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

September 2018

Preliminary and Incomplete – Comments Welcome

Abstract

We consider an increasingly utilized class of general equilibrium model in which knowledge diffusion generates positive spillovers across firms. Each period a firm "matches" with another randomly chosen 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 the remaining parameter values. We conduct a randomized controlled trial among Kenyan firms in which firms from the left tail of the profit distribution are matched one-to-one with firms from the right tail, then use the empirical results to estimate these parameters. Our quantitative results imply an important role for knowledge diffusion in a series of policy experiments in the model. Removing a labor market distortion, for example, increases real income by 69 percent at our estimated parameters, compared to 33 percent in an identical model with no productivity transmission. We show that the identification results extend to a number of alternative modeling assumptions, including occupational choice and endogenous search intensity, using the same empirical moments.

*This paper is still preliminary and incomplete, and results are subject to change. Thanks to participants at Notre Dame and the Society for Economic Dynamics for helpful comments. *Contact Information:* Brooks, wbrooks@nd.edu; Donovan, kevin.donovan@yale.edu; Johnson, tjohns20@nd.edu

1 Introduction

A growing literature recognizes that knowledge or skill transfer among firms can potentially play an important role in the development process.[1] Yet quantifying the effects of knowledge transfer is difficult, as it requires identifying parameters of the knowledge transfer process. This is important for two reasons. First, it is necessary to understand the quantitative importance of this channel for growth. Relatedly, since knowledge is non-rival and generates positive spillovers, the model parameters are critical for implementing proper policy. Our goal in this paper is to take this class of models to the data, identify key diffusion parameters, and take seriously the quantitative implications of this knowledge transfer channel.

We do so by constructing a dynamic general equilibrium model in which entrepreneurial ability is affected by the distribution of productivity among operating firms, similar to Jovanovic and Rob (1989). In particular, every period agents receive a chance to learn from a randomly selected operating firm. Agents meet one-to-one, and if useful, they can internalize some portion of their partner's ability. The quantitative importance of this channel depends critically on the parameters governing the diffusion process. This includes the extent to which individuals can internalize their partner's productivity, the extent to which search can be directed at high productivity firms, and finally, the extent to which this learning is persistent – and therefore useful – over time.

One can immediately see the identification problems inherent in measuring these parameters. An agent's skill is explicitly time varying, which eliminates fixed effect methods used in other work using panel data, such as Nix (2016) and Cornelissen et al. (2017). In this paper we take a different approach. We prove that these parameters are uniquely identified in the presence of randomized changes in the likelihood and quality of matches, then conduct such a randomized controlled trial (RCT) to estimate model parameters and conduct counterfactuals.

Specifically, we show theoretically that if we vary who one randomly selected group (i.e., the treatment group) draws matches from, we can use differences in profit to iden-

[1] See Lucas (2009), Lucas and Moll (2014), Perla and Tonetti (2014), and Buera and Oberfield (2017) and Perla et al. (2015) in the context opening of trade.

tify the diffusion parameters. The intuition for our results follows closely the intuition behind running any RCT – we need relatively little information about the underlying process by which the control group makes decisions, and instead, need to focus only on how the treatment group changes decisions relative to the control. Similarly here, the identification of diffusion parameters does not require knowing who the control group matches with, only that providing the treatment group with systematically better matches generates changes in outcomes. An important function of this result is that the uniqueness result is robust to a variety of different assumptions about model features (i.e., including occupational choice or not) and diffusion processes (i.e., endogenous or exogenous search intensity), and furthermore, all utilize identical moments from the data. Second, these parameter estimates are independent of remaining model parameters. Both result from the fact that the RCT data essentially "nets out" many of the identification problems that would occur without randomized variation, and allows the results to be useful in a variety of different modeling scenarios.

To take our theoretical results to the data and study the quantitative importance of this channel, we first conduct a randomized controlled trial in Nairobi, Kenya. In this field experiment, low-profit business owners are matched one-to-one with higher profit owners, then compared to a control group. The matches are random, conditional on industry. The goal was to stay as close as possible to the theoretical construct of "meeting" used in these models, and to that end, individuals were provided with no prompts or specific topics to discuss. Our experiment simply gave one business owner the phone number of another, and let them know that both parties had agreed to meet to discuss business issues. This experiment generates the proper variation to identify the model. In addition to being able to track the average difference between treatment and control, the random match "quality" provides the variation required to identify the intensity with which productivity transmits to the less productive firm in the match. The empirical results are broadly consistent with the importance of knowledge diffusion, and the full set of reduced form results are available in our companion paper Brooks et al. (forthcoming). Profits are 23 percent higher in the treatment group relative to the control.[2] This treatment impact is increasing in the relative profit gap

[2] In principal, a change in profit could be driven by any number of channels, including profit sharing, loans, bulk discounts, etc. We consider a number of alternative possibilities in Brooks et al. (forthcoming) and reject them.

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.

We then conduct a series of experiments in the model to measure the importance of the diffusion channel given our estimated parameters. First, we consider the tax rate on firm ownership that maximizes *ex ante* welfare. Low productivity firms generate an externality, as others randomly match with them. Thus, welfare increases by incentivizing those entrepreneurs to instead become workers. The welfare gains from this tax, relative to an economy with no tax, are **X percent**. We furtherstudy distortions considered in the literature, including a tax on wages relative to business operation, designed to incentivize marginal entrepreneurs into business operation. Similar to our first exercise, the impact is large. Eliminating this distortion increases income 69 percent when the diffusion process is parameterized using data from the randomized controlled trial, compared to 33 percent in the no-diffusion case. Lastly, we consider a counterfactual where a pool of high productivity agents is added to the initial pool of agents. These new firms increase labor demand, which increases the equilibrium wage. All else equal, this lowers profits from existing operating firms. However, it also allows existing firms to learn from new productive firms, which increases profitability. The net impact is therefore a quantitative question. In the model without diffusion, the second effect is absent, and we find that the profits of operating firms fall by 7 percent after the introduction of high productivity of firms. Many of these existing firms exit and their owners become workers in response. However, in the calibrated model with diffusion, we find that the first effect dominates the second and firms in the initial group of agents increase in profitability. This exercise demonstrates that the diffusion effect is quantitatively important and, for parameters identified by experimental evidence, is capable of more than offsetting negative general equilibrium effects.[3]

[3] This becomes important in situations such as opening to trade, where some subset of domestic firms get access to more productive foreign firms. Atkin et al. (2017) highlight this learning-by-exporting channel with a randomized controlled trial in Egypt, but the experiment is not designed to study the domestic spillovers onto control firms. Perla et al. (2015) and Buera and Oberfield (2017) include such margins, but the dominant quantitative channel requires identification of model parameters, which we do here.

