Identification and Quantitative Implications of an Equilibrium Diffusion Model*

Wyatt Brooks
University of Notre Dame

Kevin Donovan
Yale University

Terence R. Johnson
University of Notre Dame

September 2019

Abstract

An increasingly utilized class of general equilibrium models includes inter-firm knowledge spillovers through diffusion. Standard methods to calibrate critical diffusion parameters require making assumptions about the economic environment, then using the resulting structure to map these parameters onto more easily observed empirical moments. Within this class of models, we prove that randomly varying interactions uniquely identifies a small set of parameters characterizing the diffusion process independent of the remaining economic environment. We provide an application of our results in Kenya, where we conduct a randomized controlled trial matching firms from the left tail of the profit distribution to those from the right. Despite matching the quick fade out of the small-scale experimental treatment effect, the model simultaneously implies a large general equilibrium diffusion externality. Key is that critical parameters push the partial and general equilibrium magnitudes in different directions. This matters: if a policy-maker selected economies in which to implement optimal policy based solely on the magnitude of their experimental impact, she would in fact minimize the possible welfare gains. Thus, the ability to properly estimate such parameters is critical not only for measuring the equilibrium importance of diffusion but also for the interpretation and extrapolation of smaller-scale empirical studies.

*Thanks to conference and seminar participants at Arizona State, Columbia, Illinois, Notre Dame, NYU, Stanford, UBC, the Firms in Emerging Economies Workshop at Jinan University, the NBER Growth Meeting, the Society for Economic Dynamics, and Y-RISE for helpful comments, especially Hugo Hopenhayn, Melanie Morten, Jesse Perla, Tommaso Porzio, Kevin Song, Meredith Startz, and Mike Waugh. 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, wbrooks@nd.edu; Donovan, kevin.donovan@yale.edu; Johnson, tjohns20@nd.edu

1 Introduction

A growing literature focuses on how knowledge or skill transfer among firms could play an important role in the development process (??????). However, quantifying the importance of this channel is difficult. The broad notion of productivity used in these models contains a number of skills and information that can be diffused, implying that the diffusion process cannot be directly observed. The diffusion process, therefore, is governed by functions whose inputs have no direct empirical counterparts.

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 aggregate firm exit rates. In this paper we ask the extent to which properly targeted empirical work can limit the structure required to identify diffusion parameters in this class of models. To do so, we lay out sufficient variation in the data-generating process that identifies key diffusion parameters independent of such assumptions, then apply our procedure among Kenyan firms. We build these results into commonly-used general equilibrium models to study the quantitative magnitude of diffusion at scale, and use the results to highlight the importance of parameter estimates for interpretation of small-scale experimental findings relative to the general equilibrium importance of diffusion.

A full accounting of diffusion requires taking into who individuals meet with, the static benefit from a match (how much one's own productivity changes in response to a match), and the dynamic effects (the propagation of that static effect to other economic agents). Under some commonly used assumptions detailed in Section ??, we show how exogenous variation in how matches are created can be used to measure these forces. Specifically, we propose the following hypothetical data-generating process. We first randomly divide a population of firms into a treatment and control group. The treatment group is then guaranteed access to a random draw from a set of highly productive firms drawn from the same population. This two-step randomization guarantees both the standard exclusion restriction between control and treatment, but also a second within the treatment group.

We then show how this variation identifies key diffusion forces. To fix ideas in the main body of the paper, we generally discuss the intuition for our results in the context of a random search model, in which agents randomly meet and costlessly internalize some portion of their match's productivity. We do so only to fix the economic interpretation of certain model forces in a commonly used framework (e.g.

????). As we emphasize in Section **??**, the theoretical results apply to a wide range of models with various diffusion approaches, including those that swap out random search for non-random assignment, bargaining between matched agents, and congestion in the vein of **?**. The interpretation of some parameters differ in these various cases, and thus we err on the side of simpler intuition and clarity by focusing the discussion on one model, leaving the details of others for the Appendix.[1]

With that in mind, we identify the static effect by comparing treatment firms of identical productivity who receive different matches. If those matches are independent of initial conditions – guaranteed by the randomization within the treatment – we can identify the immediate impact of match by comparing differences in *ex post* profit. The dynamic effect then depends on the persistence of these static gains, which allows firms to propagate gains from past meetings as they interact with new firms in the future. This parameter is identified similarly to how one would identify the persistence of an exogenous AR(1) process, with one complication. With diffusion, there are two reasons a firm with high profit today would have high profit tomorrow: high persistence or matching with a high productivity firm. The latter force is of course irrelevant in a world without diffusion, which is what allows a lagged regression to identify persistence. Thus, diffusion introduces a potential bias into this procedure. However, because the treatment involves creating and observing matches directly, we can correct this bias and utilize a similar procedure. As we formalize in Section **??**, it turns out that measuring both forces requires only two coefficient estimates from a single linear regression, thus allowing a clean map between empirics and model parameters.

The explanation above focuses on forces that govern outcomes conditional on a match. We still need to deal with the issue of who actually meets. For example, even within a random search framework, one might imagine that matches are drawn from some particular part of the overall firm productivity distribution (i.e., firms may "direct" search).[2] The complication here comes from the fact that we cannot observe control matches. However, recall that the treatment guarantees a high quality match. Our insight is that a small average treatment effect tells us something about the control group: that they must already be quite good at generating high quality matches. Therefore, the average treatment effect reveals important information about the search process of the control group, despite not being directly observable.[3]

[1] Specifically, we emphasize this in Section **??** by laying out only the assumptions required for identification. Any model that satisfies them can be identified by our procedure. Focusing on one model serves only to fix the intuitive properties of the identification results and quantitative interpretations.

[2] The papers cited here generally assume matches are uniform draws from the existing distribution of firm productivities. This is a special case of our assumption.

[3] Again, while this intuition holds only the case within the framework of the model we utilize in the main text, the general idea is broader. In a model with bargaining, for example, we can interpret this same parameter as the bargaining

An implication of this procedure is that the parameters are identified independent of much of the remaining model structure. That is, we do not take a stand on equilibrium (e.g., whether the economy is in a steady state or on a transition path), or other features of the economic environment, such as occupational choice. While our results naturally require some assumptions to map the discussion above to estimable parameters, it substantially reduces the required structure in the remaining economic environment. We emphasize this in Section **??** by only laying out the assumptions required for identification, leaving the details of the full model for Section **??**, when we require them for the quantitative results.

