From Micro to Macro in an Equilibrium Diffusion Model*

Wyatt Brooks
Arizona State University

Kevin Donovan
Yale University

Terence R. Johnson
University of Virginia

October 2020

Abstract

We show how exogenous variation in individual-level interactions identifies critical diffusion parameters in a wide class of general equilibrium firm diffusion models. We implement our procedure in Kenya with a randomized controlled trial in which we shock firm learning opportunities. Embedding these results in a full general equilibrium model generates a quantitatively important diffusion externality. Micro treatment effects are crucial for estimating the diffusion parameters, but have no direct relationship with the quantitative importance of equilibrium diffusion. The results highlight the complex relationship between micro and macro effects when the externality operates primarily through the aggregate state of the economy.

JEL Classification Codes: O1, O4
Keywords: diffusion, equilibrium, productivity

*Thanks to conference and seminar participants at Arizona State, Columbia, Duke, Illinois, Notre Dame, NYU, Penn State, Stanford, University of British Columbia, the Firms in Emerging Economies Workshop at Jinan University, the NBER Growth Meeting, the Temple Macro-Growth Conference, and the Y-RISE Macro and Growth Meeting for helpful comments, especially Daron Acemoglu, Hugo Hopenhayn, Joe Kaboski, Melanie Morten, Jesse Perla, Tommaso Porzio, Duncan Thomas, Kevin Song, Meredith Startz, Mike Waugh, and Daniel Xu. We thank the Ford Family Program and the Helen Kellogg Institute for International Studies for financial support, including Bob Dowd and Dennis Haraszko for their help coordinating the project, and Jackline Aridi, Lawrence Itela, and Maurice Sikenyi for managing the project in the field. Brian Mukhaya provided outstanding research assistance. *Contact Information:* Brooks, wjbrooks@asu.edu; Donovan, kevin.donovan@yale.edu; Johnson, terencejohnson.trj@gmail.com

1 Introduction

A growing literature focuses on the role of productivity transfer among firms in the development process (e.g., Perla and Tonetti, 2014; Perla et al., 2020; Buera and Oberfield, 2020; Hopenhayn and Shi, 2018; Lucas and Moll, 2014). However, quantifying the importance of this channel is difficult. The broad notion of productivity used in these models – designed to capture to myriad of factors that make firms more productive and profitable – implies that the diffusion process cannot be directly observed. The standard approach to deal with this issue is to impose structure. Assumptions on the economic environment outside those made directly on the diffusion process (such as the distributions of shocks, how firms enter and exit, details of occupational choices, etc.) allow a mapping between theoretical parameters and easily – observed empirical moments, such as firm exit rates.

In this paper, we do three things. First, we provide a methodology to estimate key diffusion parameters using well-identified micro moments without the need for additional structural assumptions on the remaining economic environment, which holds for a broad class of diffusion processes. We then apply these results by creating the required "identified moments" (in the sense of Nakamura and Steinsson, 2018) with a randomized controlled trial that shocks the learning opportunities of individual firm owners in Kenya. Finally, we study the impact quantitatively, showing that these parameters matter both for macroeconomic magnitudes and the interpretation of micro-level evidence that is commonly used to motivate policy.

The class of models we study is those in which agents imitate other agents and thereby have the opportunity to improve productivity. We then show that two orthogonality conditions on the data generating process are sufficient to identify key diffusion parameters. Though the results are satisfied for any instrument that generates these required orthogonality conditions, they are easily understood as following from an "ideal" experiment in which we randomly shock the imitation opportunities for a subset of the population, then within this treatment group, randomly provide a one-to-one match. The former is akin to the standard exclusion restriction measuring the causal effect of a policy. The latter allows heterogeneity in the effect to be interpreted causally.

We show that these two conditions are sufficient to identify key diffusion parameters in the types of processes typically considered in the literature, and proceeds in two steps. First, the random variation in one-to-one matches allows us to identify the elasticity by which one's own productivity responds to a higher quality match. The second step focuses on understand the distribution of returns from imitation, conditional on the impact of a realized match. We introduce a parameter that governs the difference between the distributions of firm productivity and imitation draws. Despite the fact that we cannot observe control draws, our insight here is that the average treatment effect provides critical information about the average quality of imitation draws received by this group.

If the average treatment effect is large, for example, then the imitation draws provided by the treatment must be of greater average quality than those drawn by the control. Therefore, conditional on the fact that the first step identifies the parameters governing the outcome of a given match, the average treatment effect is informative about how the control group receives imitation draws without requiring us to directly observe their matches.

A key benefit of our procedure is that these parameters are identified under a wide range of assumptions on the nature of the underlying diffusion process, including random search, bargaining over knowledge transfer, the introduction of noise in the imitation process, and deterministic assignment.[1] Moreover, our procedure for identifying these parameters is independent of much of the remaining model structure. We are not required to make assumptions about whether the economy is in a steady state or transition path, we do not need to make assumptions about the process of entry or exit, and we need no assumptions about the shape of the productivity distribution. This independence result is in large part due to randomization. In applied work, the power of RCTs stems from their ability to eliminate potential confounds with minimal requirements on the data-generating process. Here, something similar occurs, but the confounds are structural. Various aspects of the modeled economic environment interact with the diffusion process, thus making identification difficult without substantial structural assumptions that map parameters to empirical moments. In both cases, using information from a control group helps limit many of the potential confounding factors. Our estimates can therefore be embedded into a variety of models with different assumptions on the remaining structure of the economy.[2] We emphasize this in Section 2 by only laying out the assumptions required for identification, leaving the details of the full model for Section 4, when we require them for the quantitative results. Of course, not every diffusion model will satisfy the assumptions we use. In Section 2 we therefore discuss in detail where our assumptions fail and how one would need to augment our procedure to satisfy more complicated economic environments.

We then provide an application of our results, in which we implement this experiment in Nairobi, Kenya. The randomized controlled trial essentially mimics the underlying microeconomic process of a simple random search model. We set up a program in which we guarantee a randomly selected set of firm owners access to a high-profit entrepreneur. Within this group, we then randomly create one-to-one matches between our set of treated firms and an individual high-profit entrepreneur. We empirically trace out the impulse response of the shock, finding that the average treatment effect on profit is 19 percent.

[1] While the interpretation of various parameters differs depends on how one chooses to model diffusion, the identification procedure does not. In the main body of the paper, we discuss them in the context of a random search model to fix ideas. In the Appendix we include examples of how to implement the procedure with other types of matching frameworks.

[2] It is worth emphasizing that the goal of this paper is not provide a test that differentiates between diffusion models, and indeed our empirical results are not suitable for such a task. The goal here is only to limit the assumptions required for identification within a class of diffusion models commonly used in the literature.

There is also substantial heterogeneity in treatment effect, in that it depends strongly on the profit gap between the two firms.[3] Moreover, we find no evidence that the more productive member of the match sees any change in profit or business skills that one might associate with higher productivity. These empirical results form the basis of our parameter estimation, and the full set of empirical results are in Brooks et al. (2018).

We embed these estimates in a model to study the scale and importance of diffusion. We build a model with random search and occupational choice, in which individuals in the economy learn from firm owners through diffusion. We incorporate the diffusion parameters estimated from our empirical results, while the remaining model parameters are calibrated to match characteristics of the environment in which these firms operate. As is well known in models where diffusion occurs via operating firms, the key externality is that marginal firms congest the learning process. Hence, a benevolent planner would wish to remove these marginal firms in order to increase the probability mass on imitation draws from high-productivity agents. We measure the magnitude of this diffusion externality by comparing the *laissez faire* equilibrium with the efficient one. Optimal policy increases average income substantially (by over 300 percent with our estimated parameters).

We then study how these gains relate to the micro-level changes induced by shocking learning opportunities within the RCT. In the model, there are two broad theoretical forces that govern the planner's ability to extract gains from policy. The first is that agents are able to seek out relatively good matches. The second is that once that matches occur, agents can reap large benefits from them. Both of these forces are parameterized by our RCT estimates. The former is governed by the average treatment effect, while the latter is governed by heterogeneity within the treatment effect. We find that the latter plays a more important quantitative role in our model, accounting for 85 percent of the total gains from policy. However, these same two forces that matter positively for macro-level policy gains are the same forces that *limit* micro-level success of the RCT. For example, while the macro gains are largest when firms can more easily seek out high quality matches, RCT-level gains are minimized in this case – the benefits of an exogenous opportunity to enter a high quality match is naturally quite small if firms can already do so. Similarly, the ability to gain from a good match will imply a faster fade-out of the RCT-level treatment effects, as it allows control firms to more quickly catch up to treated firms. Indeed, our results therefore imply that had we estimated a larger average treatment effect, we would have generated smaller gains from at-scale policy. The quantitative results further imply that macro level gains are primarily a function of heterogeneity in the treatment effect, not the average treatment effect.

The key distinction between these two cases is that while the RCT explicitly attempts to hold the aggregate state of the economy fixed to limit control group contamination,

[3] In Brooks et al. (2018) we consider a number of potential explanations for these results that are independent of diffusion, including profit sharing, loans, and bulk discounts, and find that none of them explain the results.

macro level gains are primarily due to the planner's ability to manipulate the aggregate state. That is, macro gains occur *because* of this contamination. This allows our key estimated parameters play divergent roles at the macro and micro level, and therefore has critical implications for policy. Despite the fact that our RCT essentially mimics the underlying microeconomic diffusion process, a policymaker attempting to naively extrapolate gains from at-scale diffusion policy from such RCT results would be incorrect. These results show that moving from micro empirics to macro consequences requires a way to link these experimental estimates to key structural parameters, which our methodology provides.

1.1 Related Literature

This paper joins a relatively small literature that uses causal empirical estimates to identify critical model parameters in dynamic structural models, including Todd and Wolpin (2006), Kaboski and Townsend (2011) and Brooks and Donovan (2020). Our paper shares a similar style but focuses on knowledge diffusion. Closest in this dimension are Lagakos et al. (2018a) and Akcigit et al. (forthcoming), who use the results from randomized controlled trials to, in part, identify the utility cost associated with migration and a key elasticity to measure the stock of management skills, respectively. We share a similar goal of using a randomized control trial to identify parameters not directly observable in data. Furthermore, our results emphasize caution when trying to infer general equilibrium outcomes from partial equilibrium randomized controlled trials. Buera et al. (2017), Greenwood et al. (2019), and Fujimoto et al. (2019) highlight broadly similar points in microcredit, health, and education, respectively.

Our work adds empirical evidence to the literature studying innovations and knowledge in general equilibrium models (Romer, 1986; Kortum, 1997). Most closely related to this paper is the more recent literature building on these papers, in which diffusion is modeled as a stochastic process of "imitation" including Jovanovic and Rob (1989), Lucas (2009), Alvarez et al. (2008), Lucas and Moll (2014), and Perla and Tonetti (2014). Hopenhayn and Shi (2018) highlight the importance of congestion in a model where all surplus is not captured by the recipient. Recent work has also extended these models to consider within and across firm diffusion (Herkenhoff et al., 2018; Jarosch et al., 2020), international trade (Perla et al., 2020; Buera and Oberfield, 2020), and the interaction of innovation and diffusion (Benhabib et al., 2019).

At the same time, the micro-development literature has long highlighted the importance of diffusing specific pieces of information or technology. See, for instance, work on the diffusion of new crops or high-yielding seeds (Foster and Rosenzweig, 1995; Conley and Udry, 2010), specific planting or production techniques (Atkin et al., 2017; BenYishay and Mobarak, 2018; Beaman et al., 2020), or financial information (Banerjee et al., 2013; Cai and Szeidl, 2018). Our contribution here is two-fold. First, we show that such

RCT results allow tight identification in aggregate models. Therefore, not only can tools from this literature be used to micro-found models, they can also provide useful insights directly to aggregate models in a more "top down" approach. Second, we show that the link from RCT-level evidence to at-scale policy is not obvious in diffusion models, and the two can in fact can be negatively related. This also reinforces the importance of the first point – a link between empirical results and structural parameters is a key input to understanding the link between such empirics and at-scale policy.

2 Identification of the Diffusion Process

We begin by specifying the class of diffusion processes we will study. We start here so that we can clearly lay out the class of models to which the identification results apply. The goal is to lay out the required assumptions without the details of the full model in which we will eventually embed the diffusion process, as they are both cumbersome and unnecessary for the main identification results. Along the way, we will draw attention to the required assumptions so that is clear what is required for the results, and at the end of this section, discuss in detail where the results may fail.

2.1 Setting Up the Problem

Consider a dynamic economy populated by agents with heterogeneous entrepreneurial productivities. We begin by describing how entrepreneurial productivity evolves over time.

Each period, every agent receives two types of shocks to their productivity. First, they receive an idiosyncratic imitation shock $\hat{z}$. If their own productivity z is greater than $\hat{z}$, then the imitation opportunity is useless and it has no effect on the agent's future productivity. If $\hat{z} > z$, then the imitation opportunity contains some useful information that the agent can incorporate into their own future productivity. The intensity with which this imitation opportunity transmits to the agent's productivity in the subsequent period is governed by the parameter β. Second, firms receive random shocks ε that enter the next period's productivity multiplicatively. This shock is assumed to be uncorrelated with own productivity z or the imitation draw $\hat{z}$.[4] The functional form of the subsequent productivity z' is given by Assumption 1.

Assumption 1. *Given a productivity z this period, an imitation opportunity $\hat{z}$, and a random shock ε, productivity next period z' is given by*

$$z' = e^{c+\varepsilon} z^{\rho} \max\left\{1, \frac{\hat{z}}{z}\right\}^{\beta}, \tag{2.1}$$

[4] Note that we need not assume that these are idiosyncratic shocks. They could, for example, have an aggregate and idiosyncratic component where the first affects all agents in the same way. Therefore, we need not assume these shocks are i.i.d. across agents.

where the parameter c is a constant growth term, β is diffusion intensity, and ρ is persistence.

The final term is the benefit to productivity from imitation opportunities. If $\beta = 0$, this law of motion collapses to a standard exogenous AR(1) process, $\log(z') = c + \rho \log(z) + \varepsilon$. On the other hand, $\beta > 0$ allows productivity to increase when presented with an opportunity to imitate some $\hat{z} > z$. Furthermore, notice that the max operator in the diffusion process rules out any productivity benefit accrued to a higher productivity firm from interaction with lower productivity firms (as in Jovanovic and Rob, 1989, for example). We address this issue directly in Section 3 and find no evidence that more productive firms gain profit from interaction with less productive firms.[5] For simplicity throughout, we refer to β as the "intensity" of diffusion and ρ as "persistence."

Given the notion of productivity we consider here, we cannot observe it directly. Thus, we require a link between productivity and observable variables, in the case, profit. The requirement is summarized in Assumption 2.

Assumption 2. *In any period, profits are proportional to productivity. That is, for any two firms i and j earning profits π_i and π_j, $\pi_i/\pi_j = z_i/z_j$.*

This assumption is satisfied by much of the literature on diffusion. A simple way to satisfy Assumption 2 is to assume $\pi_i = z_i$ as in Lucas (2009) and Perla and Tonetti (2014). A production function of the form $y = z^{\alpha} n^{1-\alpha}$, where n is labor, also satisfies Assumption 2 in a competitive labor market.[6][7]

Finally, we specify the assumptions on the distribution from which $\hat{z}$ is drawn. We denote the cumulative density function of $\hat{z}$ as $\widehat{M}(\hat{z}; z, \theta)$. Writing it in this way emphasizes that agents with different productivities z may draw from different distributions, and that these distributions depend on a parameter θ. In particular, this parameter is assumed to order a class of distributions in the sense of first order stochastic dominance. This is summarized in Assumption 3.

Assumption 3. *The imitation opportunity $\hat{z}$ is drawn by a firm with productivity z from a distribution characterized by the cumulative density function $\widehat{M}(\hat{z}; z, \theta)$, a known function. For every z and $\hat{z}$, $\widehat{M}$ is continuous in θ and $\theta_1 < \theta_2 \implies \widehat{M}(\hat{z}; z, \theta_2)$ first order stochastically dominates $\widehat{M}(\hat{z}; z, \theta_1)$.*

[5]The assumption of no productivity gain accruing to the more productive firm is not a critical one. We could alternatively allow for it, though looking ahead, our empirical results would require this channel to be shut down. We therefore exclude it for simplicity.

[6]Note, however, that assumption is violated in the presence of firm-specific distortions, such as those considered in Hsieh and Klenow (2009). In the Appendix we argue that such distortions would imply that our estimated parameter values are attenuated, suggesting that we are underestimating the effects of diffusion. We further show how the results change as the importance of such distortions vary.

[7]This assumption is again not strictly required, as long as a method exists to identify z. For example, using input and output data, one could estimate a production function from the control group and recover productivity as a residual. What we require is the existence of a mapping from observables to unobserved productivity z, but the exact details of that mapping are immaterial for our identification results to hold. Thus, we err on the side of simplicity here.

This assumption admits a variety of search and assignment processes. For example, one commonly used diffusion process is that agents draw randomly from the existing firms. Denoting M as the cdf of operating-firm productivity, this would imply $\widehat{M}(\hat{z}; z, \theta) = M(\hat{z})$. Even within the random search framework, Assumption 3 allows us to be somewhat broader, as agents may draw from better or worse distributions that the set of operating firms, where θ indexes how much the distribution of matches differs from the firm productivity distribution. We discuss this assumption in more detail in Section 2.3. We refer to θ as the "directedness" parameter as shorthand, with the understanding that this is a technological parameter in the model.