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. Lagakos et al. (2018) use the randomized controlled trial results from Bryan et al. (2014) to identify and quantify the welfare gains of rural-urban migration in Bangladesh. Brooks and Donovan (2017) use exogenous variation in infrastructure placement to identify a model of risky farm investment. Our paper shares a similar style but focuses on knowledge diffusion.

Our work therefore adds empirical evidence to the mainly theoretical literature studying the effects of diffusion in general equilibrium models, including Jovanovic and Rob (1989), Lucas (2009), Lucas and Moll (2014), and Perla and Tonetti (2014). These models consider the case of diffusion where an imitating firm is able to copy the productivity of the firm it is imitating, potentially at some cost. We consider a less extreme version, similar to Buera and Oberfield (2017), where agents imitate with only some elasticity, though agents are not required to pay a cost to access different or better matches. That is, we study the "passive" learning in much of the literature. Recent work has also extended these models to consider within and across firm diffusion (Herkenhoff et al., 2018) and international trade (Buera and Oberfield, 2017; Perla et al., 2015).

At the same time, there exists an important micro literature that focuses documenting the existence of social learning, including Foster and Rosenzweig (1995), Munshi (2004), Bandiera and Rasul (2006), and Conley and Udry (2010). Recent work, including Beaman et al. (2015) and BenYishay and Mobarak (2017), highlights the importance of targeting certain network nodes to maximize diffusion of new information. The importance of diffusion has been considered in a number of other development contexts, including trade (Atkin et al., 2017) and microfinance (Banerjee et al., 2013). These papers do not discuss the aggregate implications of diffusion. This is in part because tractable aggregate models of diffusion require some abstraction in the underlying diffusion process. Thus, despite the generally agreed upon importance of the topic, the two literatures do not overlap. Our goal is to take as given the aggregate models of diffusion, and assess their relevance – despite the abstractions – using a combination of RCT data and theory. 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 Model

In this section, we lay out a dynamic general equilibrium model in which individuals can learn from others in the economy and increase productivity. While we lay out a specific model here, we discuss the robustness of our results to other model assumptions in Section 7.

Time is discrete and infinite. In each period there is a unit mass of agents. Each agent has an exogenous probability δ of dying each period, while δ agents are born. Each agent is characterized by productivity z which evolves over time, and is discussed in detail below. In every period, the 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.

The profit of being an entrepreneur with productivity z is

$$\pi(z) = \max_{l \geq 0} \; p z^{\alpha} l^{1-\alpha} - wl \tag{2.1}$$

where p is the (normalized) price of output, w is the market clearing wage, and $\alpha \in (0,1)$ is the returns to skill. The value of having entrepreneurial skill z is

$$v(z) = \max\{\pi(z), w\} + (1-\delta)\mathbb{E}_{z'|z} v(z'). \tag{2.2}$$

Solving the entrepreneur's problem yields

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

Thus, 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. The entrepreneur with $\underline{z}$ is indifferent between the two occupations, which yields

$$w = \alpha \underline{z} p^{\frac{1}{\alpha}} \left(\frac{1-\alpha}{w} \right)^{\frac{1-\alpha}{\alpha}} \implies \underline{z}(w) = \left(\frac{w}{p} \right)^{\frac{1}{\alpha}} \frac{(1-\alpha)^{\frac{\alpha-1}{\alpha}}}{\alpha} \tag{2.3}$$

2.1 Endogenous Evolution of Productivity

Newly Born Agents Newly born agents draw a initial productivity z from the exogenous distribution G.[4]

Continuing Agents The continuation value $\mathbb{E}_{z'|z} v(z')$ naturally depends on the evolution of productivity z. Productivity of existing agents evolves as a result of idiosyncratic shock ε and from a meeting with another agent with probability $\hat{z}$. The law of motion, given ε and $\hat{z}$ is given by

$$\log(z') = c + \rho \log \left(\max \left[\hat{z}^{\beta} z^{1-\beta}, z \right] \right) + \varepsilon. \tag{2.4}$$

In words, each agent z meets another agent $\hat{z}$. Agent z is able to internalize some fraction β of their partner's productivity. If that new productivity $\hat{z}^{\beta} z^{1-\beta}$ is larger than current productivity z, the agent adopts the new productivity. Otherwise, she discards it and remains at z. Independent of this match quality, each agent is also subject to exogenous innovations ε. The parameter c determines long-run average productivity of agents, while ρ corresponds to the persistence of productivity innovations.

The idiosyncratic shock is a draw from an exogenous, time-invariant distribution F. On the other hand, match quality $\hat{z}$ is drawn from the endogenous distribution $\widehat{M}$, given by

$$\widehat{M}(\hat{z}) = \begin{cases} 0, & \text{if } \hat{z} < \underline{z} \\ \left(\frac{M(\hat{z}) - M(\underline{z})}{1 - M(\underline{z})} \right)^{\frac{1}{1-\theta}}, & \text{if } \hat{z} \geq \underline{z} \end{cases} \tag{2.5}$$

where M is the distribution of productivities among producing entrepreneurs in the

[4]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.

economy and $\underline{z}$ is again the cutoff productivity for entrepreneurship. Note also the parameter $\theta \in [0, 1)$. A purely random draw from existing entrepreneurs corresponds to $\theta = 0$. As $\theta \to 1$, the distribution of draws becomes more concentrated in the right tail of operating firms, which we interpret as firms having a greater ability to direct their imitation toward the most productive firms.

2.2 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 (2.2) with the assocaited decision rules for labor and occupational choice, the evolution of M is consistent with the decision rules and is given by

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

and the wage w clears the labor market, which requires a solution to the implicit equation

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

3 Identification of Productivity Transmission

The only dynamics from the model come from the evolution of productivity.[5] Our goal is to identify the parameters of the diffusion process, characterized in equation (2.4). If one could observe each match in the economy, this would be a trivial exercise. Of course, this is not the case. Our goal is to show that observing and manipulating a subset of matches in a particular way is sufficient to identify these parameters. In Section 4 we introduce and discuss a randomized controlled trial in which we induced this type of variation.

[5]This is not critical for the results, but meant only to highlight the importance of the diffusion channel.

3.1 A Hypothetical Experiment in the Model

Take an arbitrary group of agents with ability given by the observed distribution $B(z)$.[6] This could be a strict subset of the economy-wide productivity distribution M or equal to M. In equilibrium, these agents match with a distribution of productivities $\widehat{M}(\hat{z})$ (defined in equation 2.5). Our goal now is to induce the variation required to identify the model parameters (β, ρ, θ) that are included in the forward equation for diffusion.

We proceed in three steps. First, randomly allocate B agents into treatment and control groups. Second, define a subset $\widehat{M^T}(\hat{z})$that first order stochastically dominates H. Third, for every agent i in the treatment group, randomly select a firm in H^T that will match with i. That is, each i will be explicitly matched with an observable $\hat{z}_i \in \widehat{M^T}(\hat{z})$.

