Bringing Data to the Model: Quantitative Implications of an Equilibrium Diffusion Model*

Wyatt Brooks
University of Notre Dame

Kevin Donovan
Yale University

Terence R. Johnson
University of Notre Dame

February 2019

Abstract

We consider an increasingly utilized class of general equilibrium models in which knowledge diffusion generates positive spillovers across firms. Each period a firm "matches" with another firm, and can internalize some portion of the matched firm's productivity. Within this class of models, we prove that a small set of parameters characterizing the diffusion process are uniquely identified with exogenous and random variation in matches, and moreover, are independent of other parameter values and remaining model structure. We then provide an application of our results in Kenya. We conduct a randomized controlled trial in which firms from the left tail of the profit distribution are matched one-to-one with firms from the right tail, and use the empirical results to estimate these diffusion parameters. We embed the estimated process into a general equilibrium model to assess the quantitative importance of diffusion. The efficient level of learning spillovers substantially increases average income relative to a *laissez faire* policy.

*Thanks to conference and seminar participants at Arizona State, Columbia, Illinois, Notre Dame, NYU, Stanford, UBC, the Society for Economic Dynamics, and Y-RISE for helpful comments, especially 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 (Lucas, 2009; Lucas and Moll, 2014; Perla and Tonetti, 2014; Perla et al., 2015; Buera and Oberfield, 2017). 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.[1]

The standard approach to deal with this issue is to impose structure. Assumptions on the economic environment (e.g., distributions of shocks, how firms enter and exit, details of occupational choices, etc.) allow a mapping between theoretical parameters and empirical moments. 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, relying on randomized variation in "matches" between agents. We apply our procedure with a randomized controlled trial in Kenya, then measure the quantitative importance of diffusion by building the results into a general equilibrium model of diffusion.

Intuitively, a full accounting of diffusion requires taking into account both the static (how much one's own productivity changes in response to a match) and dynamic effects (the propagation of that static effect to other economic agents). Under some commonly used assumptions detailed in Section 2, we show how randomized variation in matches can be used to measure both. Specifically, we guarantee each treatment firm a randomly-selected (and therefore observable) match from the right tail of the productivity distribution. The static effect is then identified using within-treatment variation. Because of the randomized variation within the treatment, we can identify the immediate impact of match by comparing differences in *ex post* profit between two groups with different average match productivities. The dynamic effect then requires two separate pieces. The first is the persistence of these static gains, which allows firms to diffuse gains from past meetings when they interact with new firms in the future. We identify persistence with a standard lagged regression used to identify the persistence of an exogenous AR(1) process. However, in our context, this estimate includes a diffusion-induced bias term. We can correct this bias directly by utilizing

[1] In contrast, a large body of work in micro-development tracks diffusion of particular pieces of information or technology. Conley and Udry (2010) study the introduction pineapples in Ghana, Banerjee et al. (2013) study the diffusion of microfinance in India, and BenYishay and Mobarak (forthcoming) study the diffusion of pit planting in Malawi. This more narrow focus has the clear benefit of allowing one to track diffusion explicitly along the dimension of interest. Our goal here is focused on the set of models interested in the diffusion of *productivity*, a broader notion that does not admit such explicit tracking.

the observable matches within the treatment, thus guaranteeing that the methodology remains valid in the presence of diffusion. Finally, even conditional on the persistence, the dynamic effect potentially depends on who agents interact with. For example, if agents are relatively successful at directing meetings toward the most productive members of the economy, the dynamic effect will be muted relative to a world in which agents randomly select an individual to match with. To identify this force, we take a step back and utilize the average treatment effect between treatment and control. While we cannot observe individual control matches, small *ex post* differences in profit between treatment and control suggest that control firms are already quite good at "directing" their search toward to the right tail of the distribution (recall, the treatment is a guaranteed good match).

We formalize this intuitive argument in Section 2, and show that our randomized variation provides a unique mapping between the data and the model parameters governing the forces discussed in the previous paragraph. A useful additional implication of our procedure is that these diffusion 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 2 by only laying out the assumptions required for identification, leaving the details of the full model for Section 4, when we require them for the quantitative results.

The independence of the diffusion parameters from the remaining economic environment 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, making identification difficult without substantial structural assumptions that allow a mapping from parameters to empirical moments. In both cases, properly differencing from a control group helps limit many of the potential confounding factors. Our estimates can therefore be embedded into a variety of models with different assumptions on the remaining structure of the economy, and do not require us to take a stand on the full economic environment when estimating the diffusion parameters.[2]

[2]It is worth emphasizing that the goal of this paper is not provide a test that differentiates between diffusion models, and indeed our experimental results are not suitable for such a task. We aim to provide sufficient variation that identifies key parameters under a set of common model assumptions used in the literature, while highlighting the

We then provide an application of this procedure. We conduct a randomized controlled trial in Nairobi, Kenya in which low-profit treatment firms are matched one-to-one with a randomly selected high-profit firms in the economy. This experiment generates the proper variation to identify the model. The full set of reduced form results are available in our companion paper Brooks et al. (2018). Profits are 19 percent higher in the treatment group relative to the control.[3] This treatment impact is increasing in the relative profit gap between the two firms. Lastly, 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.

With the identified parameters in hand, we turn to quantifying the importance of diffusion. Measuring the importance of diffusion at our estimated parameters requires a fully-specified model, as the total diffusion impact requires solving the fixed point problem induced by this dynamic process. We therefore build our diffusion parameters into a dynamic general equilibrium model with occupational choice and market clearing wages, where diffusion occurs via random search. In this context, low productivity firms provide a negative externality by congesting the diffusion process. Thus, we measure the importance of diffusion by asking how much higher welfare is when the planner sets this externality optimally (via a wage subsidy) compared to a *laissez faire* policy.[4]

We find that average income is substantially higher in the planner's allocation than the *laissez faire* allocation, generating a 2.9 times increase in average income. The key force behind result is the complementarity between the static and dynamic forces. Even if an individual agent gains substantially from meeting with a highly productive agent, the equilibrium impact of this single meeting is large only when other agents gain from that match via future meetings. These two forces thus complement each other in generating a high equilibrium impact of diffusion.