The assumptions laid out in this section allow us to do two things. First, they let us translate a broad, unobservable notion of productivity to an observable characteristic, profit. Second, they parametrize the forces of diffusion we wish to investigate. Intensity β captures the static effect that governs how much individuals gain immediately from a match. Persistence ρ governs how much of a past match can be transmitted in the future, thus contributing to the dynamic impact of a single match. Finally, directedness θ governs who individuals regularly interact with. All three of these play a potentially important role in governing the total impact of diffusion.

Finally, while we note that all of the assumptions we have made are common in the literature, one can of course come up with models that do not satisfy our assumptions. While we emphasize that our goal in this paper is not to distinguish various diffusion models that one could conceive of (and our empirics are not well-suited to this task), we come back to this issue in Section 2.3 and discuss the limits of the structural assumptions made above to hopefully provide some broader context for our results.

2.2 Variation Required to Identify Diffusion

Section 2.1 laid out a set of assumptions on the primitives of the model. Our goal now is to identify three key diffusion parameters – the intensity of transmission β, the persistence of productivity ρ, and the parameter controlling the distribution of imitation draws θ – without imposing any additional structure on the economy. To that end, Assumption 4 summarizes variation in the data required to identify the parameters. After proving the identification results, Section 3 details a randomized controlled trial that satisfies these assumptions, thus allowing us to take the model to the data.

Assumption 4. *A set of agents with productivity distributed $H(z)$ are observed in two consecutive periods. The set of agents is partitioned into two subsets characterized by distributions $H_C(z)$ and $H_T(z)$ (i.e., "control" and "treatment"). The following conditions hold:*

1. *Agents in H_T and H_C draw their ε shocks from the same distributions*
2. *The matches for agents in H_C are not observable, and distributed $\widehat{M}(\hat{z}; z, \theta)$*

3. *The matches for agents in H_T are observable, and distributed $\widehat{H}_T(\hat{z}) \neq \widehat{M}(\hat{z}; z, \theta)$. Moreover, every match $\hat{z}$ is greater than the z to which it is matched.*

4. *For any arbitrary partition of the treatment group, characterized by $H_T^1(z)$ and $H_T^2(z)$, agents in both groups draw their ε shocks from the same distribution*

The first assumption imposes the usual exclusion restriction – that unobserved characteristics do not systematically vary across treatment and control groups. The second formalizes the intuitive notion that we cannot observe control group matches, and they proceed as defined by the $\widehat{M}$ function. That is, control group continues to match as defined by the underlying economy.[8] Finally, the third and fourth lay out what we require from our treatment. The third states that we can observe all treatment matches, and those matches are drawn from some other distribution than the control group. Moreover, we assume that treatment firms are always matched to a more productive agent.[9] Finally, the last assumption states a second exclusion restriction within the treatment group, guaranteeing that comparisons across treatment firms are unbiased.[10]

Our procedure works as follows. Using only the treatment firm data, we show how to identify β and ρ uniquely. We then add back the control data to show that θ can be identified from the average treatment effect as only a function of (β, ρ). Thus, the three parameters are uniquely identified under Assumption 4.

Using Treatment Data for (β, ρ) Identifying the intensity and persistence parameters use only variation from treatment matches. This has the benefit that the identification follows almost directly from the law of motion for diffusion. In logs, the law of motion is

$$\log(z_i') = c + \rho \log(z_i) + \beta \log\left(\max\left\{1, \frac{\hat{z}_i}{z_i}\right\}\right) + \varepsilon \tag{2.2}$$

Applying Assumption 2 ($\pi \propto z$) annd Assumption 4 ($\hat{z} > z$), this simplifies to

$$\log(\pi_i') = \widetilde{c} + \rho \log(\pi_i) + \beta \log\left(\frac{\hat{\pi}_i}{\pi_i}\right) + \varepsilon_i. \tag{2.3}$$

where π_i and $\hat{\pi}_i$ are baseline profit for individual i in the treatment group and her match and $\widetilde{c}$ is a constant equal to the structural parameter c if $\pi = z$. Equation (2.3) is a linear regression that can be estimated directly from panel data. Intuitively, there are two ways to read (2.3). The first is that β measures the impact of receiving a better match (higher $\hat{\pi}_i/\pi_i$), controlling for initial income π_i. Alternatively, one could read this as estimating

[8] Note at this point we still do not know the parameter θ. This assumption states that control matches occur via the (known) function $\widehat{M}$ indexed by some unknown parameter θ.

[9] As we discuss later, the assumption that $\hat{z} > z$ for all z in the treatment is not critical for the results but drastically simplifies the formal proof. In the Appendix, we show that Assumption 4 without the $\hat{z} > z$ assumption is still sufficient for identification in the treatment.

[10] While this paper proceeds using randomization as a way to generate variation consistent with Assumption 4, it is worth emphasizing that it is not necessarily required. Any instrument that satisfies these assumptions would be equally valid for the results to hold.

the decay in profit, measured by ρ, after controlling for variation in match quality ($\hat{\pi}_i/\pi$). Given the orthogonality built into the data-generating process by Assumption 4, (2.3) measures both forces simultaneously. Proposition 1 formalizes this.

Proposition 1. *The estimates* $(\hat{\beta}^{OLS}, \hat{\gamma}^{OLS})$ *from (2.3) identify parameter values* (β, ρ)

Proof. Follows directly from the within-treatment exclusion restriction of Assumption 4. ■

The argument laid out above relies in part on the result that one can remove the max operator from (2.2) via the assumption that $\hat{z} > z$. In the Appendix, we show that this assumption is not required to identify β and ρ. The intuition is identical to that laid out above, but the max operator introduces a bias that must be taken into account directly. This requirements a two-step procedure, and thus existence and uniqueness require a more substantive discussion.

"Directedness" of diffusion θ Now, we utilize both treatment and control groups to identify θ, which controls the distribution of imitation draws. We admit from the outset that we cannot observe individual-level matches in the control group. The critical insight here is that we can draw inference about the control group by differencing from the treatment. Since treatment firms are guaranteed a high productivity match, observing small differences in average *ex post* profit implies that control firms must also be drawing from a distribution with substantial mass on high productivity matches. Or put in our notation, $\widehat{M}$ must be indexed by a high θ. Similarly, large differences in average profit between treatment and control implies that the guarantee of a high productivity match generates a large effect precisely because high productivity matches are not usually realized. This corresponds to a low value of θ. The average difference in profit therefore allows us to infer θ, despite not observing the underlying matches in the control group. Figure 1 shows the intuition graphically.

To formalize this argument, define $\bar{z}_T$ and $\bar{z}_C$ as average *ex post* productivity the treatment and control groups. Following a similar procedure as above, the law of motion for productivity (Assumption 1), combined with the implied variation in matches (Assumption 4), implies

$$\frac{\bar{z}_T}{\bar{z}_C} = \frac{\int\int\int e^{c+\varepsilon} \max\left[z, \hat{z}^{\beta} z^{1-\beta}\right]^{\rho} dF(\varepsilon) d\widehat{H}_T(\hat{z}, z) dH_T(z)}{\int\int\int e^{c+\varepsilon} \max\left[z, \hat{z}^{\beta} z^{1-\beta}\right]^{\rho} dF(\varepsilon) d\widehat{M}(\hat{z}; z, \theta) dH_C(z)}$$

Defining $\Gamma_3 := \bar{z}_T/\bar{z}_C$ and using the orthogonality of the exogenous shocks ε, we can re-write the equation as

$$\Gamma_3 = \frac{\int\int z^{\rho} \max\left[1, \hat{z}/z\right]^{\beta} d\widehat{H}_T(\hat{z}) dH_T(z)}{\int\int z^{\rho} \max\left[1, \hat{z}/z\right]^{\beta} d\widehat{M}(\hat{z}; z, \theta) dH_C(z)}. \tag{2.4}$$

Figure 1: Identification of θ

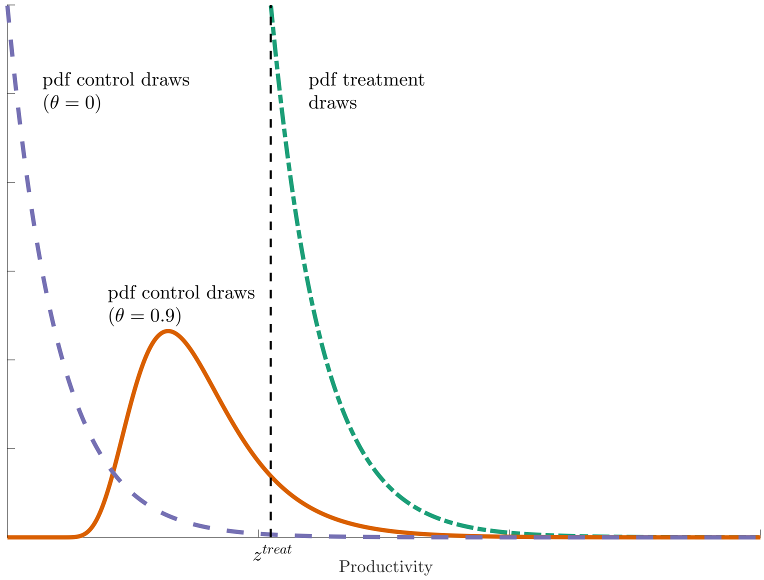


Figure notes: Figure shows the distribution of draws for treatment and control firms under different θ. The distributions are drawn Pareto, but this is for the example's sake only.

Given the values of intensity β and persistence ρ already identified, then all other parts of this equation come directly from the data (after again applying Assumption 2 that $\pi \propto z$), except for the parameter θ.[11] The assumed monotonicity of $\widehat{M}$ (Assumption 3) is sufficient to prove that any θ that solves this equation is unique. Proposition 2 formalizes the results, developing bounds to guarantee the results.

Proposition 2. *Given the values (β, ρ), the value of θ that solves (2.4) is unique if $\Gamma_3 \in [\Gamma_3^{min}, \Gamma_3^{max}]$, where*

$$\Gamma_3^{min} = \inf_\theta \frac{\int\int z^\rho \max\left[1, \hat{z}/z\right]^\beta d\widehat{H}_T(\hat{z})dH_T(z)}{\int\int z^\rho \max\left[1, \hat{z}/z\right]^\beta d\widehat{M}(\hat{z}; z, \theta)dH_C(z)} \tag{2.5}$$

$$\Gamma_3^{max} = \sup_\theta \frac{\int\int z^\rho \max\left[1, \hat{z}/z\right]^\beta d\widehat{H}_T(\hat{z})dH_T(z)}{\int\int z^\rho \max\left[1, \hat{z}/z\right]^\beta d\widehat{M}(\hat{z}; z, \theta)dH_C(z)}. \tag{2.6}$$

Proof. First, note that the only unknown on the right hand side of (2.4) is θ. This follows from Assumption 2 and Proposition 1. All that is left is to show that there exists a unique θ that solves (2.4). The right hand side is continuous in θ by the continuity of $\widehat{M}$ in Assumption 3. The intermediate value theorem then guarantees existence when $\Gamma_3 \in [\Gamma_3^{min}, \Gamma_3^{max}]$. Finally, uniqueness follows from the strict monotonicity of the right hand side in θ, which is guaranteed by the first order stochastic dominance assumed in

[11] This is conditional on Assumption 3, which assumes that $\widehat{M}$ is a known function. That is, we require that $\widehat{M}$ is a known function indexed by an unknown parameter θ. It is in this sense that this paper is not designed to distinguish between different possible matching technologies.

Assumption 3. ■

Thus, the three parameters (β, ρ, θ) are uniquely identified with the variation in the data-generating process laid out in Assumption 4.[12]

2.3 Discussion

Before turning to the estimation and RCT results, it is worth discussing some context for the preceding identification results, and laying out what the above procedure can and cannot accomplish.

2.3.1 More Detail on Identification Results

First, the empirical moments required for the estimation are are easily obtained from data, using only the average treatment effect and a properly-adjusted lagged profit regression. Thus, the moments allow for a relatively straightforward link between model and data.

Second, we note that the identification results in Proposition 1 are actually broader than we have written them. In theory, Assumption 4 allows us to semi-parametrically identify (ρ, f) in the law of motion

$$\log(z') = c + \rho \log(z_i) + f(\mathbf{x}, \widehat{\mathbf{x}}) + \varepsilon \tag{2.7}$$

where $f : X \times \widehat{X} \to Z$ is a function that takes any combination of firm $(\mathbf{x})$ and match $(\widehat{\mathbf{x}})$ characteristics and translates them into a contribution to future productivity. Again, the key is the exclusion restriction. Since, in theory, all characteristics are randomly assigned and thus orthogonal to the error term, they can be included in the regression. As a simple example, imagine the treatment effect varied in the age gap between the two firm owners. We could add interaction terms to our original regression (2.3), which would allow us to identify age bin-specific β's instead of a single β. Taking that type of intuition to its limit, we can identify the function f in (2.7).

On a more practical level, doing so requires both a substantial sample size and enough characteristic variation to be properly powered for such a test. Thus, we do not pursue this further. We note it only to highlight that the methodology itself is potentially much broader than we have written it here, and useful for other types of questions one may wish to ask in these models.[13]

[12] The bounds in Proposition 2 are only to guarantee the treatment effect stays within the set of values the model can possibly rationalize. For example, if $\beta = 0$, (2.4) shows that the model cannot rationalize any positive average treatment effect. The quantitative results are all within the required bounds for identification.

[13] Most directly, one could use the empirical results of an RCT to identify which types of characteristics play the largest role in productivity transfer, which may be useful to better design matches in future policy. Relatedly, it could provide empirical support for the inclusion or exclusion of characteristics from an individual's state variable in a model.

2.3.2 Where the Results Fail

An important question is the extent to which our identification results are robust to other economies, or put differently, where our assumptions fail.

First, the identification of β and ρ holds in a broad set of economies. The key here is the power of the (hypothetical) design. That the matches are randomized and observable within the treatment group implies only treatment firm data are required to identify (β, ρ). Thus, the details of who searches, or why, is irrelevant for the estimation (conditional on Assumptions 1 and 2). However, one way in which Assumption 2 ($\pi \propto z$) fails is if firms are subject to idiosyncratic distortions ν_i. Then, we instead have $\pi \propto z\nu$. We take this up in the Appendix and show that our results would then be a lower bound on the size of the diffusion externality.

A more subtle restriction is built into our assumption on $\widehat{M}$ in Assumption 3. Here, we require that the draw of a match $\hat{z}$ depends only on z and a parameter θ. This assumption nests as a special case work by Jovanovic and Rob (1989), Lucas (2009), and Buera and Oberfield (2020), who assume uniform draws from the existing distribution of operating firms. In that case, if M is the cdf of operating firm productivity, $\widehat{M}(\hat{z}; z, \theta) \equiv M(\hat{z})$. Lucas and Moll (2014) and Perla and Tonetti (2014) make the same uniform draw assumption, but extend these models by endogenizing a tradeoff between production and searching for a match. Models with this tradeoff generally fail Assumption 3, because the decision to search depends on the remaining details of the model and equilibrium. In this case, we lose the independence of θ from the remaining model structure.

Three things are worth emphasizing about this result though. First, assumptions on $\widehat{M}$ only affect the estimation of θ. If we followed the literature and fixed θ *ex ante*, the identification of β and ρ goes through unchanged. Second, Assumption 3 still allows for a wide set of underlying processes. We detail a number of different underlying models that satisfy our assumptions in the Appendix, and highlight the variety of interpretations one can put on θ depending on the exact model details. Finally, even if one is not willing to fix θ, the moment itself can still be quite useful. As long as the $\widehat{M}$ function still satisfies the first order stochastic dominance assumption conditional on other parameter values, there still exists a (conditional) one-to-one mapping from θ to the average treatment effect.

3 Application to Kenyan Firms

With the identification results in hand, we now turn to the data. We detail the randomized controlled trial that allows us to estimate the parameters in the previous section, then estimate these parameters. A complete description of the program and reduced-form results are available in Brooks et al. (2018), though we reproduce some of the relevant results here for simplicity's sake.

Our experimental design randomly matches older, profitable entrepreneurs with younger entrepreneurs. The younger owners were then followed for over 17 months to measure changes in business practice and profit over time. Outcomes are compared to a control group of similar firms.[14] It is important to note that while this RCT in some sense takes the random matching quite literally, this is not required to implement our procedure. Any policy intervention that generates the requisite orthogonality conditions could be similarly utilized, including natural experiments that generate the requisite change in imitation opportunities.

3.1 Details of RCT and Data Collection

The experiment took place in Dandora, Kenya, a dense urban slum on the outskirts of Nairobi. Self-employment is ubiquitous in Dandora with a huge number of street-level businesses operating in a variety of industries, such as retail, simple manufacturing, repair and other services. We began by conducting a large scale cross-sectional survey. We sampled a random cross-section of 3290 businesses. Our goal was for this sample to be representative of the population of enterprises, and it includes businesses of a variety of ages and industries. This sample is used to estimate moments of the population of operating firms.