The two steps of randomization are (1) the delineation between treatment and control and (2) the randomization of observed matches within the treatment group. This allows us to not only compare across control and treatment, but also compare individuals in the treatment group across their various (observed) match productivities. As we show below, this is sufficient to identify diffusion parameters (β, ρ, θ). We proceed as follows. First, we show that β and ρ are uniquely pinned down from empirical moments independent of any other model parameters. We then show that $\theta(\beta, \rho)$ is uniquely identified, only depending on the two parameters β and ρ. Thus, this randomized variation allows us to identify the three diffusion parameters independent of the remaining model parameters.

3.2 Joint Identification of β and ρ

The transmission parameter β is identified by comparing the effects on two initially identical participants from receiving a very high productivity $\hat{z}$ match to those receiving a relatively low $\hat{z}$ match. If those receiving a high $\hat{z}$ realize much bigger returns compared to their receiving a lower $\hat{z}$, we conclude that β is high. The persistence term ρ instead is identified by comparing two participants with different productivities, but

[6] Recall, the ability to translate ability into profits allows us to read these distributions directly off observed profit variation.

matched with $\hat{z}$ draws that are the same. The extent to which the initial disparity in productivity remains after treatment (conditional on β) measures the degree of persistence.

To formalize this intuition, we use two moments on participant firms. First, we compare participants that received a "high" productivity $\hat{z}$ draw to those receiving a "low" $\hat{z}$ draw.[7] If Ω is the set of $(z, \hat{z})$ pairs in the participant group, let Ω_H and Ω_L be disjoint subsets of Ω with probability density functions $m_H(z, \hat{z})$ and $m_L(z, \hat{z})$, such that:

$$\forall \hat{z}_0, \int_0^{\hat{z}_0} \int_0^{\infty} m_H(z, \hat{z}) dz d\hat{z} < \int_0^{\hat{z}_0} \int_0^{\infty} m_L(z, \hat{z}) dz d\hat{z}. \tag{3.1}$$

That is, the $\hat{z}$ draws within Ω_H are "better" than those within Ω_L. We will guarantee this by partitioning Ω by $\hat{z}$ so that, in fact, every element of Ω_H has a greater $\hat{z}$ than any element of Ω_L.

Now we define the first moment condition using these subsets. The average productivity after treatment for members of Ω_H compared to members of Ω_L is:

$$\Gamma_1 = \frac{\int\int\int e^{c+\gamma+\varepsilon} z^{\rho} \max\left[1, \frac{\hat{z}}{z}\right]^{\rho\beta} m_H(z, \hat{z}) dz d\hat{z} dF(\varepsilon)}{\int\int\int e^{c+\gamma+\varepsilon} z^{\rho} \max\left[1, \frac{\hat{z}}{z}\right]^{\rho\beta} m_L(z, \hat{z}) dz d\hat{z} dF(\varepsilon)} \tag{3.2}$$

where Γ_1 is the value of this ratio in the data, which is a target for these parameter values. Notice that the independence of the ε terms, along with the fact that several constants appear in the numerator and denominator, imply this can be written as:

$$\Gamma_1 = \frac{\int\int z^{\rho} \max\left[1, \frac{\hat{z}}{z}\right]^{\rho\beta} m_H(z, \hat{z}) dz d\hat{z}}{\int\int z^{\rho} \max\left[1, \frac{\hat{z}}{z}\right]^{\rho\beta} m_L(z, \hat{z}) dz d\hat{z}} \tag{3.3}$$

Since Γ_1, m_H and m_L come directly from the data, only β and ρ are yet unknown in this equation. This implies another moment from the data is required for identification.

The second moment used to identify these parameters is the relationship between initial productivity z and final productivity z' in the whole set of participants Ω with density $m(z, \hat{z})$. The identification strategy is similar to that employed in standard firm dynamics models with AR(1) processes, but much be adjusted to take into account

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

the diffusion process, which is inherently asymmetric.

$$\Gamma_2 = \frac{Cov[z, z']}{E[z]E[z']} + 1 = \frac{\int\int z^{1+\rho} \max\left[1, \frac{\hat{z}}{z}\right]^{\rho\beta} m(z, \hat{z}) dz d\hat{z}}{\int z \int m(z, \hat{z}) d\hat{z} dz \cdot \int\int z^{\rho} \max\left[1, \frac{\hat{z}}{z}\right]^{\rho\beta} m(z, \hat{z}) dz d\hat{z}} \tag{3.4}$$

The covariance normalized by the means of z and z' identify persistence. This is equal to the correlation of z and z' when divided by the coefficients of variation of both z and z'. Since levels, and not standard deviations, identify ρ, including the coefficients of variation is extraneous.

To see this directly, consider the standard exercise in models with AR(1) processes on idiosyncratic productivity and no diffusion. This is the special case of our model where $\beta = 0$. If we assume F is a normal distribution, then z is log-normal. If the observed mean of $\log(z)$ is μ and the variance is σ^2, then:

$$\frac{Cov[z, z']}{E[z]E[z']} + 1 = e^{\sigma^2 \rho} \tag{3.5}$$

Given that σ^2 is observed directly, then ρ is immediately identified from this statistic. Instead, the correlation is:

$$\frac{Cov[z, z']}{\sqrt{Var(z)Var(z')}} = \frac{e^{\sigma^2 \rho} - 1}{\sqrt{e^{\sigma^2} - 1}\sqrt{e^{\sigma^2 \rho^2} - 1}} \tag{3.6}$$

Given that the identification of ρ is much more straightforward with the mean-normalized covariance, rather than with correlation, we use the mean-normalized covariance.

For measurement, it is also useful to use the partial correlation of z and z' rather than the raw correlation, since there are likely to be nontrivial time effects. For cases where only fixed effects are removed, running a fixed effects regression, this moment is equal to: $1 + \frac{Var(z)}{E(z)E(z')}\beta^{OLS}$.

Proposition 1. *Let R be the ratio of the ex post profits in the treated "high" group and treated "low" group. Let R_M be the ratio of profits in the "high" profit matches and the "low" profit matches. Defining $C = Cov(z, z')/(\mathbb{E}(z)\mathbb{E}(z')$ and $cv = 1 + [\sigma_z/E(z)]^2$. Then if R is in $[1, R_M]$, and C is in $[1, cv]$, then (ρ, β) is in $[0, 1]^2$, and is uniquely identified by as the solution to (3.3) and (3.4).*

Identification of θ, the directedness of imitation, comes from comparing treated and untreated firms. In particular, an untreated firm receives a draw from $\widehat{M}$, determined by the population distribution of firms, while a treated firm receives a draft from a known distribution $\widehat{H}$. Denoting profits of untreated firms as π^U and of treated firms as π^T, the average profit of untreated firms is given by:

$$E[\pi^U] = cE[e^{\varepsilon}] \int \int \max[z, \hat{z}^{\beta} z^{1-\beta}]^{\rho} d\widehat{M}(\hat{z}) dH(z) \tag{3.7}$$

Then taking the ratio of average profits in the treated and untreated groups yields:

$$\frac{E[\pi^U]}{E[\pi^T]} = \frac{\int \int \max[z, \hat{z}^{\beta} z^{1-\beta}]^{\rho} d\widehat{M}(\hat{z}) dH(z)}{\int \int \max[z, \hat{z}^{\beta} z^{1-\beta}]^{\rho} d\widehat{H}(\hat{z}) dH(z)} \tag{3.8}$$

The parameter θ only appears here in the distribution $\widehat{M}$, which guarantees identification as shown in the next proposition.

