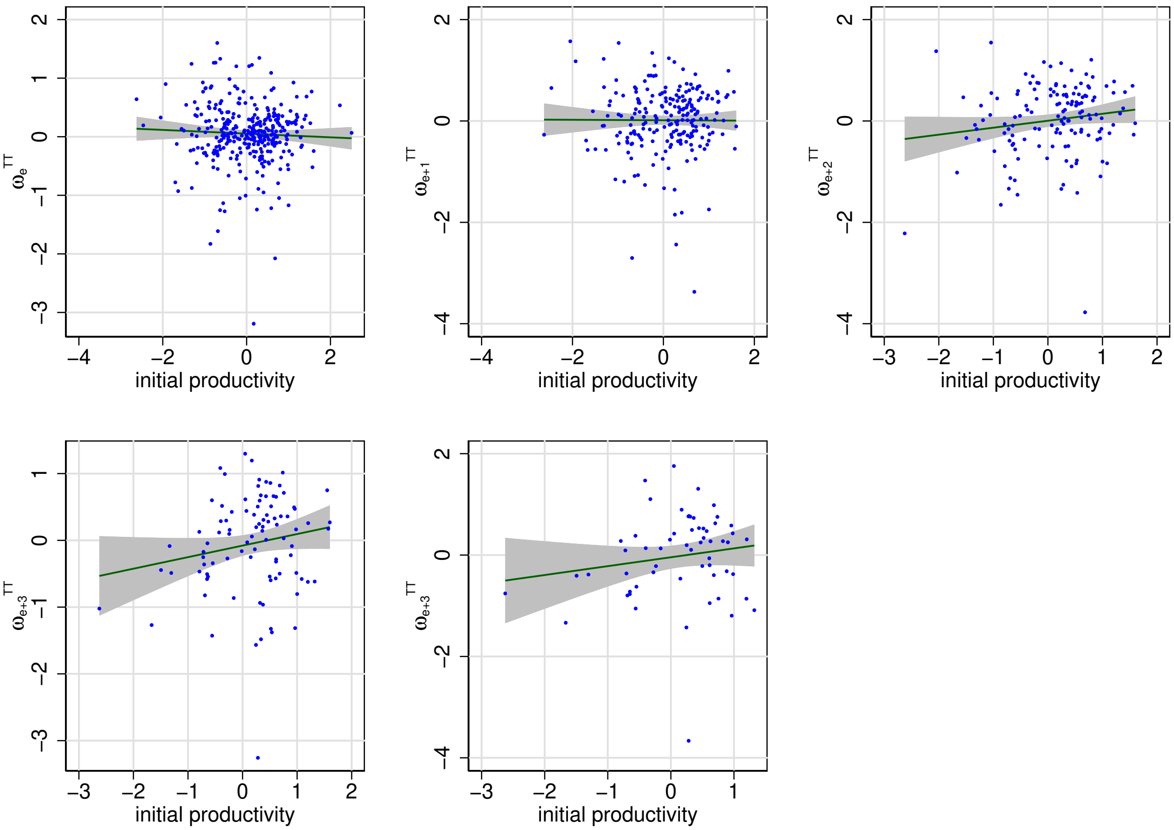
Industry	$Control < Treated$	$Treated < Control$	Combined K-S
Equipment Manufacturing	0.0458	-0.0406	0.0458
	(0.541)	(0.617)	(0.919)
Electronics Manufacturing	0.0373	-0.0759	0.0759
	(0.816)	(0.431)	(0.794)
Healthcare Manufacturing	0.0508	-0.0315	0.0508
	(0.138)	(0.467)	(0.276)
Print & Paper	0.0137	-0.0715	0.0715
	(0.934)	(0.154)	(0.308)
Food & Bevarage	0.0175	-0.0292	0.0292
	(0.825)	(0.584)	(0.951)
Metal Processing	0.0086	-0.0703	0.0703
	(0.912)	(0.002)	(0.004)
Chemical Synthesis	0.0375	-0.0515	0.0515
	(0.468)	(0.238)	(0.47)

Note: The p-values for the test statistics are in parentheses.

C.4 Supplementary Empirical Results

Scatter Plots of Treatment Effects and the Initial Productivity As a supplement to the analysis in the main text, Figure C.2 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 C.2: 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 C.4 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 C.4: 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.

D Additional Results

D.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 consecutively untreated and treated observations as in Assumption 3.3. Such data requirements can be satisfied when the panel satisfies an absorbing treatment 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.

D.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 statuses every two periods must coincide. To form moment conditions for an independently-productivity process, we impose the following assumption:

Assumption D.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 D.1 is satisfied for the 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 D.1. *Suppose Assumption D.1 holds and the productivity process satisfies Example 3, then the following two moment conditions hold:*

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

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

Moment conditions (D.1) and (D.2) allow us to use a larger fraction of firms in the dataset. However, we recommend combining moment conditions (D.1) and (D.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 (D.1) and (D.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 (D.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. However, 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.

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

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

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 D.2 are useful when the panel is short or when we only observe one period after the treatment status changes. Corollary D.2 requires the transition period to be treated separately from the consistent treatment status period. Moment condition (D.3) is imposed to identify the positive transition process h^{+}.

D.2 Identifying CATT with an Alternative Assumption

Assumption D.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 D.1. *Under Assumption 3.2, 4.2, and D.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 D.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 D.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 D.2 can be restrictive.

D.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 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[19] first took place in the 18 experimental cities and rolled out to the 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:

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

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 $t_0 + s$.

D.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 D.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*

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

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 D.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 D.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 D.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} and ϵ^1_{it+s} once the firms are treated. However, for the counterfactual treatment effects, we need to simulate both the treated and untreated future productivity.

Proposition D.3. *Under Assumption 4.2, 4.3, and D.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*

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

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

Remark D.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 $ATE^{count}_{s,0}$.*

D.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 D.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 D.4 is the high-level assumption on potential productivity distribution. There are two constraints embedded in D.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 D.4. *Suppose Assumptions D.3 and D.4 hold. The counterfactual treatment effect is identified as*

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

Proof. We first note that

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

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

Then the counterfactual treatment effect is identified as

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

The result follows. □

The identified counterfactual treatment effect in Proposition D.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 $\omega_{it_0+s+\ell}$. 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.