Qualitative Evidence on the Importance of Learning To begin, Figure 2 plots business scale measures based on self-reported learning methods from the baseline survey. Fifty-five percent of all firms claimed they were self-taught, while the rest claimed to learn either from another business operator, in school, or through an apprenticeship. Figure 2a shows that the self-taught earn less profit at any point over the lifecycle. The average profit of a self-taught firm is 18 percent less than firms that learn from others, while Figure 2b show that self-taught firms pay a smaller total wage bill.

Selection and Randomization We start from a sample of female business owners who have been in operation for less than 5 years.[15] We then randomly select a subset of these business owners to randomly match with an older, more experienced owner. In this way, we guarantee a high quality match for these business owners (in an intent-to-treat sense). Thus, the randomization allows us to compare the owners chosen into the treatment against those other business owners who were not.[16] Firms were then surveyed

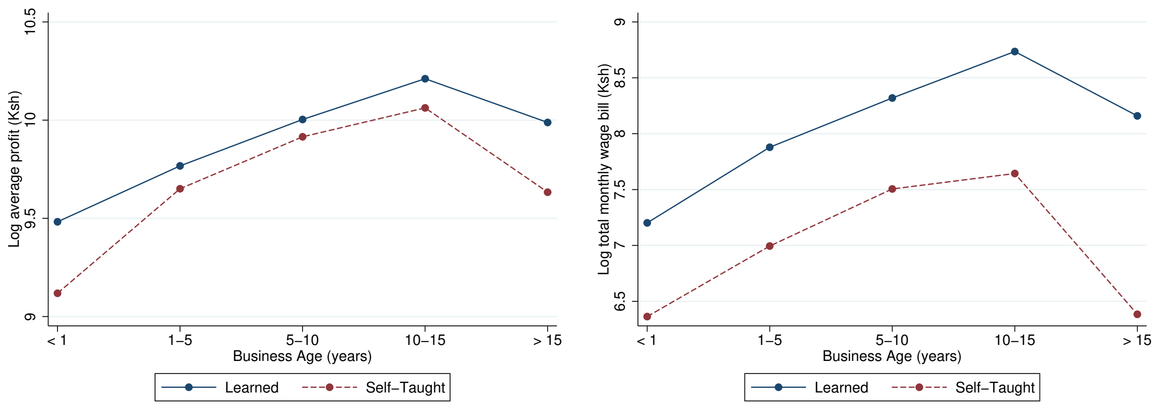
[14]In Brooks et al. (2018), we further randomize another group into formal business training classes. While we do not utilize this classroom training treatment arm here, it is interesting to note that the results differ substantially across these treatment arms. We show that this to the fact that matching with local firms provides specific information about the local economy whereas classroom training provides information on topics that are designed to be orthogonal to the market in which they are deployed (accounting, marketing, etc.).

[15]The sex selection criteria is to limit heterogeneity outside the model. Note, however, that females make up 65 percent of business owners in Dandora and 71 owners with businesses open less than 5 years.

[16]Note that this procedure satisfies all the requirements in Assumption 4. We do not assume we can observe control matches (part 2 of Assumption 4), but the randomization immediately satisfies the exclusion restriction (part 1). Our selection the treatment matches satisfies the final aspect of Assumption 4 when combined with Assumption 2.

Figure 2: Self-Reported Learning Methods and Business Scale

(a) Average profit (b) Total wage bill

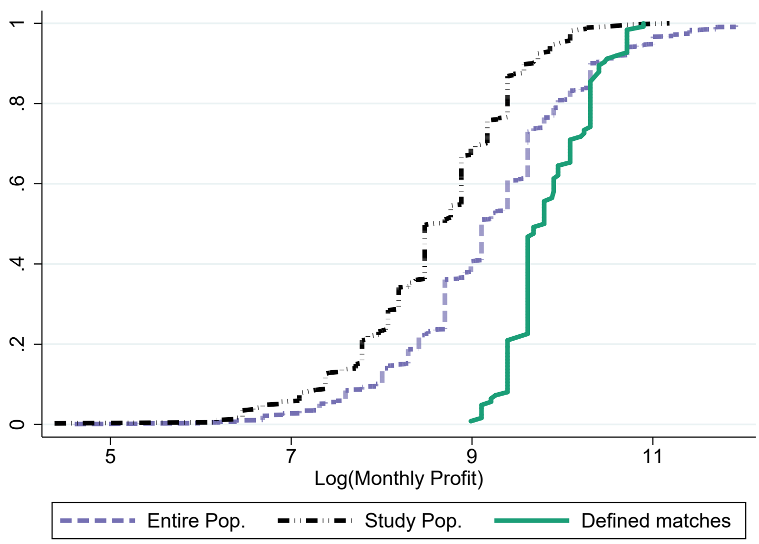


over 6 quarters to track the time series of treatment.

The older business owners who entered into a match were selected from those businesses with owners over 40 years old and at least 5 years of experience. This was to minimize the importance of "luck" in baseline profit realizations to allow us to focus on truly productive business owners. We then recruited business owners with the highest profit until we had a sufficient number for matches. Of those contacted, 95 percent accepted. Matches with the treatment firms were randomly created conditional on industry.

To summarize, Figure 3 plots the cumulative distribution function of baseline profit for the entire sample, the population we study, and the selected matches. One can see that our study population is somewhat poorer than the entire population, while the matches are drawn from the far right tail of the baseline profit distribution.

Figure 3: Baseline Profit Distributions



Details of a "Match" What does it mean to enter into one of our matches? We designed the program to remain as truthful to the theoretical counterpart of the model as possible. First, matches were designed to only last for one month, though of course there was no restriction on meeting after the formal end of the program.

The program was pitched to both sides of the match as a mentee-mentor relationship, and thus was explicitly focused on business success. The older, more successful business owners were the "mentors," while the younger owners were the "mentees," consistent with both their profitability and time engaged in business. The mentors were told they could potentially help other business owners learn the requisite skills required to operate in Nairobi. However, we provided no topics to discuss, instead preferring that the content of any discussions was self-directed. As we show later, there is substantial heterogeneity in topics discussed. After signing up mentors we simply provided the mentees with the mentor's phone number and told them that a prominent business owner in Dandora was willing to discuss business questions with them. Whether they contacted the mentor, or ever met, was their decision. However, all matches met at least once in the official month-long treatment period.[17] For simplicity and ease of reference to the more detailed discussion in Brooks et al. (2018), we refer to these two groups as mentees and mentors throughout. We emphasize, however, that they should more generally be thought of as the more and less productive members of a match.

3.2 Balance

Since our theoretical results rely on two layers of randomization, we need to verify balance both on between control and treatment and within treatment. Brooks et al. (2018) shows that the control and treatment groups are balanced. Here, we conduct a second balance test

$$y_{i0} = \alpha_0 + \alpha_1 \mathbf{M}_i + \varepsilon_i$$

where $\mathbf{M}_i$ is an indicator denoting that firm i is a treatment firm matched with a bottom 25th percentile (denoted M_L), 25-75 percentile (M_M), or top 25 percentile firm (M_H) in terms of baseline profitability.[18] Table 1 reports the results. The only significant difference is in age, and the magnitude is small.

3.3 Estimating Diffusion Parameters from the RCT Results

In Brooks et al. (2018), we show that in the pooled regression over 6 quarters mentees see a statistically significant increase in profit relative to control, and moreover, the impact

[17] One might be concerned that we indirectly primed mentees to believe these matches would be beneficial. We can do little to rule this out completely. We note, however, that evidence of the mentor's business success are easily visible to the mentee. Mentors had substantially more physical capital and workers, and had a fixed, relatively large buildings from which they conducted business. Moreover, the first meeting took place at the mentor's business. Thus, that the mentor was "good" at running a business would likely have been understood with or without us.

[18] We have experimented with a number of different ways to compute the balance table, and all show the same results.

Table 1: Balancing Test at Baseline

	Control Mean (1)	M_L - Control (2)	M_M - Control (3)	M_H - Control (4)
Firm Scale:				
Profit (last month)	10,054	-732.65	-1337.06	-760.08
		(1314.56)	(1393.38)	(2128.41)
Firm Age	2.39	0.04	-0.19	0.08
		(0.28)	(0.30)	(0.46)
Has Employees?	0.25	-0.10	-0.07	0.10
		(0.07)	(0.07)	(0.11)
Number of Emp.	0.23	-0.05	0.00	0.18
		(0.08)	(0.08)	(0.13)
Business Practices:				
Offer credit	0.74	-0.07	0.04	-0.03
		(0.07)	(0.08)	(0.12)
Have bank account	0.30	-0.04	-0.05	0.06
		(0.07)	(0.08)	(0.12)
Taken loan	0.14	-0.07	-0.06	0.03
		(0.05)	(0.05)	(0.08)
Practice accounting	0.01	-0.01	0.01	-0.01
		(0.01)	(0.02)	(0.02)
Advertise	0.07	0.04	0.01	0.11
		(0.05)	(0.05)	(0.07)
Sector:				
Manufacturing	0.04	-0.02	-0.04	-0.04
		(0.02)	(0.03)	(0.04)
Retail	0.69	-0.03	0.00	-0.10
		(0.08)	(0.08)	(0.12)
Restaurant	0.14	-0.06	0.00	0.03
		(0.05)	(0.06)	(0.09)
Other services	0.17	0.09	0.02	0.07
		(0.06)	(0.07)	(0.10)
Owner Characteristics:				
Age	29.1	0.92	-1.88	0.50
		(0.79)	(0.84)**	(1.28)
Secondary Education	0.51	0.02	-0.08	0.13
		(0.08)	(0.09	(0.13)

Table notes: Columns 1-4 are the coefficient estimates from the regression above, with column one being the estimate of the constant $\widehat{\alpha}_0$. Statistical significance at 0.10, 0.05, and 0.01 is denoted by *, **, and, ***. All constants are significant at one percent.

is increasing in mentor profit. Here, we restrict attention to the baseline and the survey wave 3 months post-treatment. We will estimate parameters off these two quarters. Later we test whether the impulse response of the model-computed RCT results match the remaining empirical time series, and find that it does.

Denoting the set of individuals in the treatment as $\mathbf{M}$ and control as $\mathbf{C}$, we first estimate equation (2.3) on treatment firm data,

$$\log(\pi_i') = \tilde{c} + \rho \log(\pi_i) + \beta \log\left(\frac{\hat{\pi}_i}{\pi_i}\right) + \varepsilon_i \qquad \text{for } i \in \mathbf{M}.$$

As discussed in Section 2, this identifies β and ρ. We then require the average treatment effect to measure θ,

$$\pi_i' = \gamma + \nu \mathbb{1}[i \in \mathbf{M}] + \xi_i \qquad \text{for } i \in \mathbf{M} \cup \mathbf{C}.$$

Both regression results are provide in Table 2.

Table 2: Identification Moments

	(1)	(2)
β	0.538	
	(0.273)**	
ρ	0.595	
	(0.273)**	
Treatment		891.990
		(280.720)***
R^2	0.053	0.047

Table notes: Standard errors are in parentheses. The top and bottom one percent of dependent variables are trimmed. Statistical significance at 0.10, 0.05, and 0.01 is denoted by *, **, and, ***.

The estimates of β and ρ in column (1) can be directly ported into the model as their structural counterparts. The results show that the technological component of productivity persistence is $\rho = 0.595$, while the intensity parameter is $\beta = 0.538$. The former is lower than persistence estimated in rich countries. This is consistent with the wide variety of topics and problems discussed during business interactions, and provides an interesting avenue for misallocation where persistence plays an important role (Buera and Shin, 2011; Moll, 2014). Comparisons of β are naturally more difficult to come by.[19] We study the importance of both parameters in the quantitative results.

The average treatment effect in column (2) of Table 2 is not directly equal to θ. This requires us to take a stand on the functional form of the imitation draw distribution, $\widehat{M}$. We assume that $\widehat{M}$ takes the form

$$\widehat{M}(\hat{z}) = M^f(\hat{z})^{1/(1-\theta)}$$

where M^f is the c.d.f. of the existing firm distribution (which we observe directly in the data). That is, we assume that draws are random from the set of existing firms, adjusted by the parameter θ. When $\theta = 0$, this is the usual uniform random matching assumption. $\theta \to 1$ implies that all mass in the distribution of imitation draws concentrates on the upper bound of the productivity distribution. As $\theta \to -\infty$, imitation draws come from the lowest z firms, thus implying that no operating firm receives a useful opportunity from imitation. Note that this formulation allows for the possibility that our intervention has an effect on members in the match (via $\beta > 0$), but none of those gains diffuse in

[19] Models studying balanced growth in this context usually require a condition similar to $\beta = 1$ to guarantee the the tail of the productivity distribution maintains it shape over time. Because we estimate $\beta < 1$ the model is generically not consistent with balanced growth and thus cannot be compared to those model parameters.

equilibrium (via $\theta \to -\infty$). Using the results in column (2) implies $\theta = -0.417$. These results form our diffusion parameter estimates.

3.4 Impact on Higher Profit Business Owner

Though not directly related to the estimation, we note that the diffusion process in Section 2 assumes that there should be no gains to the more productive members of the match via the use of the max function in law of motion for productivity (Assumption 1). These individuals were not randomly selected relevant to their peers, and thus cannot be directly compared to a control group. However, our design allows us to use the selection procedure to identify the causal impact of being chosen using a regression discontinuity after resurveying both those chosen for the program and those just below the cutoff for selection. We find no change in profitability, scale, or any practices one may associate with productivity (e.g., better book keeping, more marketing). The details and robustness of these results are available in Brooks et al. (2018) but we reproduce them in the Appendix for simplicity's sake.

3.5 Discussion During Meetings

As part of the study, we recorded the topics discussed during meetings between mentors and mentees. Discussions varied both within and across matches, highlighting the wide variety of skills and problems that make up firm-level productivity (and, at a more micro level, the difficulty inherent in designing firm training curriculum). Figure 4a plots the share of businesses that discuss each of 10 topics with their mentors. Topics include attracting customers, keeping records, lowering costs, and types of products. Moreover, Figure 4b show that this is not only a cross-firm phenomenon, but within-firm as well. Over 50 percent of treatment firms discuss at least 5 of the 10 listed topics with their mentors, with nearly 20 percent discussing all 10.[20]

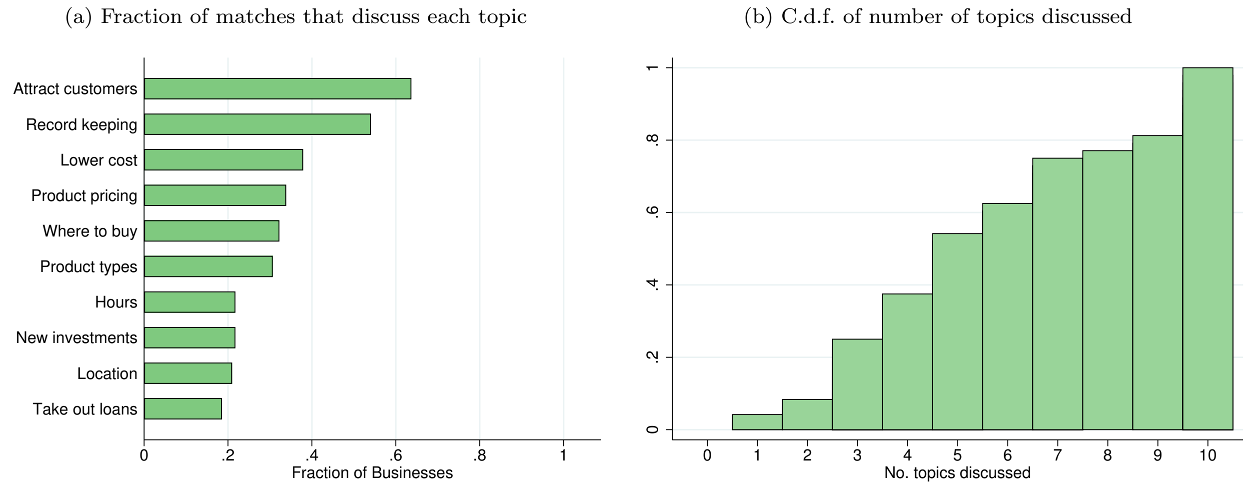
We view this as evidence that there is not one specific issue that constraints these firms, but instead that z is collectively comprised of various issues that face firms in developing countries. It also highlights the role played by open-ended knowledge transfer, instead of focusing on a specific topic.

4 Full Model

We now close the model to study the quantitative importance of diffusion. As we have emphasized throughout, this is only one potential model in which one could deploy these results. However, because measuring the impact of diffusion requires the solution to a fixed point problem, the remaining structure is required to compute the effect. Naturally,

[20]We included an "other" option that allowed for any additional topics that may have arisen, but it was rarely selected. These 10 topics seemed to cover the vast majority of discussion in matched firms.

Figure 4: Discussion Topics within Matches



the exact policy levers used and the quantitative magnitudes change depending on the model specification, but the estimated diffusion parameters do not.

We build a model in which agents can act as firms or workers, re-optimizing their occupation each period. To model diffusion, we assume random search, adjusted by the directedness parameter θ. The rationale for this is because random search mirrors closely our RCT, in which we randomly match firms. As we will show later, even in this similar context, RCT and GE results can look quite different.

Model Basics Time is discrete and infinite. In each period there is a unit mass of risk-neutral agents. Each agent has an exogenous probability δ of dying each period, while δ agents are born. Each agent is characterized by productivity z which evolves over time via the diffusion process laid out in Assumption 1.