Proposition 2. *Given average profits in the untreated group $E[\pi^U]$ and the treated group $E[\pi^T]$, parameters ρ and β, and for known distributions of mentors $\widehat{H}$, the population from which the study population is drawn H, and the set of operating firms $\widehat{M}$, then the value of θ that solves equation (3.8) is unique.*

The critical benefit of the RCT data is the ability to identify β and ρ independently of the remaining model parameters. Thus, while Proposition 2 only identifies the function $\theta(\beta, \rho)$, Proposition 1 guarantees that that only one (β, ρ, θ) satisfies both moment equations (3.3) and (3.4).

It is worth highlighting the RCT benefits in this identification strategy. The critical benefit is the ability to identify β and ρ independently of the remaining model parameters. Thus, while Proposition 2 only identifies the function $\theta(\beta, \rho)$, Proposition 1 guarantees that that only one (β, ρ, θ) satisfies both moment equations (3.3) and (3.4). More specifically related to ρ, one might think that the time series of profit in the control group could identify ρ. The issue here is that we do not know match quality for members of the control group, and thus would potentially conflate match quality with persistence. The RCT data lets us sidestep this issue by using the persistence of the *treatment* group instead. Lastly, as we show later in Section 7, the RCT data allows our results to be utilized under other model assumptions. We can eliminate the occupational choice margin, if one prefers to focus on an economy with no labor market, and can also accommodate endogenous search intensity. Indeed, the result is so stark that the proof essentially goes through unchanged.

4 Field Experiment and Reduced-Form Results

The previous section shows that we can identify diffusion parameters using specific data. We therefore run a randomized controlled trial that generates such data, and use it to identify model parameters. A complete description of the program and reduced-form results are available in our companion paper Brooks et al. (forthcoming), though we reproduce some of the relevant results here.

Our experimental design randomly matches older, profitable entrepreneurs with younger entrepreneurs. The younger entrepreneurs were then followed for over a year to measure changes in business practice and profit. For comparison, a separate group was randomized into a formal business training course that we arranged to be taught by faculty from a local university. The outcomes of both groups were compared to a control group that received neither a mentor nor classes.[8]

4.1 Details of RCT and Data Collection

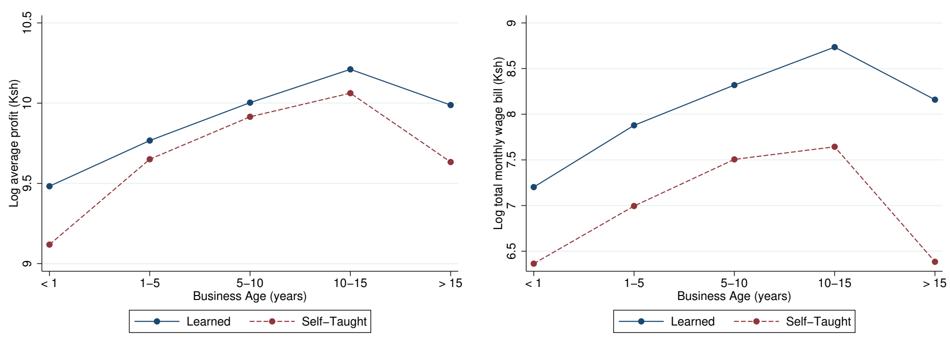
The experiment took place in Dandora, Kenya, a dense urban slum on the outskirts of Nairobi. Self-employment is ubiquitous in Dandora with a huge number of street-level businesses operating in a variety of industries, such as retail, simple manufacturing, repair and services. We began by conducting a 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.

[8] 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. In Brooks et al. (forthcoming) we attribute 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.).

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 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

(a) Average profit (b) Total wage bill



Selection and Randomization We start from a sample of female business owners who have been in operation for less than 5 years.[9] 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 $t = 1, 2, 3, 4, 7, 12, 17$. Looking ahead, these dynamics will help us identify the persistence term in the diffusion process.

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. We then recruited business owners with the highest profit until we had a sufficient number for

[9]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.

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. Therefore, the can be viewed as representing the upper half of profitable firms operated by older, experienced entrepreneurs. Note, however, that this selection procedure does not allow the simple causal identification derived from the RCT. Our selection procedure allows us to identify the impact on these more productive owners through a regression discontinuity design, which we discuss more below.

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.[10]

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. Ninety-five percent of those contacted as mentors agreed to participate. However, we provided no topics to discuss, instead preferring that the content of any discussions was self-directed. Literally, 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. (forthcoming), 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.

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

4.2 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

$$\pi_{it} = \alpha + M_i\beta + y_{i0}\delta + X_i\eta + \theta_t + \epsilon_{it}. \tag{4.1}$$

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.

As discussed in detail in Brooks et al. (forthcoming), 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 model from Section 2 assumes that there should be no gains to these individuals, which is implicitly assumed through the use

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

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

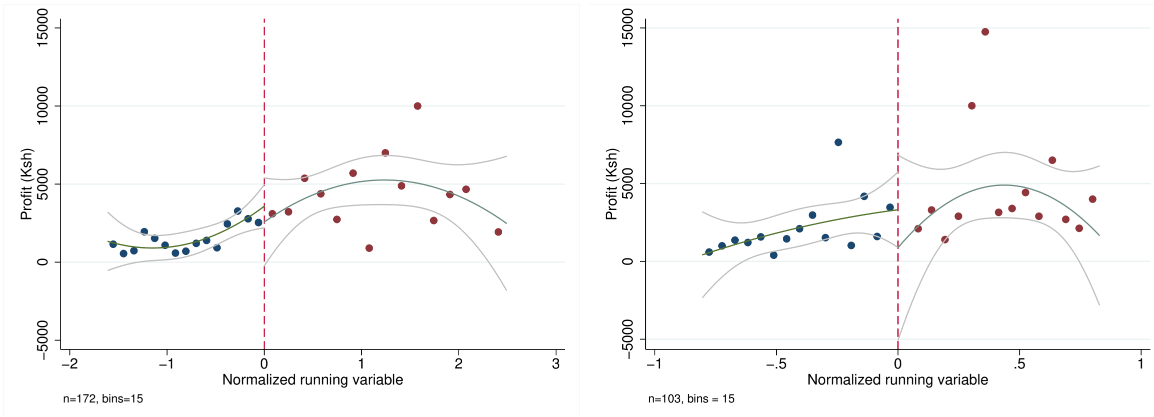
of the max function in law of motion for productivity. These individuals were not randomly selected relevant to their peers, and thus cannot be directly compared to a control group as the others were. However, our design allows us to use the selection procedure to identify the causal impact of being 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 2 plots profit along with a fitted quadratic and its 95 percent confidence interval. Figure 2a uses the entire sample, while Figure 2b uses the Imbens and Kalyanaraman (2012) procedure to choose the optimal bandwidth. Both use 15 bins on either side of the cutoff. Figure 2 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