1.1 Related Literature

This paper joins a relatively new literature that uses causal empirical estimates to identify critical model parameters in macroeconomic models. Kaboski and Townsend (2011) use exogenous variation in micro-loans to structurally estimate a model of borrowing constraints and entrepreneurship. Brooks and Donovan (2017) use exogenous

flexibility it provides when laying out the remaining economic environment. Given that, it is not particularly difficult to come up with models that do not satisfy our assumptions. We take this up in Section 6, where we lay out alternative models in which our results fail, then show which additional assumptions guarantee our procedure holds. In all such cases, our procedure generates a substantial reduction in structure required to identify diffusion parameters.

[3] We consider a number of possibilities in Brooks et al. (2018), including profit sharing, loans, and bulk discounts, and find that none of them explain the results.

[4] Naturally, the exact policy lever utilized here depends on the model structure.

variation in infrastructure placement to identify a model of risky farm investment. Our paper shares a similar style but focuses on knowledge diffusion. Closest in this dimension is Lagakos et al. (2018), 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.

Our work therefore adds empirical evidence to the mainly theoretical literature studying innovations and knowledge in general equilibrium models (Romer, 1986; Kortum, 1997). Most closely related to this paper is the more recent literature building on these papers, in which diffusion is modeled as a stochastic process of "imitation," including Jovanovic and Rob (1989), Lucas (2009), Alvarez et al. (2008), Lucas and Moll (2014), and Perla and Tonetti (2014). Recent work has also extended these models to consider within and across firm diffusion (Herkenhoff et al., 2018) and international trade (Perla et al., 2015; Buera and Oberfield, 2017).

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 (Foster and Rosenzweig, 1995; Conley and Udry, 2010), specific planting or production techniques (Atkin et al., 2017a,b; BenYishay and Mobarak, forthcoming), or financial information (Banerjee et al., 2013; Cai and Szeidl, 2018). The broadness of the more "macro" measure of productivity highlighted here makes it less feasible to track diffusion via a single piece of information or technology. Our goal is to utilize tools from this more micro literature, combined with theory, to identify and quantify the importance of equilibrium diffusion models. We emphasize that the quantitative importance of diffusion is in no way guaranteed by our field experiment, and is entirely a function of the magnitude of the empirical results.

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 $\hat{z}$.[5] The functional form of the subsequent productivity z' is given by Assumption 1.

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

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

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 Jovanovic and Rob, 1989, for example). We address this issue directly in Section 3 and find no evidence that more productive firms gain profit from interaction with less productive firms.[6]

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

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

This assumption is satisfied by much of the literature on diffusion. A simple way to satisfy Assumption 2 is to assume $\pi_i = z_i$ as in Lucas (2009) and Perla and Tonetti

[5] 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.

[6] 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.

(2014). A production function of the form $y = z^{\alpha} n^{1-\alpha}$, where n is labor, also satisfies Assumption 2 in a competitive labor market.[7]

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

Assumption 3. *The imitation opportunity $\hat{z}$ is drawn by a firm with productivity z from a distribution characterized by the cumulative density function $\widehat{M}(\hat{z}; z, \theta)$, a known function. For every z and $\hat{z}$, $\widehat{M}$ is continuous in θ and $\theta_1 < \theta_2 \implies \widehat{M}(\hat{z}; z, \theta_2)$ first order 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 3 allows us to be somewhat broader, as agents may draw from better or worse distributions that the set of operating firms, where θ indexes how much the distribution of matches differs from the firm productivity distribution. We discuss this assumption in more detail in Section 2.3.

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.

All of the assumptions we have made are common in the literature, including Lucas (2009), Perla and Tonetti (2014); Perla et al. (2015), and Buera and Oberfield (2017), among others, and in some instances allow us to generalize. Note that 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

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

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 2.3 and discuss the limits of the structural assumptions made above to hopefully provide some broader context for our results.

2.2 Variation Required to Identify Diffusion

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

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

1. *Agents in H_T and H_C draw their ε shocks from the same distributions*
2. *The matches for agents in H_C are not observable, and distributed $\widehat{M}(\hat{z}; z, \theta)$*
3. *The matches for agents in H_T are observable, and distributed $\widehat{H}_T(\hat{z},) \neq \widehat{M}(\hat{z}; z, \theta)$. Moreover, every match $\hat{z}$ is greater than the z to which it is matched.*
4. *For any arbitrary partition of the treatment group, characterized by $H_T^1(z)$ and $H_T^2(z)$, agents in both groups draw their ε shocks from the same distribution*

The first assumption imposes the usual exclusion restriction – that unobserved characteristics do not systematically vary across 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.[8] Finally, the third and fourth lay out what we require from

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

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.[9] Finally, the last assumption states a second exclusion restriction *within* the treatment group. We can break the treatment group into two arbitrary groups, and comparisons across these two groups are unbiased.[10]

Our procedure works as follows. Using only the treatment firm data, we show how to identify β uniquely. We then continue with the treatment firm data and show that ρ is uniquely identified as a function of β. Finally, we take a step back and utilize both treatment and control data to show that θ can be identified as only a function of (β, ρ). Thus, the three parameters are uniquely identified under the assumptions laid out in the previous sections.

Intensity of Diffusion, β The idea behind identifying β is easiest explained by first considering two agents with the same level of productivity z, but one receives a high $\hat{z}$ match and the other a low productivity match. If (for some reason) they both received the same exogenous shock ε, then any *ex post* profit difference is solely driven by variation in match quality. A large *ex post* profit gap therefore implies agents internalize a large portion of match productivity (high β), while a smaller one would suggest a lower β. Of course, there is no reason for ε draws to be the same across two individual agents. Repeating this exercise with a large number of agents, however, allows the exogenous shocks to average out across the two groups and thus allows us to appeal to the same intuition.