Occupational Choice and Recursive Formulation In every period, each agent can choose to be a worker or an entrepreneur. Workers sell their labor to entrepreneurs for the market clearing wage w, while entrepreneurs produce an undifferentiated consumption good using their skill and hired labor. Worker wages are taxed at rate τ.[21]

An entrepreneur's profit is

$$\pi(z) = \max_{l \geq 0} \; z^{\alpha} l^{1-\alpha} - wl \tag{4.1}$$

where w is the equilibrium wage. Recursively, the value of having entrepreneurial skill z is

$$v(z, M) = \max\{\pi(z), w(1-\tau)\} + (1-\delta)\gamma \mathbb{E}_{z'|z} v(z', M') \tag{4.2}$$

[21] This distortion is set to generate the correct share of entrepreneurs and workers in the economy. It could, for example, be a stand-in for search costs. Alternatively, one could assume individuals differ in some non-pecuniary benefit between the two occupations, such as entrepreneurship providing a more flexible work schedule.

where M is the equilibrium distribution of productivity, and is the aggregate state of the economy. Solving the entrepreneur's problem yields

$$\pi(z) = \alpha \left(\frac{1-\alpha}{w} \right)^{\frac{1-\alpha}{\alpha}} z$$

which satisfies Assumption 2 on the proportionality of profit and productivity. This further implies that agents face a cutoff rule to determine their occupation. For a given wage w, there is a $\underline{z}(w)$ such that any agent with $z < \underline{z}$ becomes a worker, while agents with $z \geq \underline{z}$ become entrepreneurs.

Diffusion Continuing agents have productivity that evolves according to our Assumption 1,

$$\log(z') = c + \rho \log(z) + \beta \log \left(\max\{1, \hat{z}/z\} \right) + \varepsilon. \tag{4.3}$$

As discussed in Section 3.3 (and with a slight change in notation to remain consistent with the model), we assume matches are drawn from the operating firm distribution, adjusted by θ:

$$\widehat{M}(\hat{z};\theta) = \begin{cases} 0, & \text{if } \hat{z} < \underline{z} \\ \left(\frac{M(\hat{z}) - M(\underline{z})}{1 - M(\underline{z})} \right)^{\frac{1}{1-\theta}}, & \text{if } \hat{z} \geq \underline{z} \end{cases} \tag{4.4}$$

Note that a key difference from the earlier discussion is that $\widehat{M}$ now depends on the economy-wide productivity distribution M and the cutoff $\underline{z}$, both equilibrium objects. It must therefore be consistent with the the diffusion process in the economy. The law of motion for M is

$$\begin{aligned} M'(z') \quad &:= \quad \Lambda(M(z')) = \qquad (4.5) \\ \delta G(z') \quad &+ \quad (1-\delta) \int_0^\infty \int_0^\infty F(\log(z') - \log(\max\{\hat{z}^\beta z^{\rho-\beta}, z^\rho\}) - c) d\widehat{M}(\hat{z};\theta) dM(z) \end{aligned}$$

where G is the exogenous distribution from which new entrants draw productivity.[22]

4.1 Definition of Equilibrium

A competitive equilibrium of this economy is a wage function w, a distribution of productivities M, and a value function v such that v satisfies (4.2) with the associated decision rules for labor and occupational choice, the evolution of M is consistent with the decision rules and is given by (4.5), the wage w clears the labor market, which requires a solution

[22] Other papers, such as Luttmer (2007) and Da Rocha and Pujolas (2011), assume that G varies with the existing distribution of productivity. This would have no effect on any of our identification results, and thus we exclude it for simplicity.

to the implicit equation

$$w = (1-\alpha)\left(\frac{\int_{\underline{z}(w)}^{\infty} z dM(z)}{M(\underline{z}(w))}\right)^{\alpha}.$$

A stationary competitive equilibrium is a competitive equilibrium in which the distribution M^* is such that $\Lambda(M^*) = M^*$.

4.2 Calibration of Remaining Parameters

The remaining parameterization of the model follows relatively standard calibration procedures and we choose parameters to match moments of the same set of firms in which the experiment was conducted. We make use of both the baseline field data that conducted on a random subset of firms in Dandora, Kenya. Care was taken in collecting this data that it be representative of the whole population of operating firms in the area, and we use it here to measure the distribution of operating firms.

The model parametrization can be broken into three different parts that can be considered separately. First, as we showed previously, the diffusion parameters are independent of the remaining model parameters. Thus, we can simply impose our estimated parameters $\beta = 0.538$, $\rho = 0.595$, and $\theta = -0.417$.

The remaining parameters are the death rate of agents δ, the labor share of output α, the growth term c, the exogenous distribution of shocks F, and the exogenous distribution of entrants G. We assume that G is log-normally distributed with parameters μ_0 and σ_0, and that F is log-normally distributed with parameters μ and σ. We normalize $\mu_0 = 0$. We note that c and μ are not separately identified, so we choose $\mu = -\sigma^2/2$ so that $E[e^{\varepsilon}] = 1$. We set $\alpha = 0.67$.

The death rate δ is used to match the average age of the population under study, which is 34. Because agents in the model can move between working and entrepreneurship frequently over the course of their lives, we match the age of the agent rather than the age of the firm. Moreover, we interpret a new agent in the model to be an eighteen year old in the data, so an average age of 34 in the data corresponds to 16 in the model. Because the rate of death is constant in the model, the age distribution is geometrically distributed with a mean equal to the reciprocal of δ. Moreover, a period in the model is interpreted as a quarter in the data. Therefore, to match an age of 16 years (or 64 quarters), we set $\delta = 0.016$.

Our remaining parameters are σ, σ_0, c and τ. The parameter σ_0 is matched to the variance of log-profit among new firms open less than one year (0.961). The remaining three parameters match three moments jointly: the standard deviation of log-profit in the overall population of operating firms (1.400), the ratio of the average profit of firms overall to the average profit of new entrants (1.558), and the fraction of agents that operate as workers (28.7 percent).[23] These moments and parameter values are reported in Table 3. The model matches these moments well.

[23]While jointly calibrated, each moment has a clear intuitive counterpart. σ matches the standard deviation of log profit. c is the exogenous growth in productivity, thus governs the relative average profit ratio. Finally, τ matches the share of workers in the economy by lowering the value of working relative to business operation.

Table 3: Targets and Parameter Choices

Model Parameter	Description	Parameter Value	Target Moment	Source	Target Value	Model Value
Group 1						
β	Intensity of diffusion	0.538	Estimated parameter from regession (2.3)	RCT results	0.538	0.538
ρ	Persistence of productivity	0.595	Estimated parameter from regession (2.3)	RCT results	0.595	0.595
θ	Directedness of search	-0.417	Treatment effect in quarter 2 (as % above control)	RCT results	0.403	0.403
Group 2						
σ	St. dev. of exogenous productivity shock distribution	0.877	Variance of log profit in all firms	Baseline survey	1.400	1.440
c	Growth factor in productivity evolution	-3.107	Ratio of average profit of all firms to new entrants	Baseline survey	1.558	1.559
τ	Tax on wage earnings	0.999	Fraction of agents employed as workers	Gollin (2008)	0.287	0.287
Group 3						
δ	Death rate of firms	0.016	Average age of baseline business owners	Baseline survey	34	34
σ_0	St. dev. of new entrant productivity distribution	0.961	Variance of log profit among new entrants	Baseline survey	0.961	0.961
α	Cobb-Douglas exponent on labor	0.67	Standard value	–	–	–

Table notes: Group 1 is jointly chosen from the experimental data. Parameters in Group 2 are calibrated to jointly match moments. Group 3 are also set to match baseline data moments, but match 1-1 with target moments.

5 The Quantitative Impact of Diffusion

We next turn to the quantitative results. First, we ask the extent to which the model predicts that diffusion matters at scale and study how our estimated parameters drive that result. Second, we compare general equilibrium policy gains and our partial equilibrium RCT results. Informed by our first set of results, we show that while many sets of parameters can generate the same impulse response to the treatment they generate drastically different gains from equilibrium policy, suggesting that understanding the underlying structural parameters is critical for policy.

5.1 Impact of Diffusion and Gains from Policy

The equilibrium in this model is inefficient due to the externality arising from the fact that individuals do not internalize the impact of their occupational choice on learning. Marginal firms decrease the likelihood of an individual learning from the right tail of the knowledge distribution. The planner thus wishes to assign marginal firm owners to be workers to increase average firm quality. We measure the size of this equilibrium externality by solving for the efficient allocation of agents between workers and entrepreneurs and compare the stationary equilibrium when agents choose optimally and without intervention. We refer to this as the *laissez faire* equilibrium.

The planner maximizes total production in the stationary equilibrium, subject to consistency with the law of motion for productivity,

$$\max_{\underline{z}} \quad \int_{\underline{z}}^{\infty} y(z) dM^*(z) \tag{5.1}$$

$$s.t. \quad M^*(z') = \delta G(z') + (1-\delta) \int_0^{\infty} \int_0^{\infty} F(\log(z') - \log(\max\{\hat{z}^{\beta} z^{\rho-\beta}, z^{\rho}\}) - c) d\widehat{M}(\hat{z};\theta) dM^*(z)$$

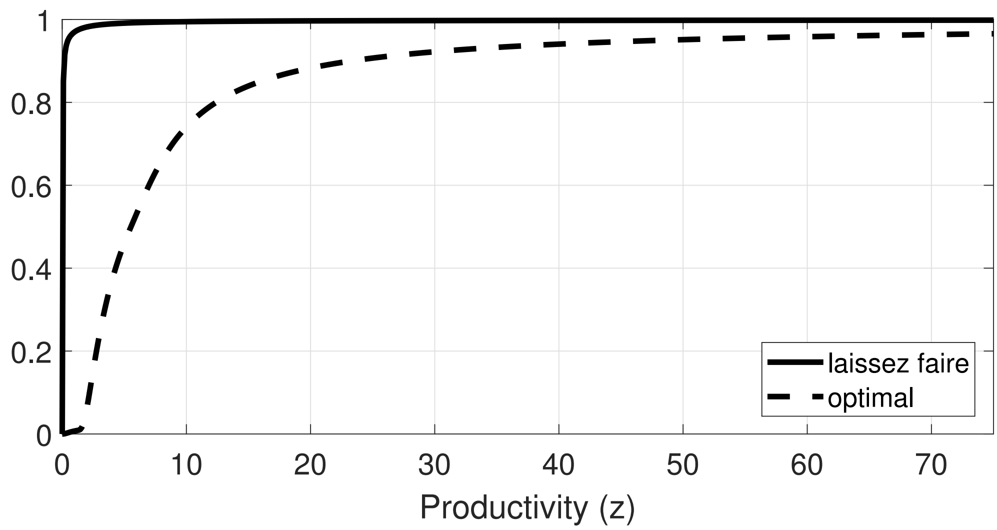
$$\widehat{M}(\hat{z};\theta) = \begin{cases} 0, & \text{if } \hat{z} < \underline{z} \\ \left(\frac{M^*(\hat{z}) - M^*(\underline{z})}{1-M^*(\underline{z})}\right)^{\frac{1}{1-\theta}}, & \text{if } \hat{z} \geq \underline{z}. \end{cases}$$

The solution to (5.1) balances more production from setting a lower productivity cutoff $\underline{z}$ for a given distribution M^* with the fact that lower $\underline{z}$ also reduces the average quality of imitation opportunities.[24] Compared to the *laissez faire* equilibrium, the planner internalizes this second effect and chooses a greater value of $\underline{z}$. Figure 5 plots the c.d.f. of equilibrium productivity across the two economies and confirms this intuition.

Table 4 then compares key moments between the *laissez faire* and efficient economies.

[24] Note that we model θ here as a technological parameter, meaning the planner cannot change it. In reality, this parameter may be made of up distortions in the matching process (e.g. Beaman and Dillion, 2018). This does not affect the interpretation of our identification results, as the procedure can applied identically to both interpretations of θ. However, it may be interesting to interpret θ as a policy parameter and consider policies that may change its value. Of course, the model implies that average income is higher when agents can more easily access the best imitation draws. However, in the Appendix we show that the quantitative gains from this additional margin are orders of magnitude smaller than the gains from policy conditional on θ.

Figure 5: Stationary Productivity CDF



We find optimal policy has a large effect. At our estimated parameters, the efficient economy has an average income 309 percent higher than the *laissez faire* equilibrium. The efficient equilibrium has only 1 percent of the population engaged in entrepreneurship compared to 71 percent in the baseline economy. Correspondingly, average entrepreneurial productivity increases by over 800 percent. The overall gains in the economy then come from two places. The first is the 48 percent increase in the wage. The second is that – through the equilibrium learning externality – the gains to the most productive entrepreneurs increase substantially.

Table 4: Equilibrium Moments

	Laissez faire	Efficient	Δ Outcome
Average Income	0.14	0.59	3.45
Fraction working	0.31	0.99	2.17
Average entrepreneurial productivity	1.72	17.13	8.97
Wage	1.37	2.02	0.48

Table notes: Equilibrium changes are measured as the ratio of efficient to laissez faire minus one. Multiply by 100 for percentage change.

These effects are large because of a feedback effect in diffusion. Improving the set of imitation draws causes the distribution of firm productivities to improve, which makes imitation draws still better, which causes the distribution of firm productivities to improve further, and so on.[25]

[25] Indeed, many of the papers using similar diffusion models achieve balanced growth through this mechanism. In our context, the fact that $\rho < 1$ keeps the economy stationary.

5.2 Importance of Average and Heterogeneous Treatment Effects

What parameters drive the results in Table 2, and how do they relate to our regression estimates from which we derive them? In this section we study this question, and relate our empirical moments to the aggregate consequences of diffusion.

To begin, we vary β and study the response under the re-calibrated model. The re-calibration does the following: we fix β, re-estimate the remaining diffusion parameters ρ and θ, then re-calibrate the three additional parameters σ, c, and τ to hit the same three moments discussed above. The results are in Table 5. We find that the lower value of β is associated with less than half of the aggregate gains from optimal policy compared to the baseline case, showing that β plays a strong role in determining aggregate consequences of diffusion.

Next we conduct a comparative static exercise where we change each diffusion parameter individually to see how aggregates change. As in the fully recalibrated case, we see that reducing the value of β alone causes the gains to fall by a similar magnitude. However, changes in either ρ or θ have comparatively small effects on aggregate gains from optimal policy. From this we conclude that the parameter β, which is determined by heterogeneity in treatment effects, plays the strongest role in determining gains from policy.

Table 5: Variation in Equilibrium Response with Diffusion Parameters

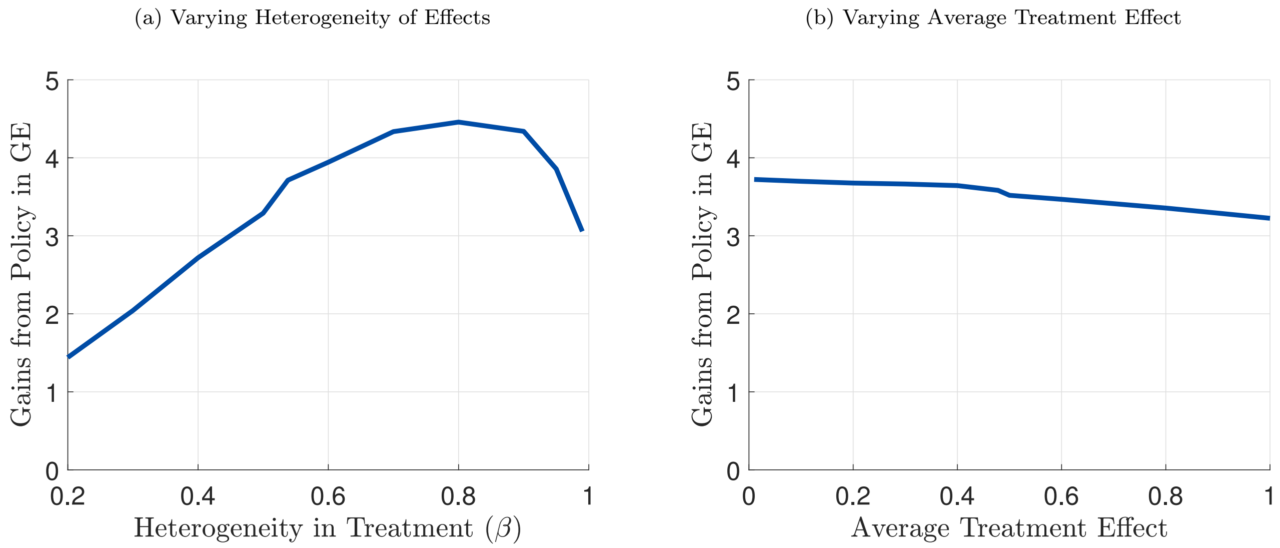
Assumed value	Implied estimates		Δ Equilibrium Outcomes			
β	ρ	θ	Average income	Fraction working	Avg. Entrepreneur z	Wage
Baseline:						
0.538	0.595	-0.417	3.45	2.17	8.97	0.48
Full recalibration with low β:						
0.25	0.314	-2.578	1.56	2.24	1.33	0.02
Varying diffusion parameters individually:						
0.25	0.595	-0.417	1.84	1.69	2.98	0.16
0.538	*0.314*	-0.417	3.94	5.76	3.97	0.01
0.538	0.595	*−2.578*	3.19	1.81	9.20	0.55

Table notes: Equilibrium changes are measured as the ratio of efficient to laissez faire minus one. Multiply by 100 for percentage change. In the last three rows, we vary individual diffusion parameters holding fixed the remaining calibration at its baseline values. The italicized parameter in the last three rows is the parameter that is varied.