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, except 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, properly differencing from a control group helps limit many of the potential confounding factors. This allows us to use moments like the average treatment effect to identify model parameters without specifying the complete economic environment. Our estimates can therefore be embedded into a variety of models with different assumptions on the remaining structure of the economy.[4]

With the theoretical results in hand, we then ask how quantitatively important these parameter estimates are. This first requires estimating these parameters in a particular application. To do so, we conduct a randomized controlled trial in Nairobi, Kenya to generate the proper variation, matching low-profit treatment firms with a randomly selected high-profit firm. The full set of reduced form results are available in **?**. Profits are 19 percent higher in the treatment group relative to the control and the effect is increasing in the relative profit gap between the two firms.[5] Moreover, we show that the more productive member of the match sees no change in profit or business skills that one might associate with higher productivity. These empirical results form the basis of our parameter estimation.

One key empirical finding is that the treatment effect is large at first, then quickly fades. After 18 months, there is no difference between treatment and control. Since

weight for the less productive member of the match. If low bargaining power limits the amount of productivity one can extract from a match, identifying this parameter turns out to be quite similar in practice to one in which the agent internalizes all the surplus but draws from a distribution with low average productivity.

[4] 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.

[5] We consider a number of possibilities in **?**, including profit sharing, loans, and bulk discounts, and find that none of them explain the results.

our study population was approximately 0.2% of the population, this result is unlikely to be driven by spillovers to the control group.[6] One possible interpretation is that the "partial equilibrium" failure to generate long-term benefits implies diffusion has a limited role at scale. To study this, we build our parameter estimates in a full general equilibrium model of diffusion, with the remaining parameters calibrated in a standard fashion. The model accurately predicts the quick fade-out of the average treatment effect in the simulated RCT. This is a partial equilibrium effect, in the sense that we hold fixed the distribution of control group matches, therefore explicitly ruling out control group contamination as an explanation. Yet, we simultaneously show that the equilibrium diffusion externality is large. Average income is 175 percent higher in the planner's allocation than the *laissez faire* allocation.[7]

These seemingly divergent results highlight the importance of proper parameter estimates for interpretation of and extrapolating from experimental results. Not only do the partial and general equilibrium results rely differentially on parameters we estimate, the value of critical parameters actually push the PE and GE magnitudes in opposite directions. Indeed, the same parameters that allow us to match the lack of long-run PE effects are the same that generate the large GE effect. As we show in Section **??**, our parameter estimates imply that while any single individual benefits from a match for only a short period of time, *many* individuals benefit from that same match through diffusion. The results show that one need not generate long-term effects among individual firms to imply diffusion is an important margin for economic development. Even these short-run gains get dispersed through other agents in the economy through diffusion, implying that the general equilibrium impact swamps the partial equilibrium effect.

Moreover, the results suggest caution when attempting to extrapolate partial equilibrium outcomes for general equilibrium significance. We further show in Section **??** that if a policy-maker were to run our same small-scale experiment in multiple economies, then use the long-run treatment effect to select the economies in which the gains from optimal policy were the largest (or, put differently, where the diffusion externality is the largest), she would instead select exactly the opposite. That is, the strategy to extrapolate the potential general equilibrium policy impact from a partial equilibrium experiment would cause a policy-maker to minimize the potential welfare gains. Unbiased parameter estimates are therefore critical for understanding the

[6] We set up the experiment to limit spatial spillovers by spreading out the firms selected for the study. We also test for spatial spillovers to control firms and find none.

[7] We view this as a way to measure the size of the diffusion externality, not necessarily the true policy one should implement to maximize diffusion spillovers. That naturally depends on the specifics of the model on which our results have little to say. Instead, we take our modeling cues from the literature, and deploy our estimated parameters in existing models.

magnitude of diffusion externalities in general equilibrium, but also for extrapolating smaller-scale empirical studies.

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 ?, ? and ?. Our paper shares a similar style but focuses on knowledge diffusion. Closest in this dimension is ?, who use the results from a randomized controlled trial to, in part, identify the utility cost associated with migration. 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. ? and ? highlight similar points in microcredit and education, respectively.

Our work adds empirical evidence to the literature studying innovations and knowledge in general equilibrium models (??). 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 ?, ?, ?, ?, and ?. ? 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 (??), international trade (??), and the interaction of innovation and diffusion (??).

At the same time, there exists an important micro-development literature documenting the diffusion of specific pieces of information or technology, including new crops or high-yielding seeds (??), specific planting or production techniques (??), or financial information (??). This more narrow focus has the clear benefit of allowing one to track diffusion explicitly along the dimension of interest. Our goal here is to focus on a set of equilibrium models interested in the diffusion of productivity while highlighting how similar tools can be deployed to provide insights into these broader models. That is, not only can these micro tools be useful for micro-founding aspects of the diffusion process, they can also provide useful insights directly to aggregate models in a more "top down" approach.

2 Identification of the Diffusion Process

We begin by specifying the class of diffusion processes we will study. The goal here 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 our main identification results. Therefore, along the way, we will draw attention to the required assumptions so that is clear what is required for the results.

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 $\hat{z}$.[8] The functional form of the subsequent productivity z' is given by Assumption **??**.

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*

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

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

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 **?**, for example). We address this issue directly in Section **??** and find no evidence that more productive firms gain profit from interaction with less productive firms.[9]

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,

[8] 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.

[9] 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.

profit. The requirement is summarized in Assumption **??**.

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 **??** is to assume $\pi_i = z_i$ as in **?** and **?**. A production function of the form $y = z^{\alpha}n^{1-\alpha}$, where n is labor, also satisfies Assumption **??** in a competitive labor market.[10]

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 **??**.

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 stochastic dominates $\widehat{M}(\hat{z}; z, \theta_1)$.*

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 **??** 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 **??**.

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. β captures the static effect that governs how much individuals gain immediately from a match. 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, θ governs who individuals regularly interact with. All three of these play a potentially important role in governing the total impact of diffusion.

[10] Note, however, that assumption is violated in the presence of firm-specific distortions, such as those considered in **?**. 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.

Finally, we note that all of the assumptions we have made are common in the literature, including **?**, **??**, and **?**, among others, and in some instances allow us to generalize. While we naturally require some structure on the diffusion process, we have not specified the remaining economic environment. In the remainder of this section we discuss the data variation required to uniquely identify these parameters without this additional structure. On the other hand, 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 **??** 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 **??** 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 **??** summarizes variation in the data required to identify the parameters. That is, while Section **??** makes assumptions about the model primatives, Assumption **??** lays out assumptions about the data required for identification in the context of that model. After proving the identification results, Section **??** 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 the two 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.[11] Finally, the third and fourth lay out what we require from our treatment. The third states that we can observe all 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.[12] Finally, the last assumption states a second exclusion restriction *within* the treatment group, guaranteeing that comparisons across treatment firms are unbiased.[13]

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 **??**.