To formalize this idea, partition treatment agents into two subsets labeled "low" and "high," denoted $H_T^L(z) \subset H_T(z)$ and $H_T^H(z) \subset H_T(z)$. The imitation draws are drawn from similarly denoted $\widehat{H}_T^L(\hat{z})$ and $\widehat{H}_T^H(\hat{z})$. These $\widehat{H}$ distributions are constructed to formalize the idea of "high draw" agents and "low draw" agents, and thus $\forall \hat{z}, z, \widehat{H}_T^L(\hat{z}) > \widehat{H}_T^H(\hat{z})$.[11]

Now, as discussed intuitively above, we compare *ex post* productivity of the two groups. To that end, denote average productivity of the high-treated firms as $\bar{z}_T^H$ and low-treated firms was $\bar{z}_T^L$. Plugging this into the law of motion for productivity

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

[10] Note that the assumption does not necessarily require randomization. A valid instrument would satisfy the same requirement.

[11] For example, one could define $\widehat{H}_T^H$ as all matches with higher than median values of $\hat{z}$, though any partition satisfying this restriction is sufficient.

(Assumption 1) implies

$$\frac{\bar{z}_T^H}{\bar{z}_T^L} = \frac{\int\int\int e^{c+\varepsilon} z^\rho \max\left[1, \hat{z}/z\right]^\beta dF(\varepsilon) d\widehat{H}_T^H(\hat{z}) dH_T^H(z)}{\int\int\int e^{c+\varepsilon} z^\rho \max\left[1, \hat{z}/z\right]^\beta dF(\varepsilon) d\widehat{H}_T^L(\hat{z}) dH_T^L(z)}.$$

Utilizing Assumptions 2 ($\pi \propto z$) and the assumption that $\hat{z} > z$ in the treatment (Assumption 4) this becomes

$$\Gamma_1 := \frac{z'^H}{z'^L} = \frac{\int \hat{z}^\beta d\widehat{H}_T^H(\hat{z})}{\int \hat{z}^\beta d\widehat{H}_T^L(\hat{z})} \tag{2.2}$$

where the exogenous shocks ε cancel out via the within-treatment exclusion restriction in Assumption 4. The left-hand side of (2.2) is simply a heterogeneous treatment effect, and thus observable. Thus, (2.2) has a unique solution in β, as $\hat{H}_T^H(\hat{z}) < H_T^L(\hat{z})$ for every z. This identification procedure for β is formalized in Proposition 1, with one caveat relative to the intuitive discussion.

Proposition 1. *There exists a unique β that solves (2.2) if*

$$\Gamma_1 \in \left[1, \frac{\int\int \hat{z} d\widehat{H}_T^H(\hat{z}) dH_T^H(z)}{\int\int \hat{z} d\widehat{H}_T^L(\hat{z}) dH_T^L(z)}\right]$$

The condition in Proposition 1 essentially requires the effect to be within the realm of possible outcomes defined by the productivity law of motion in Assumption 1. Recall that Γ_1 is the relative impact on those who get a high-productivity compared to a low-productivity match. For example, the requirement $\Gamma_1 \geq 1$ simply states that better imitation opportunities do not lead to worse firm performance. If that were the case, clearly no parametrization of β could match that result given the assumed law of motion. A useful aspect of this requirement, however, is that it is testable in the data. Looking ahead to our application, we verify the relevant treatment effect falls in this interval. Thus, Γ_1 uniquely identifies β.[12]

Persistence of Productivity, ρ The second key parameter is the persistence of productivity, ρ. In a model without diffusion, this amounts to estimating the persistence of an exogenous AR(1) process. Our strategy is similar, with one important difference: in the presence of diffusion, the imitation opportunity must be accounted for. A standard lagged regression would treat both the exogenous shock and the imitation opportunity as part of the error term. Because the imitation term explicitly varies with z (due

[12] If we eliminated the part of Assumption 4 that requires $\hat{z} > z$ in the treatment, this result no longer holds. Instead of a unique value of β, we would instead need to prove uniqueness of a joint (β, ρ) that solves two moment conditions. This complicates the math, but still goes through with identical conditions. See Appendix B for more details.

to the max operator), this technique is biased. We therefore utilize treatment group data where imitation opportunities are observed. Because they are observed, we can control for them directly and thus eliminate the bias.

To operationalize this insight, we define the "normalized covariance" of pre- and post-intervention profit as our second moment, given by

$$\Gamma_2 := \frac{Cov[z,z']}{E[z]E[z']} = \frac{\int\int\int e^{c+\varepsilon} z^{1+\rho} \max\left[1, \frac{\hat{z}}{z}\right]^{\beta} dF(\varepsilon) d\widehat{H}_T(\hat{z},z) dH_T(z)}{\int z dH_T(z) \cdot \int\int\int e^{c+\varepsilon} z^{\rho} \max\left[1, \frac{\hat{z}}{z}\right]^{\beta} dF(\varepsilon) d\widehat{H}_T(\hat{z},z) dH_T(z)} - 1.$$

This second moment, Γ_2, is an adjusted measure of a standard AR(1) regression and is thus observable.[13] Some algebra on right-hand side of this equation gives

$$\Gamma_2 = \frac{\int\int z^{1+\rho} \max\left[1, \frac{\hat{z}}{z}\right]^{\beta} d\widehat{H}_T(\hat{z},z) dH_T(z)}{\int z dH_T(z) \cdot \int\int z^{\rho} \max\left[1, \frac{\hat{z}}{z}\right]^{\beta} d\widehat{H}_T(\hat{z},z) dH_T(z)}. \tag{2.3}$$

Together with our previously-estimated β and the observability assumptions in Assumption 4, the only unknown in (2.3) is ρ, and it is straightforward to show that the solution is unique.