Overall, these results show that the quantitative magnitude of the diffusion externality is governed primarily by the elasticity β parameter that translates matches into innovations in productivity. The other diffusion parameters play a smaller role. One interesting implication of these results is that it suggests that when attempting to identify economies with the largest gains from policy, observing substantial heterogeneity in the treatment effect is more important than observing a large average treatment effect. This follows from the fact that the former identifies β while the latter identifies θ.

We now study this directly with the following exercise. We counterfactually vary the empirical moments we observe and ask how different values of the average treatment effect and heterogeneous treatment effect (which imply different values of the diffusion parameters) translate to different gains from optimal policy. The results of this are given in Figure 6. Figure 6a shows the results from varying β (heterogeneity) while holding fixed the average treatment effect. This varies both β directly and θ to hold fixed the average treatment effect.[26] Figure 6b then varies the (counterfactually assumed) observed average treatment effect from 0 to 100 percent while keeping the heterogeneity fixed. This varies θ while holding others parameters fixed, as for example, β is independent of the average treatment effect.

Figure 6: Impact of Empirical Moments on Size of Diffusion Externality



The vertical axes in this figure have the same scale for easy comparison of the magnitudes of changes as empirical moments vary, ranging from 0 to 500 percent gains from optimal policy. As expected given the results in Table 5, the gains vary substantially with the value of β, governing the heterogeneity in treatment effects generated by heterogeneity in match quality.[27]

On the other hand, given a level of observed heterogeneity, the gains from policy are actually declining in the average treatment effect. The intuition for this result follows from the divergent roles played by θ, which as discussed through, controls the average treatment effect. In terms of the RCT results, the average effect is maximized when

[26]Recall that θ can only be identified on an interval of moment values conditional on β. Given our baseline calibration, this occurs at $\beta = 0.154$, so we focus on $\beta \geq 0.2$ here.

[27]This non-monotonicity seen at high levels of β comes from the fact that efficient income is roughly linear in β while the *lassez faire* income level is convex at high levels of β. This generates the shape in Figure 6a. See the Appendix for more details and results on the mechanics of the model.

control firms are unable to find good matches. That is, it is largest when θ is low (see the Appendix for these results). However, this is exactly the same force that *limits* the ability of the planner to direct matches toward good firms with her available policy levers: agents learn only from the worst remaining firms, increasing the cost for the planner to generate an average match of a given (high) quality. The quantitative magnitudes, however, are small.

This exercise demonstrates that average treatment effects are not sufficient to understand the magnitude of potential gains in general equilibrium. If the empirical result on Table 2 had shown a small average treatment effect, one might naively conclude that diffusion is unimportant and that introducing policy to address diffusion externalities would have limited effects. Instead, it is the strong heterogeneity in effects across firms with different baseline profit gaps that indicates the potential for substantial gains. Conversely, if there had been large average treatment effects with small heterogeneous effects, reliance on the average treatment effect to infer the importance of diffusion would lead to overestimates of the equilibrium role of diffusion.

5.3 Persistence of Effects from Treatment

Up until this point, we have spent little time discussing the dynamics of these effects. One could be concerned, for example, that our results come at the expense of matching the dynamic path of the treatment effect. Here, we show that such results only reinforce our results above.

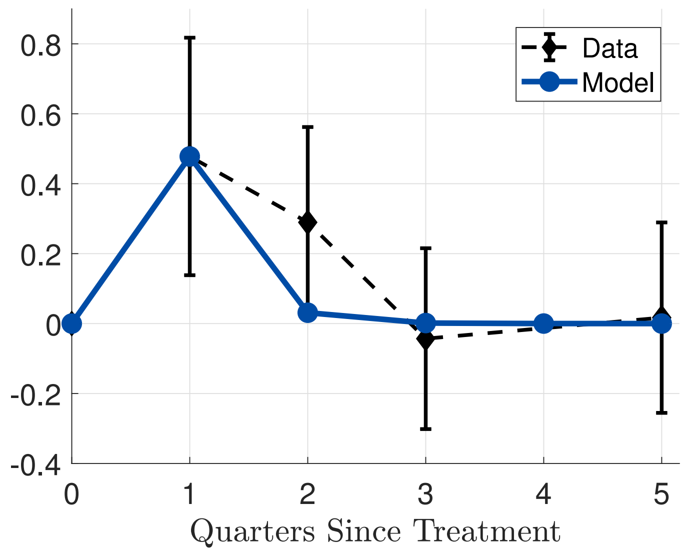
We begin by implementing the RCT in the model, tracing out the impulse response and comparing it to our empirics. To do so in the model, we start the economy from the stationary equilibrium. We then create the control and treatment group, along with a group of "mentors," and shock the treatment with a one-period draw from this high-profit group. We then trace out the dynamics of the response to this one-time shock. These results are partial equilibrium in the sense that we hold fixed the aggregate state M^* throughout.

These results are in Figure 7. While the first two quarters (i.e., $t = 0$ and $t = 1$) are matched by construction, both the model and data predict no treatment effect by $t = 3$. The model under-predicts the effect in $t = 2$, so if anything, the model understates these partial equilibrium RCT gains.

As discussed in Section 5.2, average effects can provide a misleading picture of the gains from optimal policy. Figure 7 re-enforces this idea by showing that gains from policy similarly cannot be measured by another commonly utilized measure of RCT success – persistent gains from the treatment. Here, both in the empirics and model, we find a quick fade-out of the average effect.

To study the relationship between this fadeout of the average effect and the critical heterogeneity parameter β, we trace the impulse response for various levels of β given

Figure 7: Dynamics of Treatment Effect on Profit



a fixed initial effect. That is, analogous to our counterfactuals in Figure 6, we vary β (heterogeneity) while holding fixed the average treatment effect, and now trace out the impulse response. Those results are in Figure 8. Figure 8a shows the results at our baseline value of ρ, while Figure 8b increases persistence to show how the results change.

Figure 8: Relationship between Treatment Persistence and Diffusion Intensity β

(a) At Estimated Persistence $\rho = 0.595$

(b) At High Persistence $\rho = 0.99$

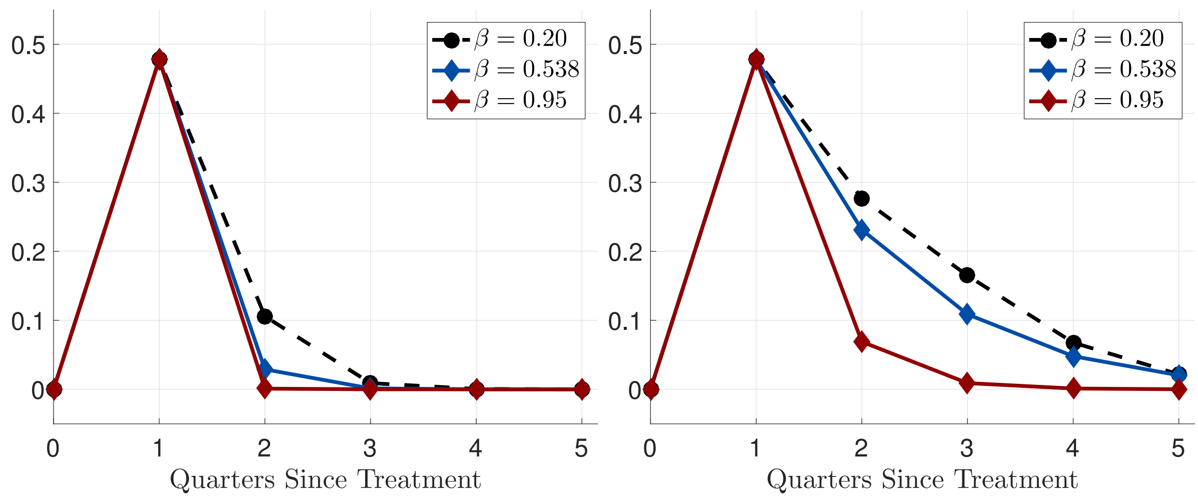


Figure 8 shows that not only does β drive the equilibrium gains from policy, it simultaneously hastens the fade-out of the average treatment effect over time. In essence, high β makes the planner's job easier by allowing agents to internalize more productivity from a given match. That agents internalize more with higher β is, of course, also true in the partial equilibrium RCT results. However, a critical feature of measuring treatment effects is to remember that the control continues to engage in matches during this time.

A higher β therefore allows control firms to quickly catch up to the treatment group – not because of contamination from treatment to control, but because they are able to internalize a large portion of a good match's productivity.

5.3.1 Does the Time Series Provide Any Useful Information Without the Model?

Our results so far point to the care that must be taken when interpreting micro level evidence from average treatment effects – either as a single value or time series – in models of diffusion. Instead, what is critical is the ability to translate those into structural parameters that govern gains from policy. To finalize this point, we show that we can *always* construct an economy that matches the empirical time series of the average treatment effect for any level of heterogeneity. To do so, we assume that the time series of average treatment effects is known. We show that a continuum of underlying diffusion processes associated with different values of β could have generated that time series, associated with a wide range of gains from at-scale policy.

Our procedure works as follows. We first fix β *ex ante*. For this given β, we search for the (ρ, θ) that minimize the sum of squared errors between the implied treatment effect and the baseline. Specifically, we solve

$$\min_{\rho,\theta} \sum_{t=1}^{t=6} \left(ATE_t(\rho, \theta; \beta) - ATE_t^{base}\right)^2 \tag{5.2}$$

where $ATE_t(\rho, \theta; \beta)$ is the implied average treatment effect at quarter t given a value of β, and ATE_t^{base} is the same moment from our baseline estimated model (the solid line in Figure 7).

We vary $\beta \in [0.2, 0.8]$, and find that the maximum error in (5.2) is 10^{-5}. Thus, the average treatment effects can be easily generated period-by-period across a wide set of economies with different underlying parameters. Figure 9 shows the implied values for ρ and θ required to match the time series (Figure 9a), while Figure 9b shows that gains from optimal policy in general equilibrium still vary widely. The gains more than double across the range of β. This shows that averages alone are insufficient to understand the equilibrium gains from at-scale policy.

6 Conclusion

We show how well-identified variation in the data generating process can identify a model of firm-to-firm productivity transmission. We implement this procedure with a randomized controlled trial in Kenya. Our results imply an important role for diffusion. The efficient level of the learning externality increases income substantially and depend critically on proper estimation of the diffusion parameters. The heterogeneity across firms in treatment effects is the crucial empirical moment for determining aggregate gains from

Figure 9: Gains from Equilibrium Policy in Models with Identical ATE Time Series

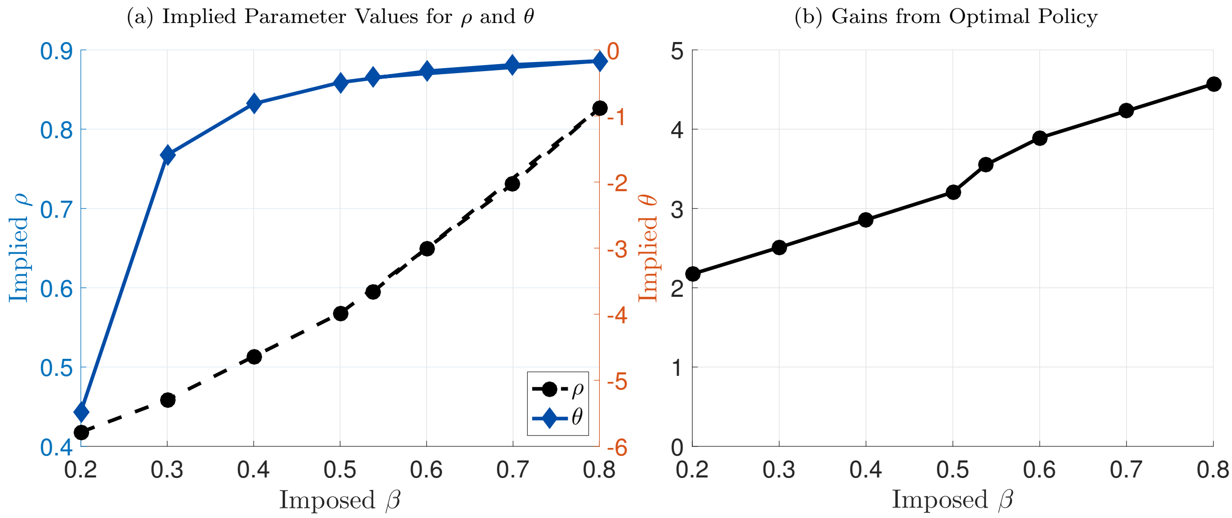


Figure notes: Each line varies the listed parameter holding all other model parameters fixed at baseline values. The points indicated by circles and diamonds are our baseline estimates.

at-scale policy while the average treatment effect plays little role. Likewise, pooled average treatment effects or even the whole time series of treatment effects are insufficient for computing aggregate gains.

We view these results as an important first step that highlights the possibilities of linking equilibrium diffusion models with causal identification. Our results show that generating required empirical moments with well-designed experiments in a "top down" approach can provide important information for equilibrium models. There are two broad implications for future work. First, we note that while our implementation took the idea of random matching quite literally, any shock to imitation opportunities that generates the correct orthogonality conditions can be similarly utilized. This opens up alternative ways to implement such a strategy, including the combination of natural experiments with necessary data (e.g. Giorcelli, 2019; Bianchi and Giorcelli, 2019).

Second, the results can be used to identify more complicated models as well. This requires a more detailed investigation of the link between model and data. For example, one question that remains unanswered both in this paper and the broader literature is why individuals do not seek out the most productive business owners to learn from, given the seemingly large benefits observed at the individual level (though, as we show, such frictions need not play a large role in equilibrium). Our model builds this in as a technological constraint, but that need not be the case. Beaman and Dillion (2018) point to frictions in the information market, while Fogli and Veldkamp (forthcoming) point out that growth-reducing network structures can be an optimal response to the possibility of detrimental flows through the network (e.g., disease). These are important questions that may, for example, help rationalize low life-cycle earnings growth in poor countries (Lagakos et al., 2018b). Different field experiments, designed with an eye toward aggregate theory, could provide more detailed information to help further refine such model choices.

References

Akcigit, Ufuk, Harun Alp, and Michael Peters, "Lack of Selection and Limits to Delegation: Firm Dynamics in Developing Countries," *American Economic Review*, forthcoming.

Alvarez, Fernando E., Francisco J. Buera, and Robert E. Lucas, "Models of Idea Flows," June 2008. NBER Working Paper 14135.

Atkin, David, Amit K. Khandelwal, and Adam Osman, "Exporting and Firm Performance: Evidence from a Randomized Experiment," *Quarterly Journal of Economics*, 2017, *132* (2), 551–615.

Banerjee, Abhijit, Arun G. Chandrasekhar, Esther Duflo, and Matthew O. Jackson, "The Diffusion of Microfinance," *Science*, 2013, *341*, 363–370.

Beaman, Lori and Andrew Dillion, "Diffusion of Agricultural Information within Social Networks: Evidence on Gender Inequalities from Mali," *Journal of Development Economics*, 2018, *133* (1), 147–161.

_ , Ariel BenYishay, Jeremy Magruder, and A. Mushfiq Mobarak, "Can Network Theory-based Targeting Increase Technology Adoption?," February 2020. Northwestern University Working Paper.

Benhabib, Jess, Jesse Perla, and Christopher Tonetti, "Reconciling Models of Diffusion and Innovation: A Theory of the Productivity Distribution and Technology Frontier," 2019. Working Paper.

BenYishay, Ariel and A. Mushfiq Mobarak, "Social Learning and Incentives for Experimentation and Communication," *Review of Economic Studies*, 2018, *86* (3), 976–1009.

Bianchi, Nicola and Michela Giorcelli, "Not All Management Training is Created Equal: Evidence from the Training Within Industry Program," 2019. Working Paper.

Brooks, Wyatt and Kevin Donovan, "Eliminating Uncertainty in Market Access: The Impact of New Bridges in Rural Nicaragua," *Econometrica*, 2020, *88* (5), 1965–1997.

_ , _ , and Terence R. Johnson, "Mentors or Teachers? Microenterprise Training in Kenya," *American Economic Journal: Applied Economics*, 2018, *10* (4), 196–221.

Buera, Francisco J. and Ezra Oberfield, "The Global Diffusion of Ideas," *Econometrica*, 2020, *88* (1), 83–114.

_ **and Yongseok Shin**, "Self-Insurance vs. Self-Financing: A Welfare Analysis of the Persistence of Shocks," *Journal of Economic Theory*, 2011, *146* (1), 845–862.

_ **, Joseph P. Kaboski, and Yongseok Shin**, "The Macroeconomics of Microfinance," 2017. Working Paper.

Cai, Jing and Adam Szeidl, "Interfirm Relationships and Business Performance," *Quarterly Journal of Economics*, 2018, *133* (3), 1229–1282.

Conley, Timothy G. and Christopher R. Udry, "Learning about a New Technology: Pineapple in Ghana," *American Economic Review*, 2010, *100* (1), 35–69.