Using Treatment Data for (β, ρ) The idea behind identifying (β, ρ) follows almost directly from the law of motion for diffusion. In logs, this 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 **??** ($\pi \propto z$) annd Assumption **??** ($\hat{z} > z$), this simplifies to

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

where π_i and $\hat{\pi}_i$ are baseline profit for individual i in the treatment group, and her match. Equation (**??**) is a linear regression that can be estimated directly from panel data. Intuitively, there are two ways to read (**??**). 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 the decay in profit, measured by ρ, after controlling for variation in match quality $(\hat{\pi}_i/\pi)$. Given the orthogonality built into Assumption **??** on the data-generating process, (**??**) measures both forces simultaneously. Proposition **??** formalizes this.

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

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

[11] 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.

[12] 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 **??** without the $\hat{z} > z$ assumption is still sufficient for identification in the treatment.

[13] While this paper proceeds using randomization as a way to generate variation consistent with Assumption **??**, 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 argument laid out above relies in part on the result that one can remove the max operator from (??) 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 So far, the estimation procedure has used only data from the treatment group. Now, we utilize both treatment and control groups to identify θ, 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.[14]. Figure ?? 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 ??), combined with the implied variation in matches (Assumption ??), 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 applying similar logic to the exogenous shocks ε as in the previous steps of the estimation, we can re-write the equation as

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

Given the values of β and ρ already identified, then all other parts of this equation

[14] All these statements are conditional on a fixed value of β. If $\beta = 0$, for instance, then the model cannot rationalize any difference in the treatment and control group, and θ is therefore not identified.

Figure 1: Identification of θ

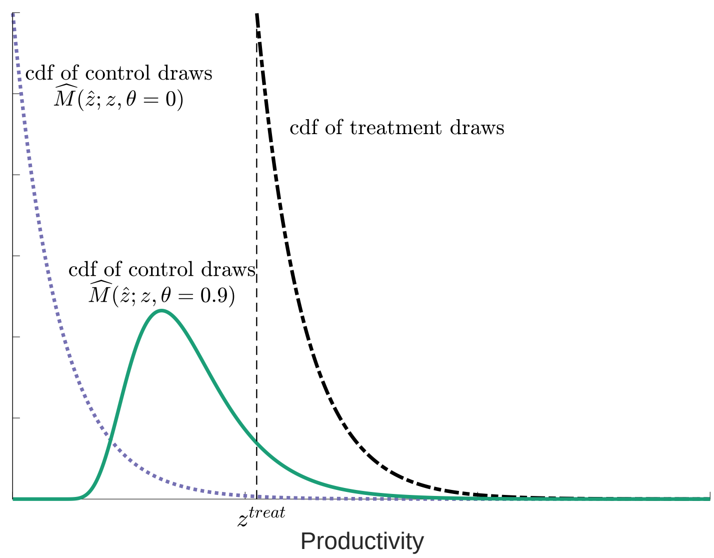


Figure notes: Figure shows the distribution of draws for treatment and control firms under different θ. The distributions are drawn Pareto, but this is for the example's sake only.

come directly from the data (after again applying Assumption **??** that $\pi \propto z$), except for the parameter θ.[15] The assumed monotonicity of $\widehat{M}$ (Assumption **??**) is sufficient to prove that any θ that solves this equation is unique. Proposition **??** formalizes the results, again developing bounds to guarantee the results.

Proposition 2. *Given the values* (β, ρ)*, the value of* θ *that solves (**??**) 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 (**??**) is θ. This follows from Assumption **??** and Proposition **??**. All that is left is to show that there exists a unique θ that solves (**??**). The right hand side is continuous in θ by the continuity of $\widehat{M}$ in Assumption **??**. 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

[15] This is conditional on Assumption **??**, 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.

stochastic dominance assumed in Assumption ??. ■

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

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.

First, note that the empirical moments required for the estimation are are easily obtained from data and use standard moments, including the average treatment effect, and what amounts to a measure of treatment heterogeneity coupled with the (normalized) coefficient from a lagged profit regression. Thus, the moments allow for a relatively straightforward link between model and data.

A second 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 nearly all 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 ?? and ??).

A more subtle restriction is built into our assumption on $\widehat{M}$ in Assumption ??. 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 ?, ?, and ?, who assume uniform draws from the existing distribution of operating firms. In that case, if M is the cdf of operating firm productivity, $\widehat{M}(\hat{z}; z, \theta) \equiv M(\hat{z})$. ? and ? make the same assumption, but extend these models by endogenizing a tradeoff between production and searching for a match. These models generally fail Assumption ??, 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.

Two things are worth emphasizing about this result though. First, assumptions on $\widehat{M}$ only affect the estimation of θ. If we followed ? and ? and fixed θ *ex ante*, the identification of β and ρ goes through unchanged.[16] Alternatively, one could add additional assumptions to guarantee our procedure still holds. In the cases outlined above, a sufficient assumption is a stationary equilibrium, which guarantees that the

[16] In our quantitative exercise, in which we utilize a random search framework, we find that varying θ has little impact on the results. Thus, at least in this context, fixing θ *ex ante* is of little importance.

value of search remains stationary. Thus, even in these cases, the amount of structure required to identify the diffusion parameters is dramatically reduced.

Second, Assumption **??** still allows for a wide set of underlying processes. For example, in addition to random search, the assumption admits pure assignment models as well. We detail a number of different underlying models that satisfy our assumptions in Appendix **??**, and highlight the variety of interpretations one can put on θ depending on the exact model details.

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 **?**, 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.[17]

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.

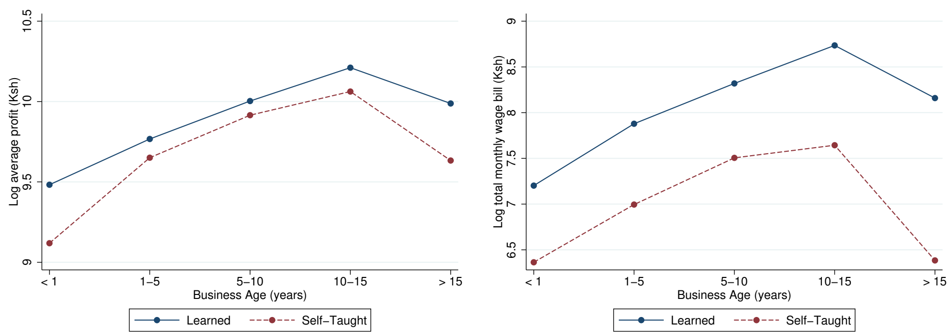
Qualitative Evidence on the Importance of Learning To begin, Figure **??** plots business scale measures based on self-reported learning methods from the baseline survey. Fifty-five percent of all firms claimed they were self-taught, while the rest claimed to learn either from another business operator, in school, or through an apprenticeship.