Proposition 2. *Given a value of β, there exists a unique ρ that satisfies (2.3) if*

$$\Gamma_2 \in \left[1, 1 + CV(z)^2\right]$$

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

"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. As we lay out in Assumption 4, we admit from the outset that we cannot observe individual-level matches in the control group. The critical insight here is that we can observe draws from the treatment group, and can use them to infer the average quality of draws in the control group by comparing the average outcomes. Take, for example, the case in which treatment draws are guaranteed to be high productivity. If we observe small differences in average *ex post* profit, it follows 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

[13] That is, if we run the panel regression $\pi_{i,t} = \eta + \gamma\pi_{i,t-1} + \nu$ to generate coefficient estimate $\widehat{\gamma}^{OLS}$, then our moment is equivalent to

$$\Gamma_2 = 1 + \frac{Var(\pi)}{\mathbb{E}[\pi]\mathbb{E}[\pi']}\widehat{\gamma}^{OLS}.$$

where we again utilize the assumption that $\pi \propto z$.

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 implies that control group firms typically receive low productivity imitation opportunities, corresponding to a low value of θ. This average difference in profit therefore allows us to infer what θ is, despite not observing the underlying matches in the control group.[14]

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

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

Defining $\Gamma_3 := \bar{z}_T/\bar{z}_C$ and 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 come directly from the data (after applying Assumption 2), except for the parameter θ.[15] The assumed monotonicity of $\widehat{M}$ (Assumption 3) is sufficient to prove that any θ that solves this equation is unique. Proposition 3 formalizes the results, again developing bounds to guarantee the results.

Proposition 3. *Given the values (β, ρ), the value of θ that solves (2.4) is unique if $\Gamma_3 \in [\Gamma_3^{min}, \Gamma_3^{max}]$, where*

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

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

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

[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.

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

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. We require only a measure of treatment heterogeneity (Γ_1 to identify β), the (normalized) coefficient from a lagged profit regression (Γ_2 to identify ρ) and the average treatment effect (Γ_3 to identify θ). 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 required 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 1 and 2).

A more subtle restriction is built into our assumption on $\widehat{M}$ in Assumption 3. Here, we require that the draw of a match $\hat{z}$ depends only on z and a parameter θ. This assumption nests as a special case work by Jovanovic and Rob (1989), Lucas (2009), and Buera and Oberfield (2017), who assume uniform draws from the existing distribution of operating firms. In that case, if M is the cdf of operating firm productivity, $\widehat{M}(\hat{z}; z, \theta) \equiv M(\hat{z})$.[16] Lucas and Moll (2014) and Perla and Tonetti (2014) make the same assumption, but extend these models by endogenizing the cost of receiving a draw. These models generally fail Assumption 3, because the decision to search depends on the remaining details of the model and equilibrium. In this case, we lose the independence of θ from the remaining model structure.[17]

First, we note that Assumption 3 still allows for a wide set of underlying matching 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 A.

Two other features about models that fail Assumption 3 are worth emphasizing. Fist, even in these cases, the identification of (β, ρ) still holds. Thus, if one was willing to accept a fixed θ (as in Lucas and Moll, 2014 and Perla and Tonetti, 2014), the remaining identification goes through unchanged.[18] Second, additional assumptions

[16] Of course, M is generally an equilibrium object. That fact is irrelevant for the identification, but matters more when we embed our estimates in a general equilibrium model in Section 4.

[17] We can, in principal, make assumptions such that the equilibrium properties of the search decision satisfy Assumption 3. The value of θ will not generically be independent of the remaining model structure, however.

[18] In our quantitative exercise, in which we utilize a random search framework, we find that varying θ has little

can guarantee our procedure still holds. In the cases outlined above, a sufficient assumption is a stationary equilibrium, which guarantees that the value of search remains stationary. Thus, even in these cases, the amount of structure required to identify the diffusion parameters is dramatically reduced.

3 Application to Kenyan Firms

Armed with the identification results, 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 our companion paper Brooks et al. (2018), though we reproduce some of the relevant results here for simplicity's sake.

Our experimental design randomly matches older, profitable entrepreneurs with younger entrepreneurs. The younger owners were then followed for over 17 months to measure changes in business practice and profit over time. Outcomes are compared to a control group of similar firms.[19]

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 pilot 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.

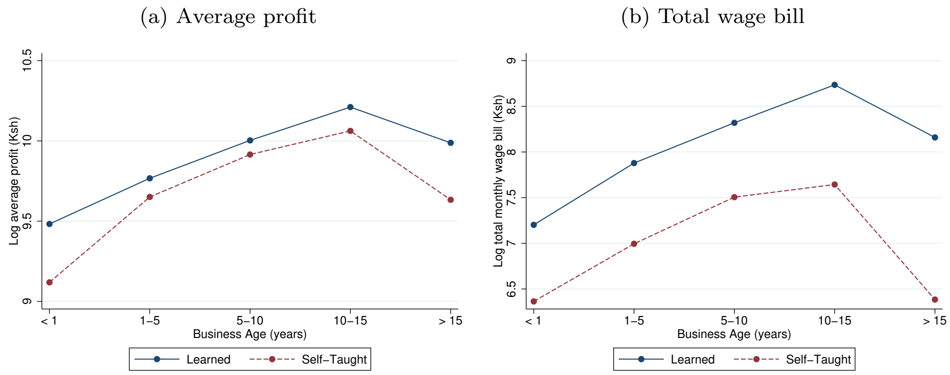
Qualitative Evidence on the Importance of Learning To begin, Figure 1 plots business scale measures based on self-reported learning methods from the baseline survey. Fifty-five percent of all firms claimed they were self-taught, while the rest claimed to learn either from another business operator, in school, or through an apprenticeship. Figure 1a shows that the self-taught earn less profit at any point over the lifecycle. The

impact on the results. Thus, at least in this context, fixing θ *ex ante* is of little importance.