Figure 2: Profit for mentors and non-mentors (from Brooks et al. (forthcoming))

(a) Full sample (b) IK Optimal Bandwidth



mentors, we run the regression

$$\pi_i = \alpha + \tau D_i + f(N_i) + \nu_i \tag{4.2}$$

where π_i is profit, $D_i = 1$ if individual i was chosen as a mentor ($\widehat{\varepsilon}_i \geq \overline{\varepsilon}$, $f(N_i)$ is a flexible function of the normalized running variable $N_i = (\widehat{\varepsilon}_i - \overline{\varepsilon})/\sigma_\varepsilon$, and ν_i is the error term. The parameter τ captures the causal impact of being chosen as a mentor. We use local linear regressions to estimate the treatment effects on profit and inventory, along with business practices of record keeping and marketing. The results are in Table 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 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 (2.4), which is assumed here and in much of the existing literature.

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 produc-

Table 2: Regression discontinuity results for mentor treatment effect (from Brooks et al. (forthcoming))

Percent of IK optimal bandwidth	Scale: Profit	Scale: Inventory	Practices: Marketing	Practices: 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.

tivity firms, there is nothing in the experimental design that guarantees this outcome.[11]

5 Model Parameterization

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 along with our information about firms that were treated or were mentors.

The model parameterization can be broken into different parts that can be considered separately. First, we identify the parameters of the productivity transmission process, θ and β, discussed extensively in the previous section. The only other parameter required to identify these parameters is the persistence of productivity ρ. That is, our identification procedure has the useful property that only one other parameter value is needed to identify β and θ. In this sense, the calibrated values of the productivity transmission process are not dependent on the calibration strategy used to derive the other parameters. Conditional on ρ, θ and β are identified from the overall

[11]For example, the mentor-mentee relationship may be consistent with collaboration or business group, where both sides gain from interacting with the other (e.g. Cai and Szeidl, forthcoming). On the other hand, if the time requirement is high enough, there could be negative effects on mentors. However, the high take-up and persistence of matches suggests this second explanation is unlikely *a priori*.

treatment effect from mentors (equal to 23 percent of control group mean profit) and the treatment effect from receiving a mentor from the top quartile of the mentor profit distribution (equal to 30 percent of control group mean profit).

We make use of the results in Propositions 1 and 2 as follows. For the purposes of identifying θ and β, we assume that firm productivity is Pareto-distributed with shape parameter γ and scale parameter x_0. We choose γ and x_0 to match the mean and variance of firm profit. The distributions of treated firms H, of all mentors $\widehat{H_1}$, and of the top quartile of mentors $\widehat{H_2}$ are likewise Pareto-distributed, but each has its own scale parameter to match the mean profitability of each respective group. Applying Propositions 1 and 2 in this way implies that $\beta = 0.15$, $\rho = 0.79$, and $\theta = 0.26$.[12]

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. Therefore, to match an age of 16, we set $\delta = 0.0625$.

The wage share of value added in the model is equal to $1 - \alpha$. Within the set of firms that have employees, we find that the wage share of value added is 0.417, which implies that $\alpha = 0.583$.

We choose the price of output p to be numeraire. Our remaining parameters are σ, σ_0, c and τ. We jointly match these four parameters to the following four moments: the

[12]Note that nothing in the identification procedure depends on the functional form assumptions. Since we observe these distributions, we could in principal feed in whatever empirical distribution matches the data. This is forthcoming in the next iteration of the paper.

variance 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 (1.56), and the fraction of the workforce that is self-employed (71.3%).[13] These moments are matched when $\sigma = 1.402$, $\sigma_0 = 1.011$, $c = 1.519$ and $\tau = 0.982$.

[13]Data on the fraction of the workforce that is self-employed comes from the 2015 World Bank Development Indicators.

Table 3: Targets and Parameter Choices

Model Parameter	Description	Parameter Value	Target Moment	Source	Target Value	Model Value
Group 1						
θ	Directedness of search	0.26	Overall treatment effect (as %)	RCT results	23	23
β	Contribution of matched firm productivity to own	0.15	Treatment effect from mentor in top quartile (as %)	RCT results	30	30
ρ	Persistence of productivity	0.79	Dynamics of treatment effect	n.a.	–	–
Group 2						
μ	Mean of exogenous productivity shock distribution	0.40	$\mathbb{E}[e^{\varepsilon}] = 1$	n.a.	1	1
σ	St. dev. of exogenous productivity shock distribution	0.42	Variance of log profit in all firms	Baseline survey	0.42	0.42
σ_0	St. dev. of new entrant productivity distribution	1.011	Variance of log profit among new entrants	Baseline survey	0.961	0.961
c	Growth factor in productivity evolution	1.519	Ratio of average profit of all firms to new entrants	Baseline survey	1.560	1.560
τ	Distortionary tax on wages	0.0982	Fraction of workforce that is self-employed	World Bank WDI	0.713	0.713
Group 3						
δ	Death rate of firms	0.0625	Average age of baseline business owners	Baseline survey	34	34
α	Wage share of value added	0.583	Wage share of value added if $n > 0$	Baseline survey	0.417	0.417
μ_0	Mean of new entrant log-productivity distribution	0	Normalization	n.a.	–	–

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.

6 Quantitative Implications

We consider two exercises in the model to study the impact diffusion in our parameterized model. In Section 6.2, we study the gains from removing the labor market distortion calibrated in the previous section. We find that the gains are twice as large in our model compared to an identical model with no diffusion. In Section 6.3 we introduce a new set of agents who have higher productivity than the existing firms in the stationary equilibrium. This idea allows an experiment in which existing firms may be negatively affected, as the new productive firms push up the wage. We again find an important role for diffusion. In our model, existing firms increase profit by 7.2 percent. In the same model without diffusion, profit declines by 7.4 percent. Thus, the equilibrium impact of diffusion more than makes up for the negative equilibrium impact of higher wages.