[17] In **?**, 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 (supplier locations, etc.) whereas classroom training provides information on topics that are designed to be orthogonal to the market in which they are deployed (accounting, marketing, etc.).

Figure ?? 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 ?? show that self-taught firms pay a smaller total wage bill.

Figure 2: Self-Reported Learning Methods and Business Scale

(a) Average profit (b) 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.[18] 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 these young owners chosen into the treatment against those other young business owners who were not. These individuals were then surveyed 7 times over 17 months, at $t = 1, 2, 3, 4, 7, 12, 17$.[19][20]

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 hopefully minimized 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 to serve as a mentor, 95 percent accepted. We reached a sufficient number of mentors at the 51st percentile of our recruitment frame. These matches with treatment firms were random conditional on industry.

To summarize, Figure ?? plots the cumulative distribution function of baseline

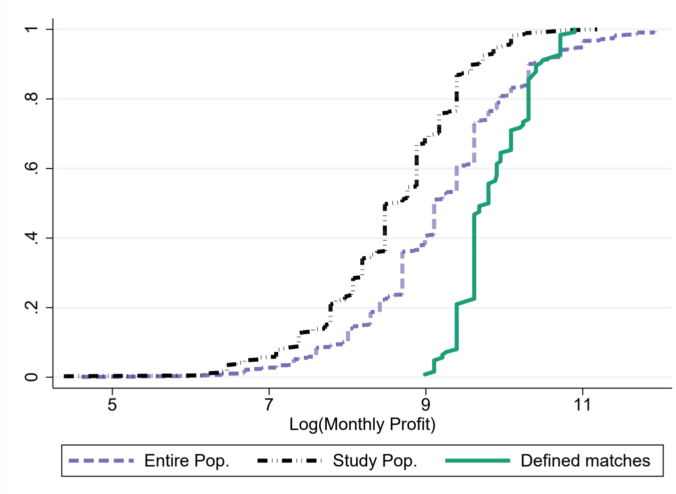
[18] The sex selection criteria is to limit heterogeneity outside the model. Note, however, that females make up 65 percent of business owners in Dandora and 71 owners with businesses open less than 5 years.

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

[20] In the context of the model, we need only that the productivity and shock distributions are the same. Of course, when we go to the data, we also want to make sure that un-modeled and unobserved characteristics are equalized across the two groups, which the randomization guarantees.

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. 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.[21] For simplicity and ease of reference to the more detailed discussion in **?**, we refer to these two groups as mentees and mentors throughout. We emphasize, however, that they

[21] 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 building from which they conducted business (most mentees did not). 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.

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. **?** 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.[22] Table **??** reports the results. The only significant difference is in age, and the magnitude is small.

3.3 RCT Results

Impact on Mentees Over the course of one year, followup surveys were conducted to measure business activity and profit among mentees as well as those that received business classes and the control group. We run the regression

$$\pi_{it} = \alpha + \mathbf{M}_i \beta + y_{i0}\delta + X_i \eta + \theta_t + \epsilon_{it}. \tag{3.1}$$

where π_{it} is the profit of firm i at time t, $M_i = 1$ if matched with another firm, X_i are firm-specific controls, and θ_t is time fixed effects. The treatment effects are summarized in Table **??**.

We found that entrepreneurs receiving a mentor realized a statistically significant increase in profit. Moreover, the increase is economically significant, representing 23 percent of the control group's mean profit. In the second column of Table **??**, we break out these results by the profitability of the more profitable member of the match. Here we divide the set of mentors by their percentile ranking within the whole set of mentors. We find that the point estimates are ordered by the profitability of each group, so that the most profitable mentors generated the largest treatment effects for their mentees.

As discussed in detail in **?**, we found that the mechanism for this increase in profit was reduced costs, such as lower prices for inputs. This is important because greater

[22] We have experimented with a number of different ways to compute the balance table, and all show the same results. We report this indicator instead of a continuous measure to increase precision to give the data the best chance at uncovering a difference, though the results are the same in either case.

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.

profits driven by greater demand may mean that the mentee group took sales away from the control group, which would bias these results upward. However, we found no significant increase in revenue, which rules out that concern. Instead, mentees learned how to achieve the same scale at lower cost, consistent with higher productivity in the sense of the model. Moreover, we are able to rule out a number of other alternatives that would be inconsistent with knowledge diffusion, such as mentors giving loans to mentees.

Table 2: Profit Effects in Randomized Controlled Trial (from **?**)

Profit	(1)	(2)
Matched	414.46 (133.07)**	
Matched with firm in ...		
$(0, 25)$ pctile		356.03 (180.17)**
$(25, 75)$ pctile		449.11 (172.01)***
$(75, 100)$ pctile		514.54 (270.54)*
Control mean	1803.48	1803.48
Obs.	1902	1902
R^2	0.091	0.090
Controls	Yes	Yes

Table notes: Standard errors are clustered at the individual level and are in parentheses. Controls include secondary education, age of owner, an indicator for any employees, and sector and wave fixed effects. The top and bottom one percent of dependent variables are trimmed. The results are robust or other (or no) trimming procedures and dropping any controls. Statistical significance at 0.10, 0.05, and 0.01 is denoted by *, **, and, ***.

Impact on Higher Profit Business Owner Finally, we consider the impact on more productive members of the match. The diffusion process in Section **??** assumes that there should be no gains to these individuals, which is implicitly assumed through the use of the max function in law of motion for productivity (Assumption **??**). These individuals were not randomly selected relevant to their peers, and thus cannot be directly compared to a control group as the others were. However, 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.