[19]In Brooks et al. (2018), we further randomize another group into formal business training classes. While we do not utilize this classroom training treatment arm here, it is interesting to note that the results differ substantially across these treatment arms. We show that this to the fact that matching with local firms provides specific information about the local economy (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.).

average profit of a self-taught firm is 18 percent less than firms that learn from others, while Figure 1b show that self-taught firms pay a smaller total wage bill.

Figure 1: Self-Reported Learning Methods and Business Scale



Selection and Randomization We start from a sample of female business owners who have been in operation for less than 5 years.[20] 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 $t = 1, 2, 3, 4, 7, 12, 17$.[21,22]

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 2 plots the cumulative distribution function of baseline profit for the entire sample, the population we study, and the selected matches. One can see

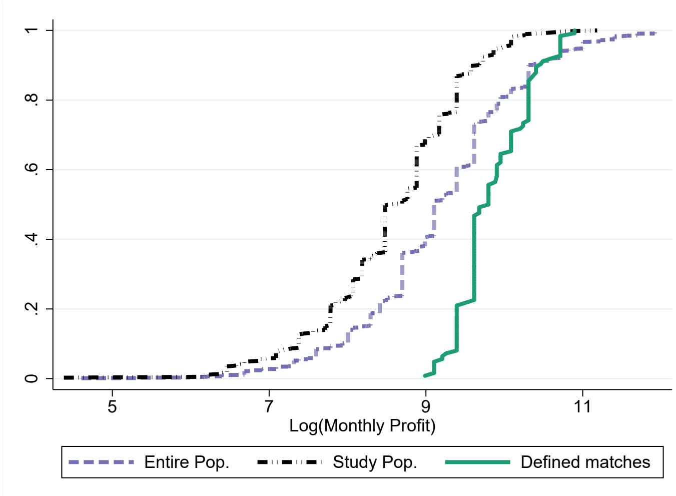
[20]The sex selection criteria is to limit heterogeneity outside the model. Note, however, that females make up 65 percent of business owners in Dandora and 71 owners with businesses open less than 5 years.

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

[22]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.

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 2: 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.[23]

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.

For simplicity and ease of reference to the more detailed discussion in Brooks et al. (2018), we refer to these two groups as mentees and mentors throughout. We emphasize, however, that they should more generally be thought of as the more and less productive members of a match.

[23]Even during the month-long treatment, they were encouraged to meet, but there was explicitly no penalty for not meeting.

3.2 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

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

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 1.

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 1, 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.

Table 1: Profit Effects in Randomized Controlled Trial (from Brooks et al., 2018)

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, ***.

As discussed in detail in Brooks et al. (2018), we found that the mechanism for this increase in profit was reduced costs, such as lower prices for inputs. This is important because greater 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.

Impact on Higher Profit Business Owner Finally, we consider the impact on more productive members of the match. The diffusion process in Section 2 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 1). 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 3 plots profit along with a fitted quadratic and its 95 percent confidence interval. Figure 3a uses the entire sample, while Figure 3b uses the Imbens and Kalyanaraman (2012) procedure to choose the optimal bandwidth. Both use 15 bins on either side of the cutoff. Figure 3 suggests no statistically discernible discontinuity around the cutoff.

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

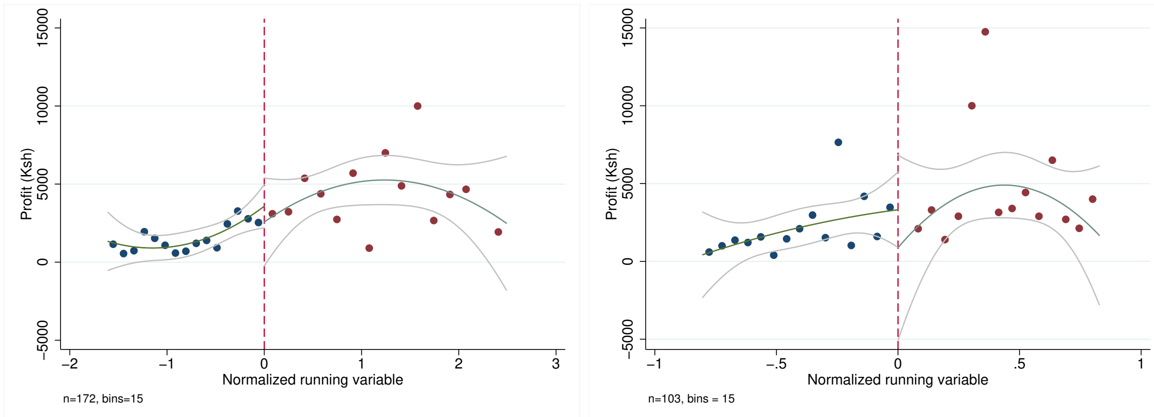
$$\pi_i = \alpha + \tau D_i + f(N_i) + \nu_i \tag{3.2}$$

where π_i is profit, $D_i = 1$ if individual i was chosen as a mentor ($\widehat{\varepsilon_i} \geq \bar{\varepsilon}$, $f(N_i)$ is a flexible function of the normalized running variable $N_i = (\widehat{\varepsilon_i} - \bar{\varepsilon})/\sigma_\varepsilon$, and ν_i is the error term. The parameter τ captures the causal impact of being chosen as a mentor. We use local linear regressions to estimate the treatment effects on profit and inventory, along with business practices of record keeping and marketing. The results are in Table 2, 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

Figure 3: Profit for mentors and non-mentors (from Brooks et al., 2018)

(a) Full sample

(b) IK Optimal Bandwidth



spending decreases, but it cannot be statistically distinguished from zero. Overall, we find little evidence that entering into a match changes either business scale or business practices for the more productive member of the match. This is consistent with the max function in the forward equation for productivity (equation 2.1), which is assumed here and in much of the existing literature.