6.1 Optimal Wage Subsidy

[**to be written**]

6.2 Gains from Removing Labor Market Distortion

As emphasized in Gollin (2008), developing countries have a substantially higher self-employment share of employment.This could be due to higher search costs, labor contracting frictions, or any number of other distortions. To capture this distortion, we introduce a wedge between the wage rate w that firms pay and the $(1-\tau)w$ that workers receive. All agents receive a lump sum transfer T that, in equilibrium, is equal to the sum of these wedges.[14] Now the cutoff value of $\underline{z}$ is determined by an indifference condition that takes the wedge into account:

$$(1-\tau)w + T = \pi(\underline{z}, w) + T \implies \underline{z} = (1-\tau)\left(\frac{w}{p}\right)^{\frac{1}{\alpha}} \frac{(1-\alpha)^{\frac{\alpha-1}{\alpha}}}{\alpha} \tag{6.1}$$

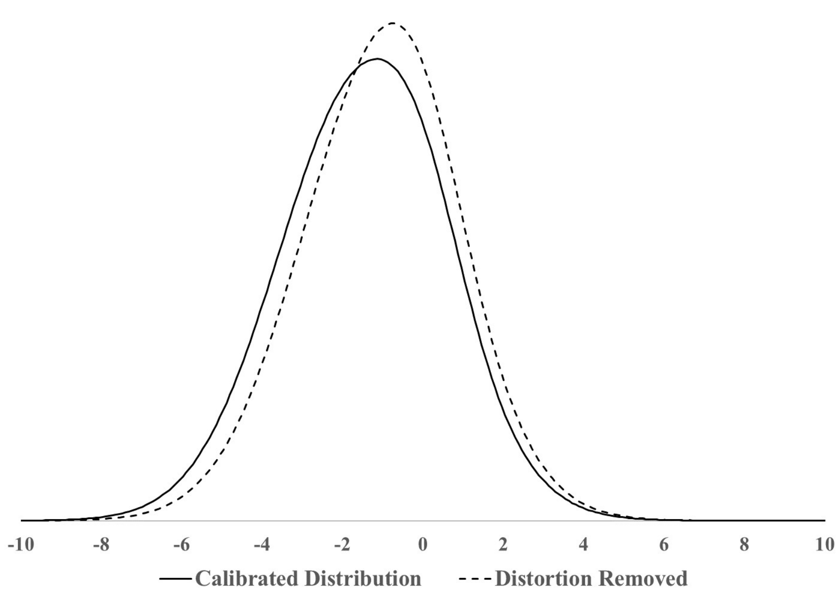
Clearly when $\tau > 0$, this reduces the value of $\underline{z}$ so that, for a given distribution M, a larger fraction of agents choose to be entrepreneurs. Moreover, within a stationary equilibrium, the fact that lower productivity firms operate generates a negative exter-

[14] That is, in equilibrium $T = \tau w M(\underline{z})$.

nality by lowering the mean of the $\widehat{M}$ distribution, which in turn causes all firms to receive, on average, lower $\hat{z}$ draws and reducing $\widehat{M}$ even more.

In this exercise, we calibrate τ to match the percentage of the workforce that operates as entrepreneurs. Our main exercise is to compute the stationary equilibrium under this calibrated value of τ, and the stationary equilibrium with $\tau = 0$, then compare total real income in the two equilibria. When β is at its calibrated value, we find that real income increases by 69% from completely removing the distortion. These gains come from two sources. First, reducing the distortion makes more labor input available for the high productivity entrepreneurs both by increasing the supply of workers, and by reducing labor demand from low productivity entrepreneurs. This allows higher productivity entrepreneurs to increase their scale and increase total output. Second, because of the transmission of productivity, the distribution of productivity across agents is endogenous. When fewer low productivity agents operate as entrepreneurs, this improves the distribution of active firms from which draws are made, and therefore improves firm growth. This causes an endogenous shift in the distribution of productivity as illustrated in Figure 3.

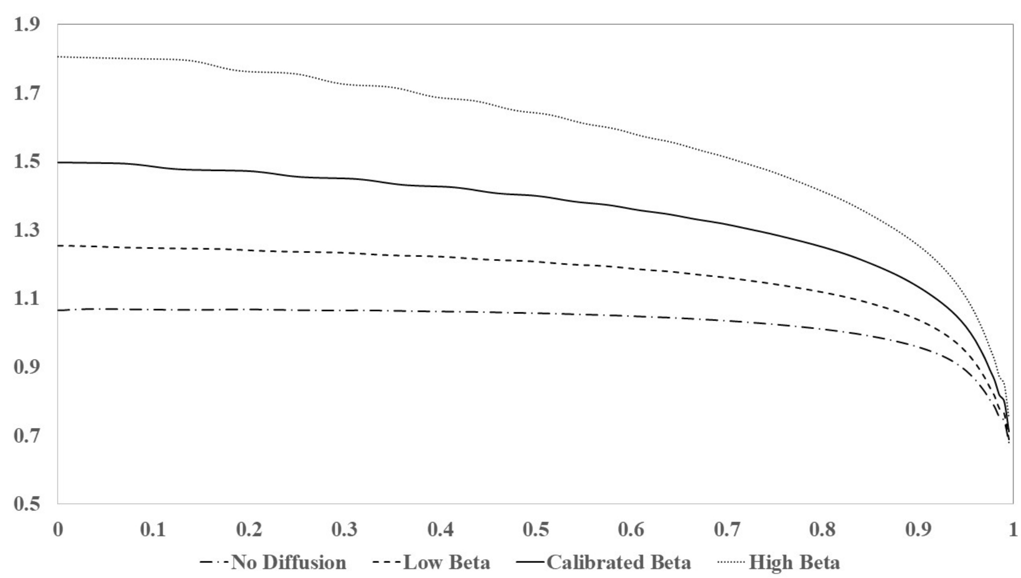
Figure 3: Log Productivity Distribution, before and after distortion removed



We compare these results to the same exercise when we turn off productivity transmission across firms by setting $\beta = 0$. In this case, real income is 33% higher in the stationary equilibrium with no distortions. As in the previous case, the reduction in τ causes more labor to be utilized by high productivity firms. However, in this case there is no change in the productivity distribution, since when $\beta = 0$ the evolution of any agent's productivity over time is only affected by exogenous, idiosyncratic shocks ε.

Because the gains to real income are nearly twice as large with productivity transmission than without (69% versus 33%), we conclude that transmission of productivity across agents generates significant amplification to the gains from removing the labor market distortion. This difference is due to the endogenous shift in the productivity distribution shown in Figure 3. As fewer agents with low productivity choose to be entrepreneurs, the distribution of $\hat{z}$ draws improves, which endogenously shifts the productivity distribution to the right.