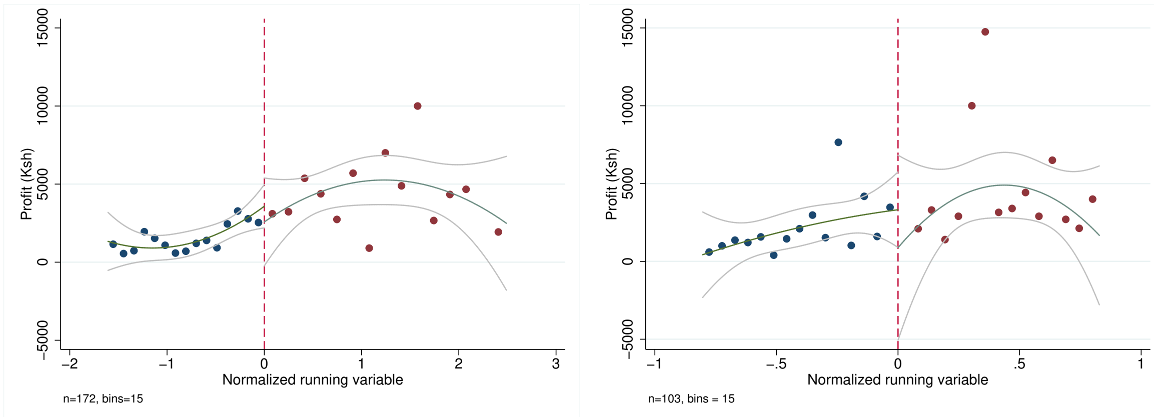
Figure **??** plots profit along with a fitted quadratic and its 95 percent confidence interval. Figure **??** uses the entire sample, while Figure **??** uses the **?** procedure to choose the optimal bandwidth. Both use 15 bins on either side of the cutoff. Figure **??** suggests no statistically discernible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

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

$$\pi_i = \alpha + \tau D_i + f(N_i) + \nu_i \tag{3.2}$$

Figure 4: Profit for mentors and non-mentors (from **?**)

(a) Full sample (b) IK Optimal Bandwidth



where π_i is profit, $D_i = 1$ if individual i was chosen as a mentor $(\widehat{\varepsilon}_i \geq \overline{\varepsilon}$, $f(N_i)$ is a flexible function of the normalized running variable $N_i = (\widehat{\varepsilon}_i - \overline{\varepsilon})/\sigma_\varepsilon$, 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 **??**, 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 **??**), which is assumed here and in much of the existing literature.

We emphasize that while this is consistent with the model described in Section **??**, where higher productivity firms receive no benefit from interaction with lower productivity firms, there is nothing in the experimental design that guarantees this outcome.[23]

[23] For example, the relationship may be consistent with a "collaboration" model in which both sides gain from interacting with the other. On the other hand, if the time requirement is high enough, there could be a cost to the more productive member of the match. However, the high take-up and persistence of matches suggests this second explanation is unlikely *a priori*.

Table 3: Regression discontinuity results for matched firm treatment effect (**?**)

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.

3.4 Diffusion Parameter Estimates

We use these results to estimate the parameters of the diffusion process. Relative to the theoretical results in Section **??**, we now have multiple periods of data. There are a number of ways to link the longer panel to the theory, and we choose a particular interpretation that allows us to use the dynamics as an "out of sample" test, allowing for some abuse of the term.

Specifically, we do the following. We pool the data into quarters, to create equally spaced time periods. The imitation opportunity (that is, the treatment) is in quarter 2. We use only quarters 1 and 2 to estimate the parameters (β, ρ, θ) from properly-defined moments detailed in Section **??**. Thus, we will exactly match the initial treatment effect. We do not, however, utilize any of the future data in our estimation. Instead, we will ask whether the model can predict the seemingly fast fade-out of the treatment effect.

Applying this strategy implies $\beta = 0.26$ and $\rho = 0.72$. Estimating θ requires specifying the functional form of $\widehat{M}$. In our quantitative results below, we assume

$$\widehat{M}(\hat{z}; z, \theta) = M(\hat{z})^{\frac{1}{1-\theta}}$$

where M is the equilibrium distribution of productivity in the economy. When $\theta = 0$, this is the usual uniform random matching assumption. $\theta > 0$ implies the imitation draws are concentrated among better firms. On the other hand, $\theta < 0$ implies draws are concentrated in the left tail of the equilibrium productivity distribution.[24] Applying our identification strategy with this function, we find that $\theta = -3.07$.

[24] More formally, $\theta \to 1$ 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. This 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 equilibrium.

4 Full Model

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

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 **??**.

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. In order to be a worker, an agent must pay f units of the final good in order to access the labor market.[25]

As before, an entrepreneur's profit is

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

where is now w the market clearing wage. Recursively, the value of having entrepreneurial skill z is

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

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

$$\begin{aligned} l(z) &= \left(\frac{1-\alpha}{w}\right)^{\frac{1}{\alpha}} z \\ \pi(z) &= \alpha \left(\frac{1-\alpha}{w}\right)^{\frac{1-\alpha}{\alpha}} z \end{aligned}$$

which satisfies Assumption **??** on the proportionality of profit and productivity. This

[25] This distortion is set simply 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.

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.[26]

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

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

In order to identify θ, we make a functional form assumption on $\widehat{M}$. A match $\hat{z}$ is drawn from

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

As discussed in the previous section, the parameter θ controls the part of the distribution from which draws arise, satisfying our Assumption **??**. In particular, when $\theta = 0$, imitation draws are uniform from the set of operating firms. However, we also allow for the possibility that draws may be better or worse than that. As θ grows toward 1, the mass in $\widehat{M}$ goes deeper into the right tail of operating firms, so that imitation draws are much better than random.[27] Alternatively, as θ goes to negative infinity, the probability mass converges to $\underline{z}$, so that no operating firm ever receives a useful imitation draw. This case captures the possibility that diffusion is completely absent.

Unlike our assumed $\widehat{M}$ in Section **??**, however, $\widehat{M}$ now depends on the economy-wide productivity distribution M, an equilibrium object. It must therefore be consistent with the the diffusion process in the economy. The law of motion for M is

$$\forall z', M'(z') = \delta G(z') + (1-\delta)\int_0^\infty \int_0^\infty F(\log(z') - \log(\max\{\hat{z}^\beta z^{\rho-\beta}, z^\rho\}) - c)\, d\widehat{M}(\hat{z};\theta)\, dM(z) \tag{4.5}$$

where G is the exogenous distribution from which new entrants draw productivity.[28]

4.1 Competitive Equilibrium

A competitive equilibrium of this economy is a wage w, a distribution of productivities M, and a value function v such that v satisfies (**??**) with the associated decision rules for labor and occupational choice, the evolution of M is consistent with the decision

[26] In Appendix **??** we embed our diffusion estimates in other models that break this result, for example by adding a utility benefit to firm operation.

[27] It is straightforward to micro-found θ. For example, if $1/(1-\theta) = K$ and K is a natural number, then this specification is equivalent to each agent receiving K draws each period and the imitation opportunity is the maximum of those K draws. See Appendix **??** for more details.