Table 2: Regression discontinuity results for matched firm treatment effect (Brooks et al., 2018)

Percent of IK	Scale		Practices	
optimal bandwidth	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.

We emphasize that while this is consistent with the model described in Section 2, where higher productivity firms receive no benefit from interaction with lower productivity firms, there is nothing in the experimental design that guarantees this outcome.[24]

[24]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*.

3.3 Diffusion Parameter Estimates

We use these results to estimate the parameters of the diffusion process. Relative to the theoretical results in Section 2, 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) in quarter 2. We use only quarters 1 and 2 to estimate the parameters (β, ρ, θ) from properly-defined moments $(\Gamma_1, \Gamma_2, \Gamma_3)$. Thus, we will exactly match the initial treatment effect. We do not, however, utilize any of the future data in our estimation. Instead, we 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.[25] Applying our identification strategy with this function, we find that $\theta = -3.07$.

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

[25] 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.

are born. Each agent is characterized by productivity z which evolves over time via the diffusion process laid out in Assumption 1.

Occupational Choice and Recursive Formulation In every period, each agent can choose to be a worker or an entrepreneur. Workers sell their labor to entrepreneurs for the market clearing wage w, while entrepreneurs produce an undifferentiated consumption good using their skill and hired labor. In order to be a worker, an agent must pay f units of the final good in order to access the labor market.[26]

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 2 on the proportionality of profit and productivity. This further implies that agents face a cutoff rule to determine their occupation. For a given wage w, there is a $\underline{z}(w)$ such that any agent with $z < \underline{z}$ becomes a worker, while agents with $z \geq \underline{z}$ become entrepreneurs.[27]

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

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

[26] 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.

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

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 3. 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.[28] 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 2, 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.[29]

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 (4.2) with the associated decision rules for labor and occupational choice, the evolution of M is consistent with the decision rules and is given by (4.5), the wage w clears the labor market, which requires a solution to the implicit equation

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

[28] 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 A for more details.

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

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 3.

Table 3: Targets and Parameter Choices

Model Parameter	Description	Parameter Value	Target Moment	Source	Target Value	Model Value
Group 1						
β	Intensity of diffusion	0.26	Γ_1: Treatment effect from top quartile	RCT results	1.16	1.16
ρ	Persistence of productivity	0.72	Γ_2: Normalized profit covariance in treatment group	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	Gollin (2008)	0.287	0.288
Group 3						
δ	Death rate of firms	0.016	Average age of baseline business owners	Baseline survey	34	34

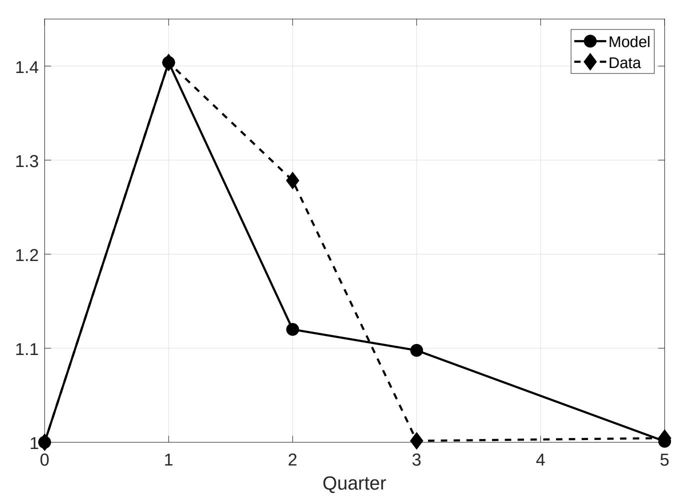
Table notes: Group 1 is jointly chosen from the experimental data. Parameters in Group 2 are calibrated to jointly match a number of moments from our baseline data. Group 3 are also set to match baseline data moments, but match 1-1 with target moments.

4.3 Dynamics of Treatment Effects

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.

As mentioned previously, we use the dynamics as as a test of the calibrated model. Figure 4 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 5, both the model and data predict no treatment effect, though the model slightly under-predicts the effect in quarter 2 while slightly over-predicting the effect in quarter 3.[30]

Figure 4: Dynamics of Treatment Effect



Intuitively, in an environment where persistence is low, then an intervention that provides a positive shock to profit will see its treatment effect degrade quickly. Our estimates imply low persistence ($\rho = 0.72$), especially when compared to standard values in more developed countries of around 0.9 (e.g. Midrigan and Xu, 2014). Therefore, although the model is not parameterized to match the rapid fade-out of effects, we have a separate piece of information (low persistence) that produces rapid fade-out endogenously.[31]

[30] A simple way to rectify this discrepancy to include aggregate shocks. The overall treatment effect co-moves positively with the level of the control profit in the data, and including such a shock helps along this dimension.

[31] For comparison, instead of employing the identification strategy discussed in this paper, we could calibrate ρ to maximize fit to the time series of effects. That exercise implies a value of $\rho = 0.78$ compared to $\rho = 0.72$ when parameterized as described in this paper.

5 Measuring the Importance of Diffusion

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. To study this result, we solve 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.

Fundamentally, the distinction between the efficient and the *laissez faire* allocation is what determines the productivity cutoff between operating as a firm or worker. In the efficient allocation, the cutoff is chosen to maximize aggregate income, as the planner internalizes the effect that the composition of operating firms has on the distribution of imitation draws $\widehat{M}$.[32] In the *laissez faire* allocation, the cutoff is given by the productivity level that makes an agent indifferent between the wage net of fixed cost and profit. When there is no diffusion (such as if $\beta = 0$), these two allocations coincide, since aggregate income is maximized when each agent maximizes individual income. However, in the presence of diffusion, the cutoff productivity is greater in the optimal allocation than in the *laissez faire* allocation.