Da Rocha, Jose Maria and Pau Pujolas, "Policy Distortions and Aggregate Productivity: the Role of Idiosyncratic Shocks," *B.E. Journal of Macroeconomicss*, 2011, *11* (1), 1–21.

Fogli, Alessandra and Laura Veldkamp, "Germs, Social Networks and Growth," *Review of Economic Studies*, forthcoming.

Foster, Andrew D. and Mark R. Rosenzweig, "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture," *Journal of Political Economy*, 1995, *103* (6), 1176–1209.

Fujimoto, Junichi, David Lagakos, and Mitch Vanvuren, "The Aggregate Effects of "Free" Secondary Schooling in the Developing World," 2019. Working Paper.

Giorcelli, Michela, "The Long-Term Effects of Management and Technology Transfers," *American Economic Review*, 2019, *109* (1), 121–152.

Gollin, Doug, "Nobody's Business But My Own: Self Employment and Small Enterprise in Economic Development," *Journal of Monetary Economics*, 2008, *55* (2), 219–233.

Greenwood, Jeremy, Philipp Kircher, Cezar Santos, and Michèle Tertilt, "An Equilibrium Model of the African HIV/AID Epidemic," *Econometrica*, 2019, *87* (4), 1081–1113.

Herkenhoff, Kyle, Jeremy Lise, Guido Menzio, and Gordon Phillips, "Knowledge Diffusion in the Workplace," July 2018. Working Paper, University of Minnesota.

Hopenhayn, Hugo and Liyan Shi, "Knowledge Creation and Diffusion with Limited Appropriation," September 2018. Working Paper.

Hsieh, Chang-Tai and Peter J. Klenow, "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 2009, *124* (4), 1403–1448.

Imbens, Guido and Karthik Kalyanaraman, "Optimal Bandwidth choice for the Regression Discontinuity Estimator," *Review of Economic Studies*, 2012, *79* (3), 933–959.

Jarosch, Gregor, Ezra Oberfield, and Esteban Rossi-Hansberg, "Learning from Coworkers," 2020. Working Paper.

Jovanovic, Boyan and Rafael Rob, "The Growth and Diffusion of Knowledge," *Review of Economic Studies*, 1989, *56* (4), 569–582.

Kaboski, Joseph P. and Robert M. Townsend, "A Structural Evaluation of a Large-Scale Quasi-Experimental Microfinance Initiative," *Econometrica*, 2011, *79* (5), 1357–1406.

Kortum, Samuel S., "Research, Patenting, and Technological Change," *Econometrica*, 1997, *65* (6), 1389–1419.

Lagakos, David, Ahmed Mushfiq Mobarak, and Michael E. Waugh, "The Welfare Effects of Encouraging Rural-Urban Migration," January 2018. NBER Working Paper 24193.

_ , Benjamin Moll, Tommaso Porzio, Nancy Qian, and Todd Schoellman, "Life Cycle Wage Growth across Countries," *Journal of Political Economy*, 2018, *126* (2), 797–849.

Lucas, Robert E., "Ideas and Growth," *Economica*, 2009, *76* (301), 1–19.

_ and Benjamin Moll, "Knowledge Growth and the Allocation of Time," *Journal of Political Economy*, 2014, *122* (1), 1–51.

Luttmer, Erzo G.J., "Selection, Growth, and the Size Distribution of Firms," *Quarterly Journal of Economics*, 2007, *122* (3), 1103–1144.

Moll, Benjamin, "Productivity Losses from Financial Frictions: Can Self-financing Undo Capital Misallocation?," *American Economic Review*, 2014, *104* (10), 3186–3221.

Nakamura, Emi and Jón Steinsson, "Identification in Macroeconomics," *Journal of Economic Perspectives*, 2018, *32* (3), 59–86.

Perla, Jesse and Christopher Tonetti, "Equilibrium Imitation and Growth," *Journal of Political Economy*, 2014, *122* (1), 52–76.

_ , _ , and Michael E. Waugh, "Equilibrium Technology Diffusion, Trade, and Growth," October 2020. Working Paper.

Romer, Paul M., "Increasing Returns and Long-Run Growth," *Journal of Political Economy*, 1986, *94* (5), 1002–1037.

Todd, Petra E. and Kenneth I. Wolpin, "Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility," *American Economic Review*, 2006, *96* (5), 1384–1417.

Table of Contents for Online Appendix

A Identification with Different Diffusion Processes

In this section, we provide additional interpretations of models that fall under our assumptions.

A.1 Multiple Draws

Suppose each period each agent takes K independent, uniform draws from the distribution M, labeled $\hat{z_1}, ..., \hat{z_K}$. The agent then has to select the most useful of these draws. Hence:

$$\hat{z} = \max\{\hat{z_1}, ...\hat{z_N}\} \tag{A.1}$$

The distribution of $\hat{z}$ then follows the well-known form:

$$\widehat{M}(c) = Prob(\hat{z} \leq c) = Prob(\max\{\hat{z_1}, ...\hat{z_N}\} \leq c) = \prod_{i=1}^{K} Prob(\hat{z_i} \leq c) = \prod_{i=1}^{K} M(c) = (M(c))^K \tag{A.2}$$

where the third inequality comes from the fact that they are independent and the fourth from the fact that each draw is from M.

Note that this example is a special case of the version considered in the body of the paper when $1/(1-\theta)$ is a natural number.

A.2 Effort Choice and Bargaining

Each period, every agent characterized by productivity z is matched to an agent that owns a potential imitation opportunity z_m as a uniform draw from the distribution of operating firms M. The agent has an effort endowment of 1 that must be divided between imitation and providing a utility benefit to the owner of the imitation opportunity z_m. If $z \geq z_m$, then no effort is put into imitation and $\hat{z} = z$. If $z_m > z$, then the agent and the owner of the imitation opportunity must first agree on the distribution of effort, then the choice of effort x and the values of z and z_m together generate the value of $\hat{z}$ for the agent in that period according to:

$$\hat{z} = \left(\frac{z_m}{z}\right)^x z \tag{A.3}$$

That is, by putting in more effort $x \in [0, 1]$ the agent is able to close the gap between their z and z_m. The benefit to the owner of z_m is given by the function $b(x)$, which is decreasing in x.

Agents and owners of imitation opportunities have one-off interactions and each receive 0 benefit if no agreement is made. They bargain over the assignment of the agent's effort between imitation and utility benefits for the owner of the imitation opportunity according to a Nash bargaining problem where the bargaining weight of the agent is θ. The bargaining problem is:

$$\max_{x\in[0,1]} \left(\left[\frac{z_m}{z}\right]^x z\right)^\theta b(x)^{1-\theta} \tag{A.4}$$

Suppose that $b(x)$ is given by $b(x) = 1 - x$. Then it is easy to show that:

$$x = \max\left[0, 1 - \frac{1-\theta}{\theta \log(z_m/z)}\right] \tag{A.5}$$

$$\hat{z} = \max\left[z, z_m e^{1-1/\theta}\right] \tag{A.6}$$

As expected, the more bargaining power that the learning agents have, the greater is x, resulting in greater $\hat{z}$.

Note that, in the model, draws of imitation opportunities $\hat{z} < z$ are not useful. Hence, the distribution $\widehat{M}$ can be written, for any value c, as:

$$\widehat{M}(c) = Prob(\hat{z} \leq c) = Prob(z_m e^{1-1/\theta} \leq c) = Prob(z_m \leq c e^{1/\theta-1}) = M(c e^{1/\theta-1}) \tag{A.7}$$

or following the notation more standard in the paper:

$$\forall z, \widehat{M}(\hat{z}, z, \theta) = M(\hat{z} e^{1/\theta-1}) \tag{A.8}$$

A.3 Noise in the Imitation Process

Here we show how the Buera and Oberfield (2020) environment maps into that considered in this paper. In their model (adapted to our notation), an agent with productivity z receives new arrivals of ideas that have two components: z_m that comes from a random match from another agent, and γ a random innovation on that idea. Then $\hat{z} = \gamma^{1/\theta} z_m$. Here, z_m is a uniform draw from the distribution of productivities. Then if γ has a cumulative density function given by Γ, then:

$$\widehat{M}(c) = Prob(\hat{z} \leq c) = Prob(z_m \leq c\gamma^{-1/\theta}) = \int M(c\gamma^{-1/\theta}) d\Gamma(\gamma) \tag{A.9}$$

A.4 Deterministic Assignment

Here we consider a case where $\widehat{M}$ arises when all agents can interact with one another and sort into relationships endogenously. Suppose that every agent with productivity $\hat{z}$ has the option to influence any other agent that has productivity z. Every agent can only be influenced by one other agent each period, and they always prefer to be influenced by the highest productivity possible.

The utility of an agent with productivity $\hat{z}$ influencing an agent with productivity z

is given by:

$$\frac{\hat{z}}{z} - 1 - \frac{1}{2\theta}\left(\frac{\hat{z}}{z} - 1\right)^2 \tag{A.10}$$

That is, the agent with $\hat{z}$ gains benefit in proportion to how large the benefit is for the other agent, but their cost is quadratic in the distance between their productivities. For example, the influencer is happy when the other agent is helped by their influence, but it takes more effort to influence when the distance between them is great. Therefore, if there is a continuous distribution of $z < \hat{z}$, the ideal agent that the influencer would like to interact with has productivity:

$$z^*(\hat{z}) = \hat{z}/(1+\theta) \tag{A.11}$$

That is, the lower is the cost of influencing low productivity firms, the deeper into the left tail of the distribution is the agent willing to go.

However, since every agent can only be influenced by one agent each period and they strictly prefer to be influenced by agents of higher productivity, it is possible that (even if the distribution is continuous) that the ideal agent for $\hat{z}$ is already matched to another influencer. Therefore, intuitively, the probability distribution over assignment between $\hat{z}$ and z is constructed by starting at the upper support of the distribution M, allowing the highest productivity firms to choose their most preferred matches, then descending down through the distribution letting each firm choose to influence its preferred firm among those remaining. Note that not all firms need have another firm to influence if their utility from doing so be negative.

Formally, the probability distribution over imitation opportunities can be constructed in the discretized case as follows, when the productivity grid takes values $z \in \{z_1, ...z_N\}$, which are ordered ($i < j \implies z_i < z_j$).

Define $\tilde{\mu}(z, \hat{z})$ as the measure of $\hat{z}$ influencing z (a $N \times N$ matrix). We can construct $\tilde{\mu}$ in the following steps given the measure μ of agents of each z type:

1. Let $U(z, \hat{z})$ be the $N \times N$ matrix of utilities of $\hat{z}$ influencing z, and $\tilde{\mu}$ be a $N \times N$ matrix of zeros. Let $\bar{\mu}$ be the $N \times 1$ vector of unassigned influencers and μ_u be the $N \times 1$ vector of unassigned imitators. Set $\bar{\mu} = \mu_u = \mu$, $n = N$, and $m = 1$.

2. Let l be the m-argmax of $U(\cdot, z_n)$. If $U(z_l, z_n) \leq 0$, set $\tilde{\mu}(z_1, z_n) = \mu_u(z_n)$ and skip to step 5.

3. If $\bar{\mu}(z_n) \leq \mu_u(z_l)$, then $\bar{\mu}(z_n) = 0$, $\mu_u(z_l) = \mu_u(z_l) - \bar{\mu}(z_n)$, and $\tilde{\mu}(z_l, z_n) = \bar{\mu}(z_n)$. Skip to step 5. Otherwise, go to 4.

4. If $\bar{\mu}(z_n) > \mu_u(z_l)$, then set $\tilde{\mu}(z_l, z_n) = \mu_u(z_l)$, $\mu_u(z_n) = 0$ and $\bar{\mu}(z_n) = \bar{\mu}(z_n) - \mu_u(z_l)$. Set $m = m + 1$ and return to step 2.

5. Set $n = n - 1$ and $m = 1$. If $n = 0$, go to step 6. Otherwise, go to step 2.

6. Set $\tilde{\mu}(\cdot, z_1) = \tilde{\mu}(\cdot, z_1) + \mu_u$, and stop.

Given this matrix $\tilde{\mu}(z, \hat{z})$, the measure of assignments $\widehat{M}$ is given by:

$$\widehat{M}(\hat{z}_i, z_j) = \frac{\sum_{k=1}^{i} \tilde{\mu}(z_j, \hat{z}_k)}{\mu(z_j)} \tag{A.12}$$

A.5 Congestion

As discussed in the main text, θ may sometimes fail to be independent of the remaining model structure. One such example of that is in a model with congestion, in which firms decide whether to be "teachers" or "students."

Let $\mathbf{X} = (x, X)$, where x is the individual state and X the aggregate state of the economy, and $o(\mathbf{X}; \theta) = 1$ be the decision rule to become a student. Individuals could choose to become teachers or students based on any number of reasons, including some warm-glow preferences or transfers made for their services, but that rationale is irrelevant here. Let $s(\mathbf{X}; \theta) = \int o(\mathbf{X}; \theta) d\mathbf{X}$ be the measure of students.

The matching function is $s^{\theta}(1-s)^{1-\theta}$. Conditional on drawing a match, the exact match is a uniform random draw from the set of teachers. Then, for any value c (and dropping the dependence on $\mathbf{X}$ for notational simplicity),

$$\begin{aligned} \widehat{M}(c; \theta) &= Prob(\hat{z} \leq c) \\ &= s(\theta)^{\theta}(1 - s(\theta))^{1-\theta} Prob(z_m \leq c) \\ &= s(\theta)^{\theta}(1 - s(\theta))^{1-\theta} M^t(c; \theta) \end{aligned}$$

where M^t is the c.d.f. of teacher productivity.

From here, there are two possibilities. The first is if we can observe who is a student and who is a teacher. In this case, there no issue and the identification of θ goes through as the main text. The second is if we cannot identify student/teacher type. In this context, as long as the last line satisfies the FOSD assumption in θ, there will be a unique mapping between model parameters and the value of θ required to match the average treatment effect. Thus, it still provides a valuable moment in estimation. However, the value of θ will generally not be independent of the remaining model structure in this context, as that structure is required to back out the distribution M^t from some overall distribution M that includes the productivity of both students and teachers.

Note that we have not mentioned β or ρ here, since those results go through identical to the main text.

B Identification without More Productive Treatment Draws

In the main body of the paper, we assumed that for all treatment firms i, their matches are more productive. That is, $\hat{z}_i > z_i$ for all i in the treatment. This assumption is not necessary for the main identification results, and we relax it here. The key difference is that β and ρ must now be jointly identified, requiring more work on the existence and uniqueness of a fixed point. Proposition 3 shows the result is not required. Below we detail the procedure.

The transmission parameter β is identified by comparing the effects on two initially identical participants from receiving a very high productivity $\hat{z}$ match to those receiving a relatively low $\hat{z}$ match. If those receiving a high $\hat{z}$ realize much bigger returns compared to their receiving a lower $\hat{z}$, we conclude that β is high. The persistence term ρ is then read off the persistence of profit among the treatment firms.

To formalize this idea, we first compare treatment that received a "high" productivity $\hat{z}$ draw to those receiving a "low" $\hat{z}$ draw.[28] Letting $\Omega(z, \hat{z})$ be the set of all realized treatment matches, we can define disjoint subsets Ω_H and Ω_L with associated probability density functions $m_H(z, \hat{z})$ and $m_L(z, \hat{z})$ such that:

$$\forall \hat{z}_0, \int_0^{\hat{z}_0} \int_0^{\infty} m_H(z, \hat{z}) dz d\hat{z} < \int_0^{\hat{z}_0} \int_0^{\infty} m_L(z, \hat{z}) dz d\hat{z}. \tag{B.1}$$

That is, the $\hat{z}$ draws within Ω_H are "better" than those within Ω_L.

Now we define the first moment condition using these subsets. Defining the average profit after treatment as $\mathbb{E}[\pi_H^T]$ and $\mathbb{E}[\pi_L^T]$ for members of Ω_H and Ω_L, our first empirical moment is

$$\Gamma_1 \equiv \frac{\mathbb{E}[\pi_H^T]}{\mathbb{E}[\pi_L^T]} = \frac{\int\int\int e^{c+\varepsilon} z^{\rho} \max\left[1, \frac{\hat{z}}{z}\right]^{\beta} m_H(z, \hat{z}) dz d\hat{z} dF(\varepsilon)}{\int\int\int e^{c+\varepsilon} z^{\rho} \max\left[1, \frac{\hat{z}}{z}\right]^{\beta} m_L(z, \hat{z}) dz d\hat{z} dF(\varepsilon)}. \tag{B.2}$$

Note that Γ_1 is simply a measure of the heterogeneity in treatment effect for some measure of "high" (H) and "low" (L) quality matches. This empirical moment can be read directly off a regression given our randomization, and thus is observable. Furthermore, given the independence of the ε terms along with the fact that several constants appear in the numerator and denominator, this can be written more simply as

$$\Gamma_1 \equiv \frac{\int\int z^{\rho} \max\left[1, \frac{\hat{z}}{z}\right]^{\beta} m_H(z, \hat{z}) dz d\hat{z}}{\int\int z^{\rho} \max\left[1, \frac{\hat{z}}{z}\right]^{\beta} m_L(z, \hat{z}) dz d\hat{z}}. \tag{M1}$$

Since Γ_1, m_H and m_L come directly from the data, only β and ρ are yet unknown in this equation. Thus, (M1) pins down β as a function of ρ. We therefore need a second moment to separate them.[29]

[28] These are only relative classifications. The "low" draws are still from the upper tail of the population distribution.