[28] Other papers, such as **?** and **?**, 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.

rules and is given by (**??**), the wage w clears the labor market, which requires a solution to the implicit equation

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

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 $\beta = 0.26$, $\rho = 0.72$, and $\theta = -3.07$.

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 $E[e^{\varepsilon}] = 1$.

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 f. We jointly match these four parameters to the following four moments: the standard deviation of log-profit in the overall population of operating firms (1.399), the variance of log-profit among new entrants (0.961), the ratio of the average profit of firms overall to the average profit of new entrants (1.557), and the fraction of agents that operate as workers (28.7%). These moments and parameter values are reported in Table **??**.

Table 4: Targets and Parameter Choices

Model Parameter	Description	Parameter Value	Target Moment	Source	Target Value	Model Value
Group 1						
β	Intensity of diffusion	0.26	Estimated parameter from regession (**??**)	RCT results	1.16	1.16
ρ	Persistence of productivity	0.72	Estimated parameter from regession (**??**)	RCT results	1.18	1.18
θ	Directedness of search	-3.07	Γ_3: Overall treatment effect in quarter 2	RCT results	1.40	1.40
Group 2						
σ	St. dev. of exogenous productivity shock distribution	1.435	Variance of log profit in all firms	Baseline survey	1.399	1.331
σ_0	St. dev. of new entrant productivity distribution	0.915	Variance of log profit among new entrants	Baseline survey	0.961	0.900
c	Growth factor in productivity evolution	0.539	Ratio of average profit of all firms to new entrants	Baseline survey	1.557	1.450
f	Fixed cost to access labor market	1.028	Fraction of agents employed as workers	**?**	0.287	0.288
Group 3						
δ	Death rate of firms	0.016	Average age of baseline business owners	Baseline survey	34	34

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

5 The Quantitative Impact of Diffusion

With the calibrated model in hand, we turn to the quantitative results. Our goal here is two-fold. First, we ask whether the model can replicate the time series of the empirical treatment effect, and we study the importance of our estimated parameters in generating this result. Second, we ask whether our model implies a large general equilibrium diffusion externality, and study its relationship to the partial equilibrium results from the first step.

5.1 Partial Equilibrium Effect

The experiment used to parameterize this model exhibited distinctive dynamics over the study period. In particular, the treatment effects were strong in the second quarter, and then decayed over time. After 4 quarters, the treatment effects were no longer detectable. We begin by asking whether the model can match this result.

5.1.1 Procedure

We study this question with the following experiment. We start the economy from the stationary equilibrium, defined by the aggregate state $M^*(z)$. We then create a control and treatment group with the same properties as our study group, including average and variance of profitability relative to the full economy.[29] Similarly, we construct a "mentor" group from M^* with the same properties as our empirical mentors. We then shock the treatment group with a one-period random draw from the mentor group. After that period, treatment firms continue to draw from M^*, while control group draws from M^* in every period.

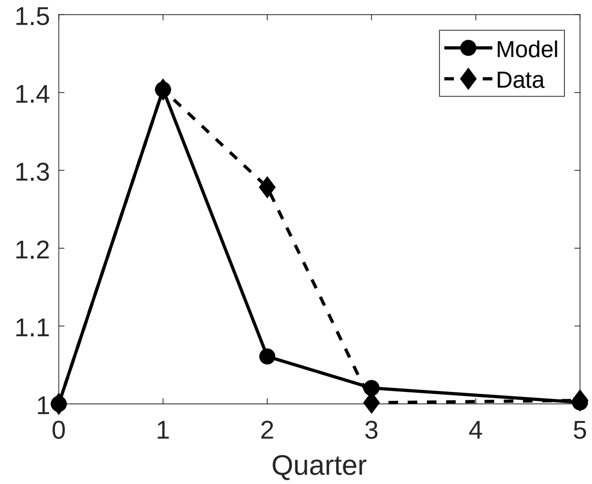
It is important to note that this is a "partial equilibrium" experiment, in the sense that the treatment does not influence the evolution of M (it remains fixed at M^*). Thus, if we observe convergence of treatment and control, it is not due to the fact that the control group is also benefiting from treatment.

5.1.2 Results

Figure **??** plots the implied dynamics of the treatment effect in the model and data. Note that the first two quarters (i.e., the pre-period and the first post-treatment period) are matched by construction. However, the model predicts the end of the effect quite well. By quarter 3, both the model and data predict no treatment effect, though the model under-predicts the effect in quarter 2.

[29] The age cutoff is irrelevant here, given the constant death probability in the model.

Figure 5: Dynamics of Treatment Effect



The key to this result is our relatively low estimate of persistence ($\rho = 0.72$), especially when compared to standard values in more developed countries of around 0.9 (e.g. **??**).[30] Thus, the impact of a one-time treatment degrades quickly despite not being targeted directly. If we instead calibrated ρ to minimize the sum of squared errors over the entire time series of Figure **??**, we would have estimated $\rho = 0.78$. Thus, the model matches the time series quite well without directly targeting it.

What drives this result? Figure **??** plots the treatment effect after 1 quarter (Figure **??**) and after 5 quarters (Figure **??**) for various combinations of β and ρ. It highlights the importance of different parameters for different moments of the PE impulse response observed in Figure **??**. Figure **??** shows that the initial spike in profit is primarily a function of β. That is, when a low productivity treatment firm meets a high productivity firm, the immediate response depends on how much how much of that firm's productivity gets immediately internalized.[31]

On the other hand, the half-life of this initial effect – and thus the "long-run" PE effect – is primarily controlled by ρ (Figure **??**). Given our estimated ρ, most values of β imply no long-run PE effect. However, a second fact arises here. For a given ρ, the long-run PE effect is *decreasing* in β. The intuition for this result is straightforward. If β is high, individuals internalize a large portion of the potential benefits from a good match. Over time, most agents meet with someone productive at least once, and thus the treatment effect is muted.