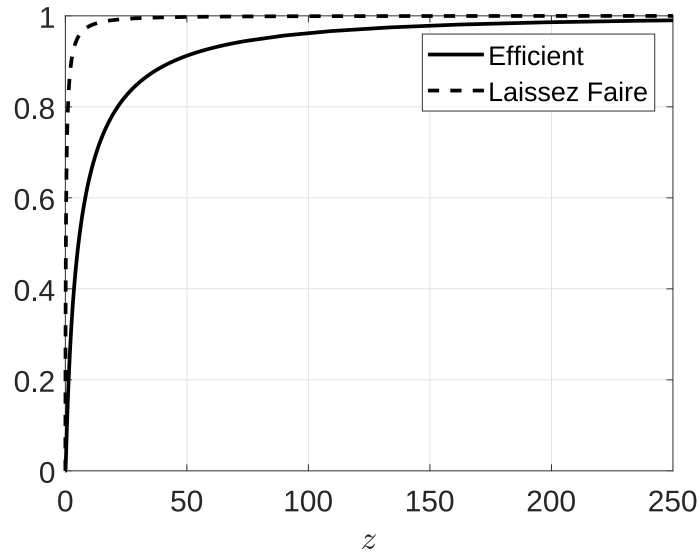
We solve for planner's allocation and compare it to the baseline economy. The difference is large. Aggregate income is 290 percent greater in the economy where workers are allocated optimally compared to that where allocation is decentralized. These gains are driven by an endogenous shift in the stationary distribution of productivities illustrated in Figure 5. 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.

Figure 6 turns to the importance the various diffusion parameters in generating these income gains. First, Figure 6 shows that the persistence in productivity ρ plays a critical role. Holding the remaining parameterization fixed, the income gained by moving to the efficient allocation changes from approximately 2 percent when $\rho = 0.2$ to 6328 percent at $\rho = 0.95$ (not shown on the graphs for the sake of readability). This implies an interesting potential role for policy – to the extent that income persistence varies across economies (countries, sectors, etc.), there is potentially much larger scope for policy gains in other economies than we find here.

Figure 6a shows that lower θ increases the gains from policy. This is because higher

[32] To be clear, aggregate income is the sum of all profit from those operating firms, the sum of all wages paid to workers and subtracts the sum of all costs paid to access the labor market.

Figure 5: Stationary Productivity CDF



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. Moving to $\theta = 0.25$ lowers the gains from our baseline 290 percent down to 215 percent. In contrast, Figure 6b shows that β can play a large role. 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. Note, however, that this need not be the case. At $\rho = 0.2$, for example, the gains decline from 2 percent to slightly below 1 percent when β decreases from 0.26 to 0.10.

The results show that the complementarity between β and ρ are critical for generating a significant role for diffusion. 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.

Finally, it is worth highlighting that the magnitude of these gains naturally depend on the model from which they are derived. Other models with different assumptions potentially generate higher or lower impact of diffusion. The estimated diffusion parameters, however, remain the same.

Figure 6: % Income Changes in Efficient Allocation

(a) Impact of θ (b) Impact of β

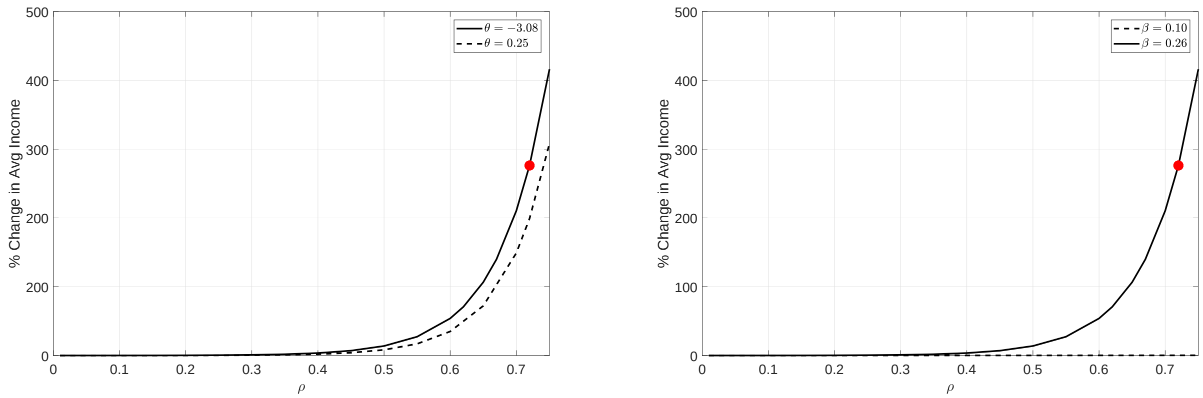


Figure notes: This figure plots the average income change in response to the optimal wage subsidy subsidy for various diffusion parameter combinations. Our estimated gains are indicated by the red dot on the solid line.

6 Discussion

6.1 Alternative Modeling Choices

As we highlighted in Section 2, 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 (net of the directedness parameter) because it allows us to link a specific framework to the broader identification results in Section 2 in a relatively transparent way. Moreover, it underlies a large portion of the literature.

In Appendix A show how a variety of different diffusion processes can be re-written as matches $\hat{z}$ drawn from a distribution $\widehat{M}(\hat{z}; z, \theta)$ that fall under the assumptions required for the identification procedure laid out in Section 2.

In Appendix C we hold the diffusion process fixed, but vary the surrounding model structure. We consider two possibilities. First, we we allow for fixed, *ex ante* heterogeneity in the utility benefit from operating a firm. This accounts for the fact that individuals potentially have different incentives for starting a business that may be unrelated to productivity, such as flexible hours. This changes the planner's problem, as not all individuals of the same productivity exit in response to a subsidy.