[29] The reason that Γ_1 only identifies $\beta(\rho)$ instead of β directly stems from the fact that Γ_1 is a measured response to a

The second moment used to identify these parameters is the relationship between initial productivity z and final productivity z' among the set of treatment participants $i \in T$. The identification strategy is similar to that employed in standard firm dynamics models with AR(1) processes, but much be adjusted to take into account the diffusion process, which is inherently asymmetric. Specifically, the moment we use is

$$\Gamma_2 \equiv \frac{Cov[z,z']}{E[z]E[z']} + 1 = \frac{\int\int z^{1+\rho} \max\left[1, \frac{\hat{z}}{z}\right]^{\beta} m(z,\hat{z}) dz d\hat{z}}{\int z \int m(z,\hat{z}) d\hat{z} dz \cdot \int\int z^{\rho} \max\left[1, \frac{\hat{z}}{z}\right]^{\beta} m(z,\hat{z}) dz d\hat{z}}. \tag{M2}$$

A simple way to highlight the empirical availability of Γ_2 is to note that we can rewrite $\Gamma_2 = 1 + \frac{Var(z)}{\mathbb{E}[z]\mathbb{E}[z']}\widehat{\gamma}^{OLS}$, where $\widehat{\gamma}^{OLS}$ is the coefficient estimate from a lagged profit regression

$$\pi_{i,t} = \eta + \gamma \pi_{i,t-1} + \nu$$

run on all treatment individuals. Thus, this moment, like Γ_1 is easily observed in the data. This moment allows us to pin down ρ as a function of β. For some intuition on why this is the case, note that in an economy with no diffusion and exogenous shocks drawn from $F \sim N(\mu, \sigma^2)$ then this moment simplifies to $\Gamma_2 = \exp(\sigma^2 \rho)$. Thus, with knowledge of the distribution of exogenous shocks, the normalized lagged profit regression coefficient identifies persistence of productivity. This result is used in a variety of firm dynamics models that do not include diffusion, and identifies the persistence of an exogenous AR(1) process.

Diffusion introduces a slight complication to this result – if we observe two individuals with different initial productivities that converge over time, it is no longer possible to conclude that persistence is low. Instead, it could be that the less productive individual was hit with a higher match productivity. Thus, we can only identify ρ conditional on the ability to internalize match productivity, β. That is, this same procedure now identifies $\rho(\beta)$.

The last step is summarized in Proposition 3, which is to find a fixed point (β^*, ρ^*) that jointly matches the moments (Γ_1, Γ_2).

Proposition 3. *If the following two conditions hold, then there exists a unique pair* (β^*, ρ^*) *that solve equations (M1) and (M2). Those conditions are:*

$$\Gamma_1^{empirical} \in \left(1, \frac{\int \pi(\hat{z}) m_H(z,\hat{z}) d\hat{z}}{\int \pi(\hat{z}) m_L(z,\hat{z}) d\hat{z}}\right) \tag{C1}$$

$$\Gamma_2^{empirical} \in \left(1, 1 + CV(z)^2\right) \tag{C2}$$

where $CV(z)$ *is the coefficient of variation of baseline productivity among treatment firms.*

treatment. Any measurement that occurs over time, such as this one, requires taking into account the decay of the effect. Thus, this moment cannot separate β from ρ. The easiest way to see this is to assume that there is no productivity decay over time, so that $\rho = 1$. In that case, Γ_1 would directly pins down β.

Proof. Define:

$$\begin{aligned} G_1(\rho,\tilde{\beta}) &= \Gamma_1 \frac{\int\int z dM(z,\hat{z}) \int\int z^{\rho} \max\left[1,(\hat{z}/z)^{\tilde{\beta}}\right] dM(z,\hat{z})}{\int\int z^{1+\rho} \max\left[1,(\hat{z}/z)^{\tilde{\beta}}\right] dM(z,\hat{z})} \\ G_2(\rho,\tilde{\beta}) &= \Gamma_2 \frac{\int\int z^{\rho} \max\left[1,(\hat{z}/z)^{\tilde{\beta}}\right] dM_L(z,\hat{z})}{\int\int z^{\rho} \max\left[1,(\hat{z}/z)^{\tilde{\beta}}\right] dM_H(z,\hat{z})} \end{aligned}$$

Then define:

$$T(\rho,\tilde{\beta}) = \begin{bmatrix} \rho G_1(\rho,\tilde{\beta}) \\ \tilde{\beta} G_2(\rho,\tilde{\beta}) \end{bmatrix}$$

Last, define:

$$B(\rho,\tilde{\beta}) = (G_1(\rho,\tilde{\beta}) - 1)^2 + (G_2(\rho,\tilde{\beta}) - 1)^2$$

The proof works as follows:

1. Prove G_1 and G_2 are strictly convex.
2. Prove $(\rho,\tilde{\beta}) \in [0,1]^2 \implies T(\rho,\tilde{\beta}) \in [0,1]^2$. This is true under the conditions above.
3. Since T is obviously continuous, then T has a fixed point in $[0,1]^2$ by Brouwer's FPT. The $(\rho,\tilde{\beta})$ that is a fixed point in T solves both moment equations above, proving existence.
4. Any $(\rho,\tilde{\beta})$ that is a fixed point of T also solves $B(\rho,\tilde{\beta}) = 0$. Since G_1 and G_2 are strictly convex, B is strictly convex. Also, clearly all values of B are weakly positive. Therefore, any zero of B is unique. Therefore, T has a unique fixed point. This proves uniqueness.

Proofs of parts 1 and 2 follow. The arguments above prove parts 3 and 4, conditional on the first two parts being true. ■

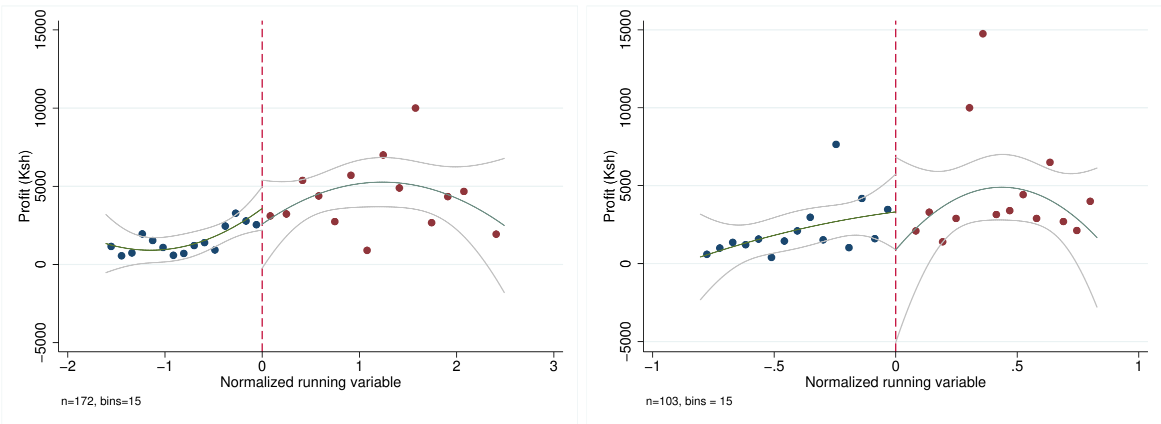
C Empirical Impact on More Productive Member of the Match

Since the more productive members of treatment matches were not randomly selected, we require a different approach to identify any effect on these business owners. Our design allows us to use the selection procedure to identify the causal impact of being chosen. Specifically, we surveyed both those chosen for the program and those just below the cutoff for selection, then employed a regression discontinuity design to study the impact of being chosen into the program.

Figure 10 plots profit along with a fitted quadratic and its 95 percent confidence interval. Figure 10a uses the entire sample, while Figure 10b uses the Imbens and Kalyanaraman (2012) procedure to choose the optimal bandwidth. Both use 15 bins on either side of the cutoff. Figure 10 suggests no statistically discernible discontinuity around the cutoff.

Figure 10: Profit for mentors and non-mentors (from Brooks et al., 2018)

(a) Full sample (b) IK Optimal Bandwidth



We next test this more formally. In particular, letting $\bar{\varepsilon}$ be the cut-off value for mentors, we run the regression

$$\pi_i = \alpha + \tau D_i + f(N_i) + \nu_i \tag{C.1}$$

where π_i is profit, $D_i = 1$ if individual i was chosen as a mentor ($\widehat{\varepsilon}_i \geq \bar{\varepsilon}$, $f(N_i)$ is a flexible function of the normalized running variable $N_i = (\widehat{\varepsilon}_i - \bar{\varepsilon})/\sigma_\varepsilon$, and ν_i is the error term. The parameter τ captures the causal impact of being chosen as a mentor. We use local linear regressions to estimate the treatment effects on profit and inventory, along with business practices of record keeping and marketing. The results are in Table 6, and we find that being a mentor has no statistically significant effect on profits. Moreover, there is no change in marketing or record-keeping practices, which one might associate with productivity. There is some evidence that inventory spending decreases, but it cannot be statistically distinguished from zero. Overall, we find little evidence that entering into a match changes either business scale or business practices for the more productive member of the match. This is consistent with the max function in the forward equation for productivity (equation 2.1), which is assumed here and in much of the existing literature.

Table 6: Regression discontinuity results for matched firm treatment effect (Brooks et al., 2018)

Percent of IK optimal bandwidth	Scale		Practices	
	Profit	Inventory	Marketing	Record keeping
100	-503.18	-3105.87	0.01	0.02
	(1321.82)	(2698.11)	(0.11)	(0.18)
150	300.19	-2585.22	0.01	0.07
	(1407.26)	(2291.34)	(0.09)	(0.14)
200	322.09	-123.59	0.01	0.10
	(1324.17)	(1964.08)	(0.08)	(0.13)
Treatment Average	4387.34	8435.79	0.08	0.85
Control Average	1794.09	4039.20	0.13	0.63

Table notes: Statistical significance at 0.10, 0.05, and 0.01 is denoted by *, **, and, ***. Profit and inventory are both trimmed at 1 percent.

D Allowing for Idiosyncratic Distortions

The model assumes throughout that $\pi \propto z$. One scenario in which this would not be the case is if individuals were subject to some unobserved distortion ν. We take this up in this section, showing how such distortion bias both the estimates of our diffusion parameters, and how that feeds into the quantitative results. We therefore augment the model and assume that on birth, agents still draw their initial productivity $z \sim G$, but now also draw an i.i.d. productivity shifter ν where $\log(\nu) \sim N(0, \sigma_\nu)$ and is fixed throughout life. This requires another state variable for the model, and profit is now defined as

$$\pi(z, \nu) = \max_{l \geq 0} \ (\nu z)^\alpha l^{1-\alpha} - wl \tag{D.1}$$

which implies an update to our original assumption ($\pi \propto z$). Now, we have $\pi \propto z\nu$. The value of having entrepreneurial skill z and distortion ν is

$$v(z, \nu, M) = \max\{\pi(z, \nu), w(1-\tau)\} + (1-\delta)\gamma \mathbb{E}_{z'|z} v(z', \nu, M'). \tag{D.2}$$

Note that this has the additional implication of changing occupational choice among model agents, who now make decisions in part based on their additional parameter ν. We assume that ν is not transmitted, in the sense that a distortion would not be transmitted across agents, and so the rest of the model is unchanged.

Our goal is to study how the quantitative gains from policy change as σ_ν varies. To do so, we re-estimate the model with $\sigma_\nu > 0$ and measure the gains from optimal policy, as in the main text. In the main text, this involves the following steps: (1) estimate (β, ρ), (2) estimate θ, (3) calibrate remaining parameters, and (4) solve for the efficient allocation. The introduction of ν changes each step, which we discuss in turn.

D.1 Bias in Diffusion Coefficients β and ρ

We first note that such distortions will bias our estimates of parameters in the law of motion for productivity, β and ρ. With $\pi \propto z$, these could be read off the main estimating regression in the text,

$$\log(\pi_i') = \widetilde{c} + \rho \log(\pi_i) + \beta \log\left(\frac{\hat{\pi}_i}{\pi_i}\right) + \varepsilon_i. \tag{D.3}$$

Now, these estimates are biased. To see this, rearrange the structural equation

$$\log(z_i') = c + \rho \log(z_i) + \beta \log\left(\frac{\hat{z}_i}{z_i}\right) + \varepsilon_i \tag{D.4}$$

under the assumption that we observe $\pi_i \propto z_i \nu_i$ to get

$$\log(\pi_i') = \tilde{c} + \underbrace{(\rho - \beta)}_{\equiv \eta} \log(\pi_i) + \beta \log(\hat{\pi}) + \Big[\varepsilon_i + \log(\nu_i) - \underbrace{(\rho - \beta)}_{\equiv \eta} \log(\nu_i) - \beta \log(\hat{\nu}_i)\Big]. \quad \text{(D.5)}$$

The term in brackets is the regression error.[30] Estimating this regression implies

$$\text{plim}\, \widehat{\eta} = \eta \left(\frac{\sigma_z^2}{\sigma_z^2 + \sigma_\nu^2} \right)$$
$$\text{plim}\, \widehat{\beta} = \beta \left(\frac{\sigma_{\hat{z}}^2}{\sigma_{\hat{z}}^2 + \sigma_\nu^2} \right)$$

and thus the estimates of β and $\eta \equiv \rho - \beta$ are both biased toward zero. This further implies that $\rho = \eta + \beta$ is biased downward as well. Thus, under a given value of σ_ν^2, unbiased structural parameters (β, ρ) require an adjustment to the coefficients we observe from the regression.

D.2 Updating the "Directedness" Parameter θ and Remaining Calibration

The "directedness" parameter θ uses only the average treatment effect. Since profit shows up only as a dependent variable in this regression, the distortions induce no direct bias in the point estimate of this regression, only increasing the standard error of the point estimate. We leave this latter issue aside here to focus in the induced bias in the structural parameters. Note, however, θ depends on the value of (β, ρ) and thus will change in response idiosyncratic distortions. Similarly, the remaining calibration will be updated to take into account the new values.

D.3 Quantitative Results

D.3.1 Quantitative Experiment and Parameter Updates

Procedure Given the discussion above, our procedure therefore works as follows. Assume some value for σ_ν. Since we observe the variance of log profit for both treated firms and their matches ($\sigma_\pi^2 = 0.671$ and $\sigma_{\hat{\pi}}^2 = 0.204$), we can use them to compute $\sigma_z^2 = \sigma_\pi^2 - \sigma_\nu$ and $\sigma_{\hat{z}}^2 = \sigma_{\hat{\pi}}^2 - \sigma_\nu$. Then use our regression results in (D.5) to back out the structural parameters (β, ρ) as

$$\beta = \hat{\beta}^{OLS} \left(\frac{\sigma_{\hat{\pi}}^2}{\sigma_{\hat{z}}^2} \right) = 0.538 \left(\frac{0.204}{\sigma_{\hat{z}}^2} \right) = \frac{0.110}{\sigma_{\hat{z}}^2}$$
$$\rho = \hat{\eta}^{OLS} \left(\frac{\sigma_\pi^2}{\sigma_z^2} \right) + \hat{\beta}^{OLS} \left(\frac{\sigma_{\hat{\pi}}^2}{\sigma_{\hat{z}}^2} \right) = 0.057 \left(\frac{0.671}{\sigma_z^2} \right) + 0.538 \left(\frac{0.204}{\sigma_{\hat{z}}^2} \right) = \frac{0.038}{\sigma_z^2} + \frac{0.110}{\sigma_{\hat{z}}^2}$$

[30] Note that we assume the max operator does come into play. This is not critical, but focuses the discussion on the distortions themselves.

From there, update the remaining diffusion parameter θ and calibrated parameters (σ, c, τ) to match the same moments in the main text.[31]

Implementation In our quantitative experiment, we set σ_ν^2 such that it induces a true parameter value of $\beta = 0.807$, implying that our regression estimate $\hat{\beta}^{OLS} = 0.538$ is two-thirds of the structural parameter β.[32] This implies that $\sigma_\nu = 0.26$. Table 7 shows the updated parameter values.

Table 7: Updated Parameter Values

Model Parameter	Description	Parameter (Baseline)	Parameter (Distortions)
Exogenously varied:			
σ_ν	St. dev. of distortions	0	0.260
Group 1			
β	Intensity of diffusion	0.538	0.807
ρ	Persistence of productivity	0.595	0.870
θ	Directedness of search	-0.417	-0.161
Group 2			
σ	St. dev. of exogenous productivity shock distribution	0.877	0.746
c	Growth factor in productivity evolution	-3.107	-2.310
τ	Tax on wage earnings	0.999	0.998
Group 3			
δ	Death rate of firms	0.016	0.016
σ_0	St. dev. of new entrant productivity distribution	0.961	0.922
α	Cobb-Douglas exponent on labor	0.67	0.67
Group 2 Sum of Squared Errors		$\mathbf{2.73 \times 10^{-6}}$	$\mathbf{4.86 \times 10^{-6}}$

Table notes: Group 1 is jointly chosen from the experimental data. Parameters in Group 2 are calibrated to jointly match moments. Group 3 are also set to match baseline data moments, but match 1-1 with target moments. Both are set to match the same set of moments discussed in the main text (see Table 3 for details). SSE measure includes only parameters in Group 2 that are jointly set.