Figure 4: Aggregate Income Varying the Distortion τ, by β

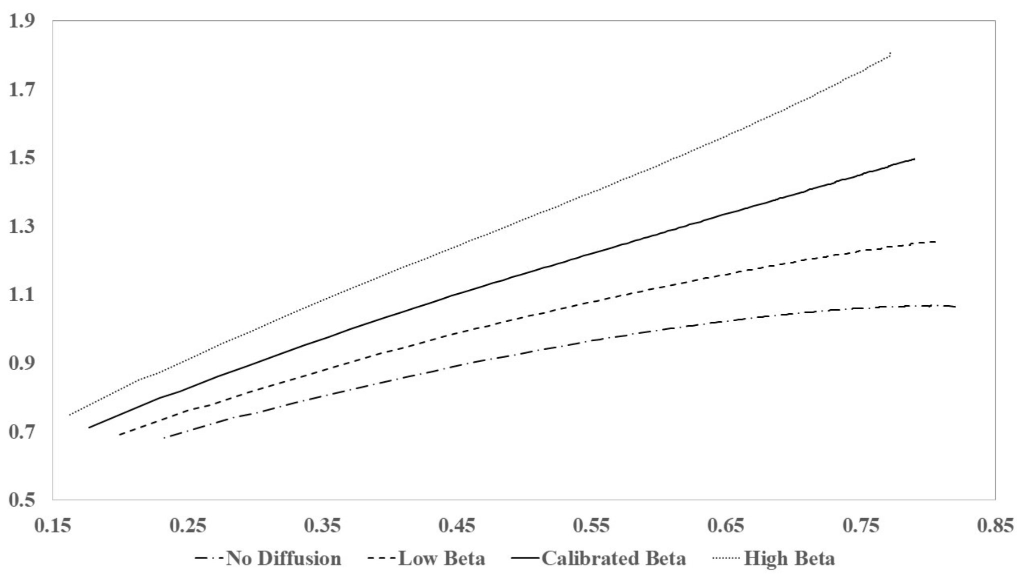


Varying Transmission Parameter Next we compare versions of the previous exercise as we vary τ and consider four values of β. We consider the calibrated value, $\beta = 0.147$, 50% less than the calibrated value, $\beta = 0.074$, 50% more than the calibrated value, $\beta = 0.221$, and no transmission, $\beta = 0$. We then vary τ from 0 to 0.995.[15] The results from this are illustrated in Figure 4.

[15]Note that there is no equilibrium when $\tau = 1$, since no agent would choose to be a worker. Therefore, labor markets cannot clear at any finite wage.

Figure 4 shows that higher values of β imply higher average income, and larger gains from reducing τ. Moreover, this demonstrates that much of the effect that τ has on total income is concentrated at higher values of τ because this is where the extensive margin decision to work or be an entrepreneur is most affected. To illustrate this, we find it informative to instead directly compare income under each value of β to the fraction of the total workforce employed as a worker, which we show in Figure 5. This shows that income is increasing in β for every self-employment share of the workforce, and that the gains to income are increasing in β from moving from one worker share of total labor to a lower one.

Figure 5: Aggregate Income Varying Share of Agents that are Workers, by β



6.3 Gains from New Ideas

For our last counterfactual exercise, we consider a market that is initially in a stationary equilibrium, then measure the effect of adding another group of agents with, on average, higher productivity. We show how the extent of diffusion affects the income of agents from the initial pool.

We calibrate the initial stationary equilibrium in exactly the same way as the previous example. Then we modify the distribution G from which agents draw their initial productivities at birth so that each period a measure δ of new agents draw from the original log-normal distribution, which we label as being of type 1. In addition, a measure 0.1δ of new agents, which we label type 2, draw from a distribution that has a mean that is 8 standard deviations greater than G.[16] Besides the distribution from which they are initially drawn, both types of agents behave in exactly the same way, and both enter the $\widehat{M}$ pool of potential draws.

The spirit of this exercise is that type 1 agents represent the original members of the market that are now being affected by the inclusion of (on average) greater productivity new, type 2 entrants. Type 1 agents are affected in two ways. First, because type 2 agents have, on average, greater productivity, they demand for labor and cause the equilibrium wage to increase. Second, they cause the $\widehat{M}$ distribution to improve relative to the economy with only type 1 agents. The first effect reduces the profit from being an entrepreneur at every productivity level, while the second causes the distribution of realized productivity for type 1 agents to shift to the right. Which effect dominates depends on the extent of diffusion.

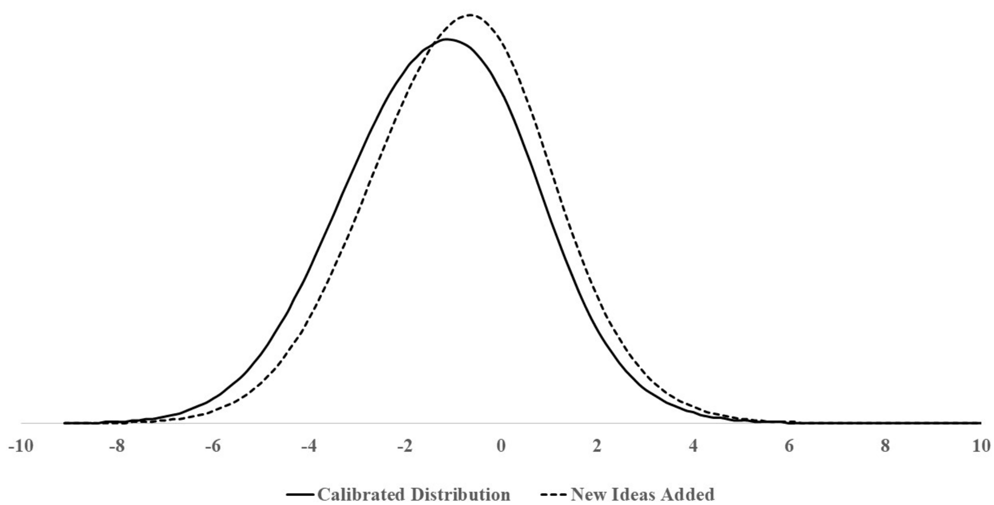
We compute the initial stationary equilibrium and calibrate it as in the previous subsection. We then compute a new stationary equilibrium that includes the type 2 agents. With diffusion, we find that the average profit of operating firms increases by 7.2%. In the same model without diffusion (where $\beta = 0$), average profit decreases by 7.4%. The difference is caused by the endogenous shift in the productivity distribution of type 1 firms in the diffusion model. This is displayed in Figure 6.

7 Discussion and Robustness

Before concluding, we discuss the robustness of our results. In Section 7.1, we consider the identification results in other diffusion models. While we hold fixed the idea that diffusion potentially generates spillovers in the economy, we change some of the other details surrounding occupational choice and the diffusion process. The identification results go through there as well. We further show how to extend the results to more complicated network structures. This requires additional data we do not have, but that has been collected in a number of other studies. Lastly, Section 7.2 compares our results to an identification strategy without using RCT data. It instead relies on only

[16] We readily admit that the choices of the measure of type 2 agents and the difference in their mean initial draw are arbitrary. Changes in these values affect the magnitude of the effects from combining the two groups, but does not affect our qualitative findings.