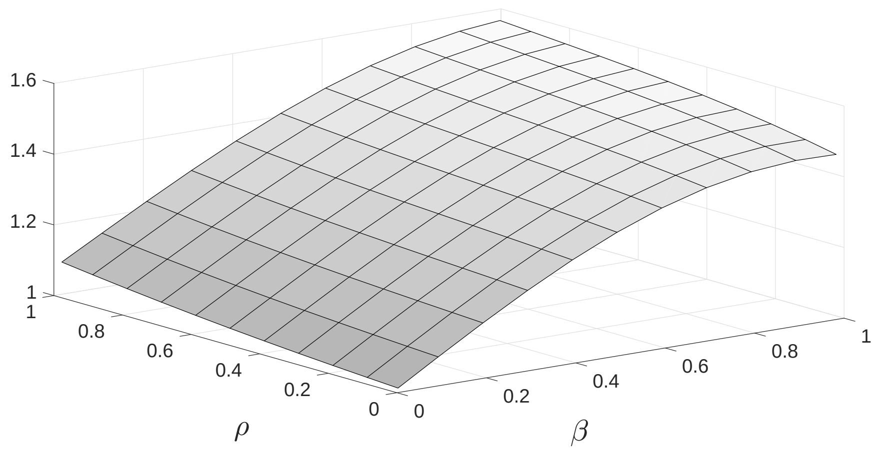
Could one use these results to deduce that diffusion is unimportant in an aggregate

[30] We discuss the rationale for low firm-level persistence in **?**.

[31] θ also matters for this result, as it governs how much better the match is relative to the average control match. We do not include this graph here to save space, as θ turns out to play on a small quantitative role once we consider general equilibrium effects.

Figure 6: Partial Equilibrium Treatment Effect

(a) After 1 Quarter



(b) After 5 Quarters

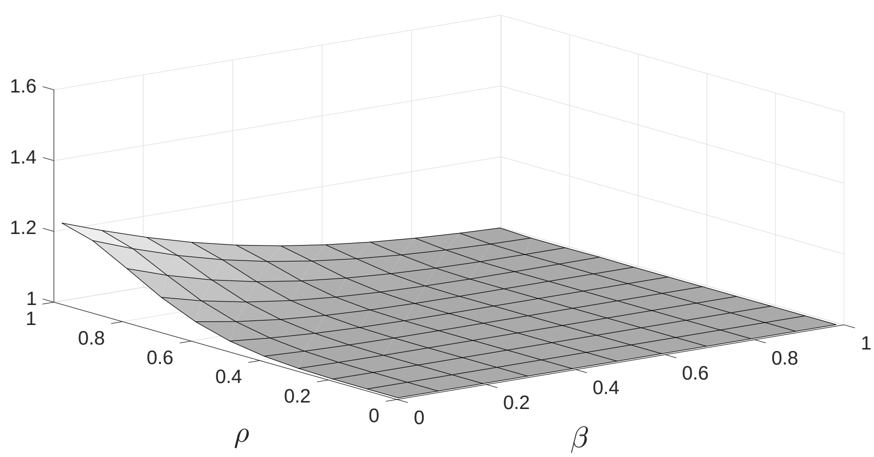


Figure notes: Average control group baseline profit is normalized to one, so a value equal to one here implies no treatment effect.

sense? Only on the condition that β is equally as unimportant for long-run general equilibrium effects as they are for long-run partial equilibrium effects. Figures **??** and **??** tell us nothing about that, and gives us no reason to make such an extrapolation. As we show in the next subsection, even more caution is warranted. Using the long-run PE effects as a measure of the aggregate importance of diffusion would in fact point in the *wrong* direction – while the partial equilibrium impact is decreasing in β, the general equilibrium effect is increasing in β.

5.2 General Equilibrium Effect

We next ask whether the general equilibrium importance of diffusion is large, even within a model that can match the quick fade of the partial equilibrium treatment effect. The equilibrium of this model is inefficient due to the negative externality generated by marginal firms. Because knowledge diffuses via random search (conditional on directedness parameter θ), marginal firms decrease the likelihood of an individual learning from the right tail of the knowledge distribution. The planner, therefore, would allocate marginal firms to instead operate as workers in order to increase the average match quality of firms. We measure the size of this equilibrium externality by solving for the optimal allocation of agents between workers and entrepreneurs and compare the stationary equilibrium when agents are assigned optimally to the *laissez faire* equilibrium.

5.2.1 Results

We solve for planner's allocation and compare it to the baseline economy. At our estimated parameters, we find that the GE diffusion externality is large. Despite the lack of any long-run partial equilibrium effect, aggregate income is 290 percent higher in the economy where workers are allocated optimally compared than when the allocation is decentralized. These gains are driven by an endogenous shift in the stationary distribution of productivities illustrated in Figure **??**. This large shift is generated by the feedback effects of increasing the average quality of $\hat{z}$ draws, which then causes the productivity distribution to shift to the right, improving the set of $\hat{z}$ draws even more.

5.2.2 Role of Estimated Parameters in Generating the GE Effect

Finally, we do a similar exericse to the partial equilibrium case, and study the importance of various parameters in generating the income gains highlighted above. Figure **??** shows the results graphically.

First, Figure **??** first shows that lower θ increases the gains from policy. This is because higher θ implies that individuals are more easily able to access high-z matches, limiting the impact of this policy designed to make marginal firms exit. However, the differences are relatively muted relative to the importance of β and ρ. Moving to $\theta = 0.25$ lowers the gains from our baseline 290 percent down to 215 percent.

Second, ρ plays a critical role, as was true in partial equilibrium as well. Holding the remaining parameterization fixed, the income gained by moving to the efficient allocation increases from 2 to 6328 percent when ρ increases from 0.2 to 0.95 (the