D.3.2 Quantitative Gains from Policy

We now study the gains from optimal policy. We assume that the planner can observe the distortions ν but cannot change them, allowing us to focus on the diffusion externality. This amounts to the planner choosing a cut-off function $\overline{z}(\nu)$ in which in all individuals with distortion ν and $z \geq \overline{z}(\nu)$ operate a firm and those with $z < \overline{z}(\nu)$ become workers.

[31] Recall, they are (1) the variance of log profit among all firms, (2) the ratio of average profit of all firms to new entrants, (3) the fraction of agents employed as workers.

[32] Or, said differently, variance in true productivity z makes up two-thirds of observed profit variation within the firms we use to match treatment firms.

Specifically, the planner's problem is now

$$\max_{\overline{z}(\nu)} \int_0^\infty \int_{\overline{z}(\nu)}^\infty y(z,\nu)dM^*(z)dH(\nu) \tag{D.6}$$

$$s.t. \quad M^*(z') = \delta G(z') + (1-\delta)\int_0^\infty \int_0^\infty F(\log(z') - \log(\max\{\hat{z}^\beta z^{\rho-\beta}, z^\rho\}) - c)d\widehat{M}(\hat{z};\theta)dM^*(z)$$

$$\widehat{M}(\hat{z};\theta) = \left(\frac{\int_0^\infty \int_0^{\hat{z}} \mathbb{1}[z \geq \overline{z}(\nu)]dM^*(z)dH(\nu)}{\int_0^\infty \int_0^\infty \mathbb{1}[z \geq \overline{z}(\nu)]dM^*(z)dH(\nu)}\right)^{\frac{1}{1-\theta}}.$$

The main results are in Table 8. Column 1 reproduces the baseline results from the main text, while column 2 covers the updated model with idiosyncratic distortions.

Table 8: Δ Outcomes

	Baseline (1)	Distortions (2)
Average Income	3.45	7.03
Fraction working	2.17	2.05
Average entrepreneurial productivity	8.97	984.70
Wage	0.48	1.63

Table notes: Equilibrium changes are measured as the ratio of efficient to laissez faire minus one. Multiply by 100 for percentage change.

The gains from policy are approximately twice as large as the baseline case. The rationale for this follows from the previous discussion. As we showed in Section D.1, our estimate $\hat{\beta}^{OLS}$ is biased downward in the presence of idiosyncratic distortions. Relative to our baseline model, our structural value of β is 50 percent higher in this case. As we showed in the main text, β plays a critical role in governing the gains from policy. Combined, these results show that our results in the main text are a lower bound on the estimated gains from policy.

E Additional Results

E.1 Income Changes by Parameters in Laissez Faire and Efficient Economies

Figure 11: Average Income Changes as Diffusion Parameters Vary

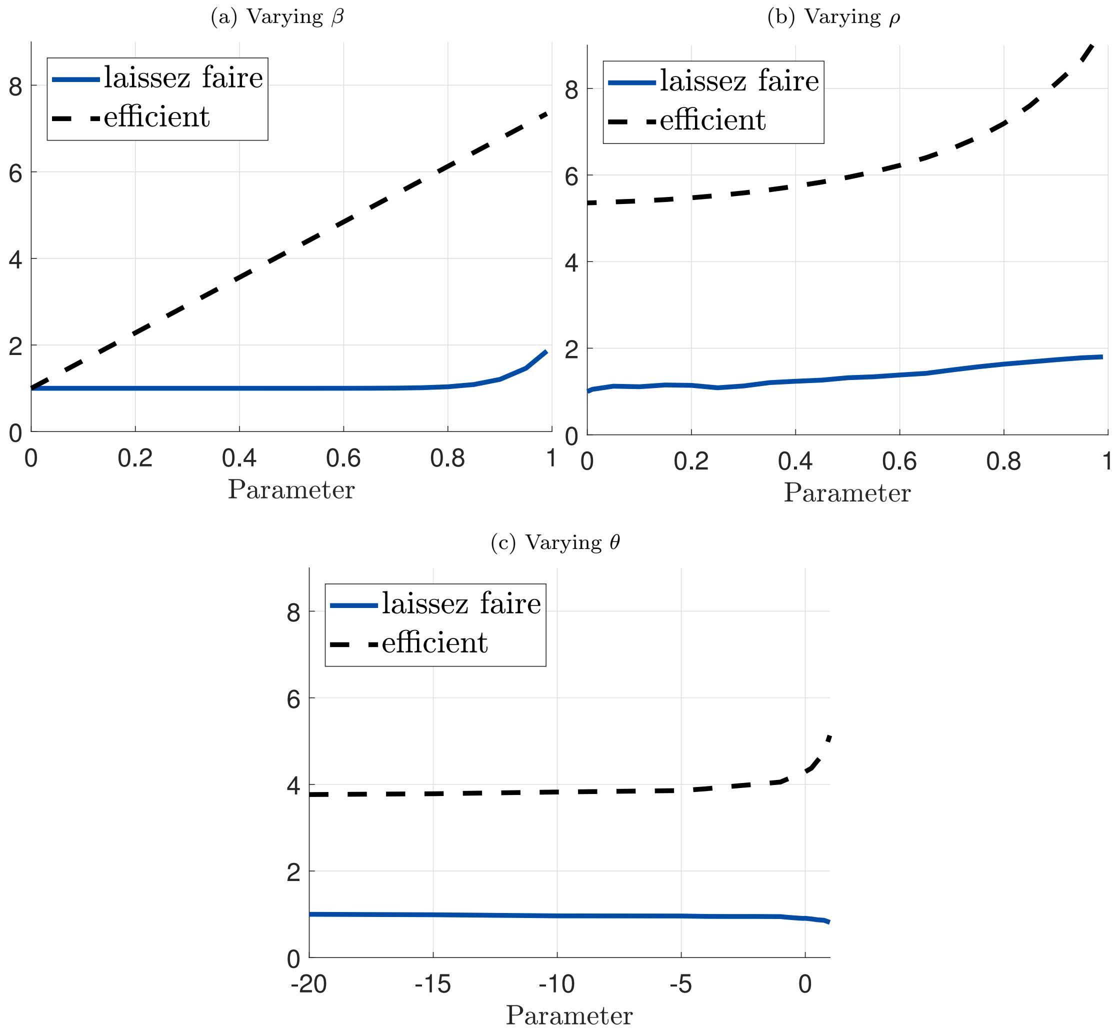


Figure notes: For each subfigure, *laissez faire* income is normalized to 1 at the smallest parameter value value.

E.2 Allowing the Planner to Change Search Frictions (via the directedness parameter θ)

Our RCT adjusts the set of matches available to the treatment group. This is similar in spirit to a change in the value of θ for the treatment group. Therefore, one might be interested in using the model to study the effects of changing θ.

Clearly any increase in θ is a technological improvement in that it makes the distribution of imitation draws better, and will therefore increase income. Here we study the interaction between a policy that increases θ and the gains from implementing optimal policy. We show that the additional gains from optimal policy relative to the baseline case is small.[33]

Of course, we have no sense of what the costs the planner faces when changing θ. We therefore take an extreme case and assume that it can be changed costlessly, studying the gains from policy when θ can costlessly be increased to a fixed higher value. We then vary this fixed higher value and show how the results change.

To be explicit, denote $L(\theta)$ and $E(\theta)$ as the level of income in the *laissez faire* and efficient economies defined by parameter θ. The object of interest in the main text is therefore $W(\theta) := E(\theta)/L(\theta)$. Here, we ask how much larger those gains are if the planer is allowed to change θ from its baseline value $\theta^* = -0.417$ to some other value θ'. The gains from policy in this case are given by

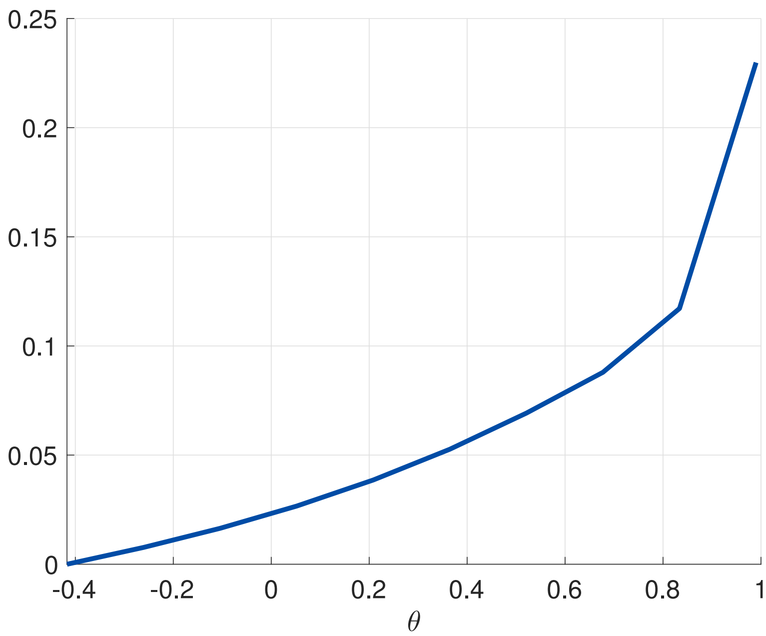
$$\widetilde{W}(\theta') = \frac{E(\theta')}{L(\theta^*)}.$$

where $\widetilde{W}(\theta^*) = W(\theta^*)$ by definition. We measure the excess gains from this additional margin of adjustment, given by

$$G(\theta') := \frac{\widetilde{W}(\theta')}{\widetilde{W}(\theta^*)} - 1 = \frac{E(\theta')}{E(\theta^*)} - 1$$

This measure G measures how much larger the gains are by allowing the planner to simultaneously achieve a higher θ relative to the baseline. The results are in Figure 12.

Figure 12: Excess Gains from Policy When θ is Allowed to Change



[33]This is essentially a corollary of the results in Section 5.2, where we show that varying θ induces small changes in the gains from policy.

Recall that the baseline gain from policy (denoted here as $\widetilde{W}(\theta^*)$) is 345 percent. The additional gains shown in Figure 12 are substantially smaller. For example, if the planner is allowed to shift θ from its baseline value $\theta = -0.417$ to $\theta = 0.20$, the gains from optimal policy increase by an additional 4 percent. Even in the extreme case, where the planner can push all matches into the right tail as $\theta \to 1$ (which we view as an unrealistic possibility), the excess gains are 25 percent. The rationale for this result is that the planner's baseline policy level – a wage subsidy to eliminate firms – is a substitute for shifting matches via the θ parameter. Thus, allowing the planner this second policy lever plays a relatively minor role.

E.3 Importance of Diffusion Parameters in Model-Run RCT

Figure 13 shows how the average treatment effect varies as one varies the diffusion parameters (β, ρ, θ) individually.

Figure 13: Impact of Estimated Diffusion Parameters on the Average Profit Treatment Effect

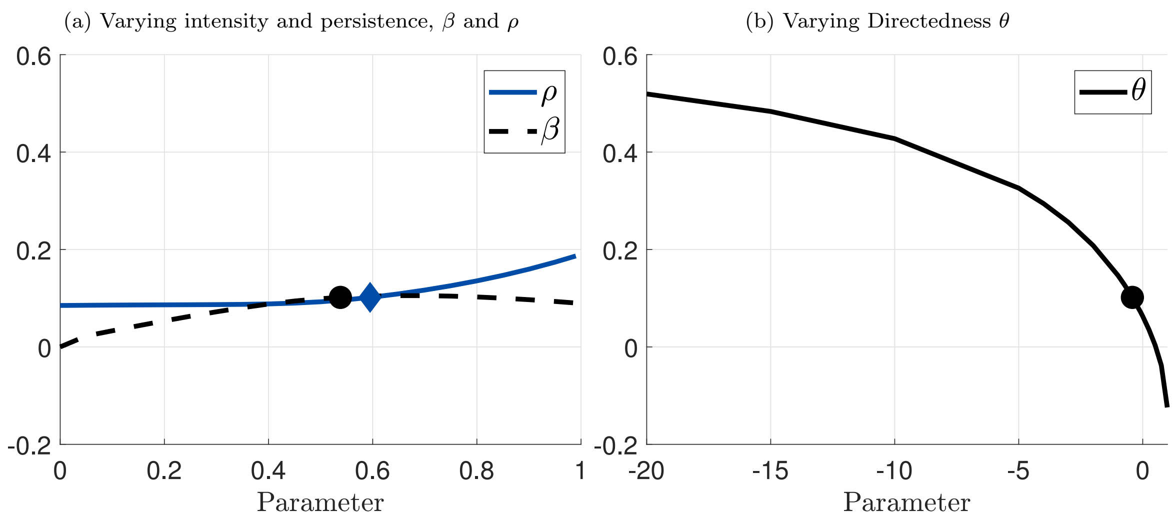


Figure notes: Each line varies the listed parameter holding all other model parameters fixed at baseline values. The points indicated by circles and diamonds are our baseline estimates.

There are a number of results that stand out here. First, θ plays a substantially larger role in generating the treatment effect than either β or ρ. This is the opposite of the impact on the GE externality, where β dominates. The impact of θ ranges from an average treatment effect of 53 percent when $\theta = -20$ to -12 percent when $\theta = 0.99$ (the rationale for the scale differences is discussed in the main text).[34]

The second important result is that the treatment effect is in fact decreasing in θ, while the gains from equilibrium policy were increasing in θ. Thus, not only do the

[34]Note that negative effects are consistent with the model. As $\theta \to 1$ all the mass is being moved to the best possible firm. Since our empirical draws include a distribution of firms, negative treatment effects are feasible in such extreme cases.

parameters generate different magnitudes between PE and GE results, but inferring GE relevance from the RCT would push a policy maker in the wrong direction. The rationale for this difference highlights the divergent roles played by θ in the partial equilibrium RCT and in general equilibrium. In PE, the treatment effect is maximized in economies in which the treatment provides the largest shock – that is, in economies in which control firms find it difficult to meet with high-productivity individuals on a regular basis. Thus, the treatment effect declines in θ. This same force, however, limits the planner's ability to extract gains from the economy. If individuals create matches by seeking out mostly agents from the left tail of the productivity distribution (i.e., low θ), the planner has limited ability to shift the distribution of matches toward the right tail. Put differently, the scope for learning increases when individuals can more easily find the best firms to learn from. This requires a high θ.

These divergent results highlight the crucial importance of understanding key structural parameters for policy making in such a context. If a policymaker naïvely looked only at treatment effects to extrapolate GE impact, she would run the risk of choosing economies based exactly on parameters that play little role at scale. Thus, not only do the parameter estimates matter for quantifying the gains from policy in equilibrium, they can similarly confuse estimates of results commonly used to identify economies in which policymakers will deploy such policy instruments. To study this in more detail, we break the model-derived RCT results into two pieces: understanding the initial impact at $t = 1$ then the fade-out in quarters $t = 2, \ldots, 5$.

F Cheat Sheet of Parameters and Assumptions

This Appendix summarizes the key assumptions and parameter values used in the main text, for simplicity and to facilitate ease of comparison within the paper. Our goal is to identify:

- "Intensity" parameter β
- "Persistence" parameter ρ
- "Directedness" parameter θ

The relationship between these parameters is detailed below in our assumptions on the economic environment.

Three Assumptions on the Economic Environment

Our economic environment requires three assumptions, Assumptions 1, 2, and 3 in the main text.

Assumption 1. *Given a productivity z this period, an imitation opportunity $\hat{z}$, and a random shock ε, productivity next period z' is given by*

$$z' = e^{c+\varepsilon} z^{\rho} \max\left\{1, \frac{\hat{z}}{z}\right\}^{\beta}, \tag{F.1}$$

where the parameter c is a constant growth term, β is diffusion intensity, and ρ is persistence.

Assumption 2. *In any period, profits are proportional to productivity. That is, for any two firms i and j earning profits π_i and π_j, $\pi_i/\pi_j = z_i/z_j$.*

Assumption 3. *The imitation opportunity $\hat{z}$ is drawn by a firm with productivity z from a distribution characterized by the cumulative density function $\widehat{M}(\hat{z}; z, \theta)$, a known function. For every z and $\hat{z}$, $\widehat{M}$ is continuous in θ and $\theta_1 < \theta_2 \implies \widehat{M}(\hat{z}; z, \theta_2)$ first order stochastically dominates $\widehat{M}(\hat{z}; z, \theta_1)$.*

Assumption on Variation in the Data-Generating Process

With the three assumptions on the economic environment, we summarize the requirements on the data generating process in Assumption 4.

Assumption 4. *A set of agents with productivity distributed $H(z)$ are observed in two consecutive periods. The set of agents is partitioned into two subsets characterized by distributions $H_C(z)$ and $H_T(z)$ (i.e., "control" and "treatment"). The following conditions hold:*

1. *Agents in H_T and H_C draw their ε shocks from the same distributions*

2. *The matches for agents in H_C are not observable, and distributed $\widehat{M}(\hat{z}; z, \theta)$*

3. *The matches for agents in H_T are observable, and distributed $\widehat{H}_T(\hat{z}) \neq \widehat{M}(\hat{z}; z, \theta)$. Moreover, every match $\hat{z}$ is greater than the z to which it is matched.*

4. *For any arbitrary partition of the treatment group, characterized by $H_T^1(z)$ and $H_T^2(z)$, agents in both groups draw their ε shocks from the same distribution*