6.2 External Validity

External validity is a concern in any randomized controlled trial, and our results are certainty subject to the same criticism. While there is little we can do to directly address this, we note that our results suggest a number of important dimensions of heterogeneity one could test across sectors or countries.[33]

For example, the estimated persistence of productivity, ρ, is low relative to other estimates in more developed countries (see Brooks et al., 2018, for a longer discussion of this result). Moreover, the quantitative results show that this persistence parameter is an important margin for the quantitative importance of diffusion. To study how our results would change when applied to other economies with more persistent productivity, we provide a number of additional results in Appendix D. We begin by showing that the treatment did not change the persistence of profit between treatment and control firms. Thus, one could in principal estimate persistence independent of available RCT data. Next, we ask what combinations of (β, θ) would rationalize the results for a fixed level of ρ and show that there are larger gains from policy at higher levels of ρ. Thus, at least from the perspective of our estimated parameters, our results suggest a lower bound.[34]

Finally, we introduce idiosyncratic distortions into the model [**to be completed**]. We calculate how much our procedure understates the importance of diffusion when they are introduced. Again though, our results are a lower bound on the importance of diffusion.

Overall, we view these results as suggestive evidence of differential impacts of diffusion across time and space. Unfortunately, without running similar experiments around the world, it is difficult to say much beyond such suggestive counterfactuals. This is, in large part, due to the nature of the problem. The inability to directly observe the diffusion process puts strict data requirements on the identification of these parameters, though we emphasize that these are important questions to study in the future. We view our results as a framework that can be utilized to answer such important questions in the future.

[33] This idea follows closely a procedure familiar to many macro-development papers. We calibrate/estimate a model to a particular economy, then allow parameters to vary to study their importance in generating quantitative results.

[34] The other parameters β and θ have no direct empirical counterpart. To our knowledge, the only other paper using the elasticity β is Buera and Oberfield (2017), who calibrate $\beta \in [0.6, 0.7]$ depending on the exact model specification, chosen to match the growth rate of aggregate TFP in the U.S.. This again suggests our estimates are a lower bound relative to the existing values utilized in the literature.

7 Conclusion

This paper uses evidence from a randomized controlled trial to identify a model of firm-to-firm transmission of productivity. Our results imply an important role for diffusion. The efficient level of the learning externality increases income by 9 percent, and the strength of diffusion is sufficient to overcome the negative pressure on incumbent firms due to the entry of more productive firms. 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.

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 (Brooks et al., 2018). Beaman and Dillion (2017) point to frictions in the information market, while Fogli and Veldkamp (2016) 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.

References

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

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

_, Azam Chaudhry, Shamyla Chaudry, Amit K. Khandelwal, and Eric Verhoogen, "Organization Barriers to Technology Adoption: Evidence from Soccer-Ball Produces in Pakistan," *Quarterly Journal of Economics*, 2017, *132* (1), 1101–164.

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

Beaman, Lori and Andrew Dillion, "Diffusion of Agricultural Information within Social Networks: Evidence on Gender Inequalities from Mali," December 2017. Working Paper, Northwestern University.

BenYishay, Ariel and A. Mushfiq Mobarak, "Social Learning and Incentives for Experimentation and Communication," *Review of Economic Studies*, forthcoming.

Brooks, Wyatt and Kevin Donovan, "Eliminating Uncertainty in Market Access: The Impact of New Bridges in Rural Nicaragua," December 2017. Working Paper, University of Notre Dame.

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

Buera, Francisco J. and Ezra Oberfield, "The Global Diffusion of Ideas," July 2017. WUSTL Working Paper.

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

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

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

Fogli, Alessandra and Laura Veldkamp, "Germs, Social Networks and Growth," July 2016. Working Paper, NYU.

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

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

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

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

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

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

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

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

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

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

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

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

Midrigan, Virgiliu and Daniel Yi Xu, "Finance and Misallocation: Evidence from Plant-Level Data," *American Economic Review*, 2014, *104* (2), 422–458.

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

_ , _ , and Michael E. Waugh, "Equilibrium Technology Diffusion, Trade, and Growth," Working Paper 20881, National Bureau of Economic Research January 2015.

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

Table of Contents for Online Appendix

A Examples of Diffusion Processes

In this section we give several examples of $\widehat{M}$ distributions that are functions of the underlying distribution of firm productivities M and are indexed by a parameter θ.

A.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.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:

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

or following the notation more standard in the paper:

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

A.3 Innovations through Imitation

Here we show how the Buera and Oberfield (2017) environment maps into that considered in this paper. In their model (adapted to our notation), an agent with productivity z receives new arrivals of ideas that have two components: z_m that comes from a random match from another agent, and γ a random innovation on that idea. Then $\hat{z} = \gamma^{1/\theta} z_m$. Here, z_m is a uniform draw from the distribution of productivities. Then if γ has a cumulative density function given by Γ, then:

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

A.4 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. Using the same moments as in the text, Proposition 4

Proposition 4. *If the following two conditions hold, then there exists a unique pair $(\beta^*, \rho^*) \in (0,1)^2$ that solve equations (2.2) and (2.3). Those conditions are:*

$$\Gamma_1 \in \left(1, \frac{\int\int \hat{z} d\widehat{H}_T^H(\hat{z}, z) dH_T^H(z)}{\int\int \hat{z} d\widehat{H}_T^L(\hat{z}, z) dH_T^L(z)}\right) \tag{B.1}$$

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

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 Other Models

C.1 *Ex Ante* Heterogeneity in Utility Benefit from Entrepreneurship

[to be completed]

D Additional Results

D.1 Re-estimating (β, θ) for fixed ρ

[to be completed]

D.2 Allowing for Idiosyncratic Distortions

[to be completed]