Figure 7: Stationary Productivity CDF

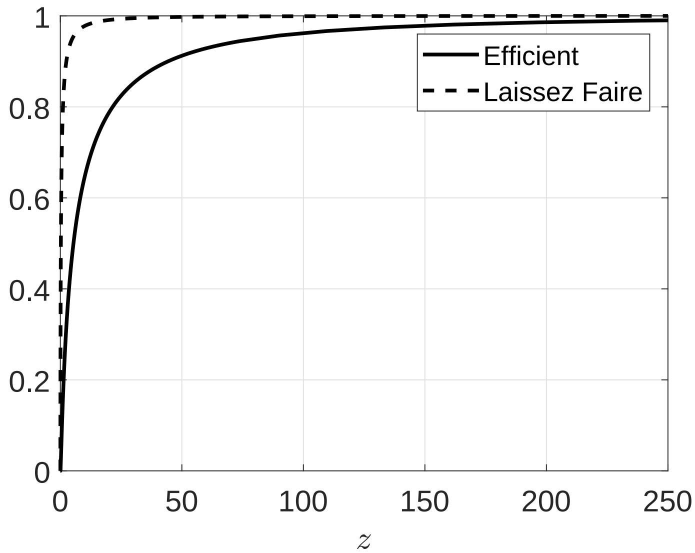


Figure 8: % Income Changes in Efficient Allocation

(a) Impact of θ

(b) Impact of β

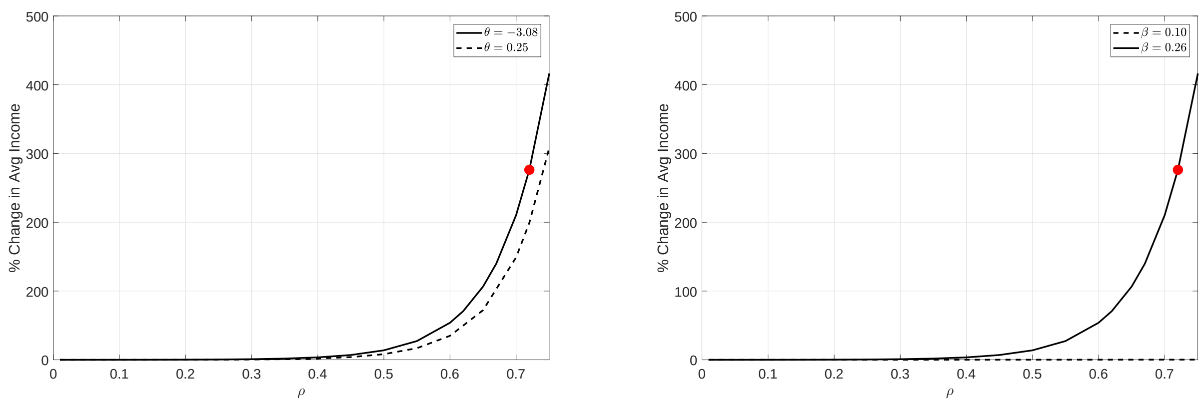


Figure notes: This figure plots the average income change in response to the optimal wage subsidy subsidy for various diffusion parameter combinations. Our estimated gains are indicated by the red dot on the solid line.

latter is not shown on the graphs for the sake of readability).

Finally, Figure **??** shows that the GE importance of diffusion is increasing in β. Decreasing β from 0.26 to 0.10 holding the remaining parameterization fixed causes the gains to plummet to from 290 percent to 2 percent. This contrasts sharply with the PE results. While the long-run GE impact is increasing in β, the long-run PE impact is decreasing in β.

This contrast with the partial equilibrium case is important. In both cases, the magnitudes rely on complementarity between β and ρ, but in different ways. The partial equilibrium magnitude is maximized with high ρ and low β. The GE effects, on the other hand, rely on high ρ and high β. For example, at $\rho = 0.2$ the gains decline from 2 percent to slightly below 1 percent when β decreases from 0.26 to 0.10, compared to the 290 to 2 percent drop at our estimated $\rho = 0.72$.

The intuition is straightforward. The ability to internalize productivity (high β) is worth relatively little from the planner's perspective if those gains die out quickly (low ρ). Thus, with low persistence, there is little to be gained regardless of β. Once the impact of a good match lasts for multiple periods, the planner has a larger role to play in managing the diffusion externality.

These results confirm the importance of exercising caution when extrapolating PE effects for GE interpretation. For example, imagine a scenario in which a policy-maker decided to run our experiment in various economies with different β, then use the long-run treatment effect to identify where diffusion played the largest role (that is, where the gains from optimal policy were largest). Using this metric, she would select economies with the lowest β. That is, she would instead minimize the potential welfare gains from policy.[32]

6 Discussion and Alternative Modeling Chocies

As we highlighted in Section **??**, the model laid out here is not the only one in which one could deploy our estimated diffusion parameters, nor the only diffusion process. We highlight the random search framework in the main body of the paper it underlies a large portion of the literature. However, the model does have its own idiosyncrasies that potentially matter for the quantitative results. In the Appendix, we re-create the results under two different scenarios.

In the first, we fix θ *ex ante* and allow for a cost of finding a match, similar to **?**.[33] As in the baseline model, θ plays little quantitative role, making this assumption relatively innocuous.

In the second, we depart from the baseline assumption that all surplus is captured by the less productive member of the match, and consider a bargaining problem when two agents meet as in **?**. In this context, we reinterpret θ as a measure of bargaining power.

[32]If economies differ in β and ρ, this would no longer be exactly true. In this case, the policy maker would choose economies with low β and high ρ. Quantitatively, however, the same results emerge. See Figure **??** and the surrounding discussion.

[33]In a model where search requires a cost, θ is not generically identified independent of the remaining model environment. See Section **??** for a longer discussion.

We note, however, that these are not the only models in which one can deploy these results. In Appendix **??**, we provide a variety of alternative assumptions that fall under the assumptions laid out in Section **??**, including models with multiple draws, draws that include noise, and deterministic assignment.

7 Conclusion

This paper uses evidence from a randomized controlled trial to identify a model of firm-to-firm productivity transmission . Our results imply an important role for diffusion. The efficient level of the learning externality increases income substantially. We emphasize that this need not be true, as both the model and experimental design *ex ante* allow for the possibility that the impact of diffusion is small or nonexistent. Moreover, our results show that the critical parameters for generating the equilibrium importance of diffusion are the same t

We view these results as an important first step that highlights the possibilities of linking equilibrium diffusion models with causal identification. The next steps require a more detailed investigation of 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 and low costs we observe (**?**). Our model builds this in as a technological constraint, but that need not be the case. **?** point to frictions in the information market, while **?** point out that growth-reducing network structures can be an optimal response to the possibility of detrimental flows through the network (e.g., disease). Put differently, here we assume a diffusion process assumed by much of the recent literature (though generalized somewhat), and estimate it. The next steps require distinguishing across diffusion processes, including any potential distortions that arise. Different field experiments, designed with an eye toward aggregate theory, could provide more detailed information to help further refine such model choices.

Table of Contents for Online Appendix

A Alternative Models [to be completed]

In this section, we replicate the results in a number of different models. After, we provide additional interpretations of models that fall under our assumptions (without the quantitative results).

A.1 Cost of Searching

A.2 Bargaining Problem During a Match

A.3 Assignment

A.4 Additional Diffusion Processes

A.4.1 Model of 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.4.2 Model with Effort Choice

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.4.3 Innovations through Imitation

Here we show how the **?** 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.5 Model with 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:

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

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 **??** 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.[34] 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

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

this equation. Thus, (**??**) pins down β as a function of ρ. We therefore need a second moment to separate them.[35]

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 **??**, 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*

[35]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 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.

*(β^*, ρ^*) that solve equations (**??**) and (**??**). 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.

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 Additional Results

C.1 Allowing for Idiosyncratic Distortions

[to be completed]