Figure 6: Log Productivity Distribution of Type 1 Agents, before and after Type 2 added



the panel data from our control group.

7.1 Alternative Model Assumptions: Endogenous Search Intensity and No Occupational Choice

To show the robustness of our results, we change two assumptions in the model: we eliminate the occupational choice, so that the model consists exclusively of entrepreneurs, and we also change the diffusion process. In particular, entrepreneurs can now use one unit of time to (1) learn independently to increase productivity and (2) search for matches to learn from others. Either can be changed independently, we simply combine them to show that the results are robust to various alternative assumptions.

The problem of an agent is now

$$\max_{s,\theta\geq 0} \int \int e^{c+\gamma s+\varepsilon} z^{\rho} \max\left[1, \frac{\hat{z}}{z}\right]^{\rho\beta} d\widehat{M}(\hat{z},\theta) dF(\varepsilon) \tag{7.1}$$

$$\text{subject to:} \quad \theta + s \leq 1$$

Note that s increases productivity directly, with γ governing the effectiveness of this independent learning. Choosing to search for work, on the other hand, changes the distribution of draws from $\widehat{M}$, where $\widehat{M}$ has the same interpretation as in the main text. Taking the first order conditions for an interior solution and rearranging yields

$$\gamma = \frac{\int \max\left[1, \frac{\hat{z}}{z}\right]^{\rho\beta} \frac{\partial \widehat{m}(\hat{z},\theta)}{\partial \theta} d\hat{z}}{\int \max\left[1, \frac{\hat{z}}{z}\right]^{\rho\beta} \widehat{m}(\hat{z},\theta) d\hat{z}} \tag{7.2}$$

We interpret our experiment to do two things. First, participants do not have to pay the cost θ to search for their $\hat{z}$ draw. Therefore, we assume that the treated group chooses $s = 1$. Second, we observe the complete distribution $M(z, \hat{z})$ of matches.

Proposition 3. *The diffusion parameters are uniquely identified under the same RCT data, and utilize the same moments, as the main text.*

7.2 Identifying Model Parameters without RCT Data

[to be written]

8 Conclusion

This paper uses evidence from a randomized controlled trial to identify a model of firm-to-firm transmission of productivity. We find that the transmission of productivity across firms generates quantitatively important amplification to the gains from removing distortions that affect the extensive margin choice between being a worker or an entrepreneur. This is particularly important given the prevalence of self-employment in the developing world. These results provide a stronger link between high levels of self-employment and low levels of income through learning externalities. In particular, the elimination of a tax on wage work more than doubles average income at our estimated parameters relative to a model with no diffusion. We note that many other types of distortions also affect that margin indirectly, such as variable cost distortions or size-dependent policies.

Our procedure starts by directly assuming a class of models used in the recent literature, and shows that (1) experimental data is consistent with the qualitative predictions of such models and (2) the quantitative implications are large. 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. We view this an important first step. The next steps require a more detailed link between model and data. For example, one question that remains unanswered both in this paper and the broader literature is why individuals do not seek out the most productive business owners to learn from, given the seemingly large benefits and low costs we observe (Brooks et al., forthcoming). Beaman and Dillion (2017) point to frictions in the information market, for example. Different field experiments, designed with an eye toward aggregate theory, could provide more detailed information on the law of motion for productivity help further discipline model choices.

References

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.

Bandiera, Oriana and Imran Rasul, "Social Networks and Technology Adoption in Northern Mozambique," *Economic Journal*, 2006, *116* (514), 869–902.

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.

_, Ariel BenYishay, Jeremy Magruder, and A. Mushfiq Mobarak, "Can Network Theory-based Targeting Increase Technology Adoption?," June 2015. Northwestern University Working Paper.

BenYishay, Ariel and A. Mushfiq Mobarak, "Social Learning and Incentives for Experimentation and Communication," March 2017. Yale University Working Paper.

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 Johnson, "Mentors or Teachers? Microenterprise Training in Kenya," *American Economic Journal: Applied Economics*, forthcoming.

Bryan, Gharad, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak, "Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh," *Econometrica*, 2014, *82* (5), 1671–1748.

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*, forthcoming.

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

Cornelissen, Thomas, Christian Dustmann, and Uta Schönberg, "Peer Effects in the Workplace," *American Economic Review*, 2017, *107* (2), 425–456.

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.

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.

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.

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.

Munshi, Kaivan, "Social learning in a heterogenous population: technology diffusion in the Indian Green Revolution," *Journal of Development Economics*, 2004, *73* (1), 185–213.

Nix, Emily, "Learning Spillovers in the Firm," April 2016. Working Paper, USC.

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.

Table of Contents for Online Appendix

A Proofs

A.1 Proposition 1: (β, ρ) in baseline model

We require the following two conditions on the moment values Γ_1 and Γ_2 to prove that $(\rho, \tilde{\beta}) \in (0,1)^2$:

$$\Gamma_1 \in \left(1, 1 + CoefVar(z)^2\right) \tag{A.1}$$

$$\Gamma_2 \in \left(1, \frac{\int \max\left[1, \hat{z}/z\right] dM_H(z,\hat{z})}{\int \max\left[1, \hat{z}/z\right] dM_L(z,\hat{z})}\right) \tag{A.2}$$

where $CoefVar(z)$ is the coefficient of variation of z in Ω.

Define:

$$G_1(\rho, \tilde{\beta}) = \Gamma_1 \frac{\int\int z dM(z,\hat{z}) \int\int z^{\rho} \max\left[1, (\hat{z}/z)^{\tilde{\beta}}\right] dM(z,\hat{z})}{\int\int z^{1+\rho} \max\left[1, (\hat{z}/z)^{\tilde{\beta}}\right] dM(z,\hat{z})} \tag{A.3}$$

$$G_2(\rho, \tilde{\beta}) = \Gamma_2 \frac{\int\int z^{\rho} \max\left[1, (\hat{z}/z)^{\tilde{\beta}}\right] dM_L(z,\hat{z})}{\int\int z^{\rho} \max\left[1, (\hat{z}/z)^{\tilde{\beta}}\right] dM_H(z,\hat{z})} \tag{A.4}$$

Then define:

$$T(\rho, \tilde{\beta}) = \begin{bmatrix} \rho G_1(\rho, \tilde{\beta} \\ \tilde{\beta} G_2(\rho, \tilde{\beta}) \end{bmatrix} \tag{A.5}$$

Last, define:

$$B(\rho, \tilde{\beta}) = (G_1(\rho, \tilde{\beta}) - 1)^2 + (G_2(\rho, \tilde{\beta}) - 1)^2 \tag{A.6}$$

Will proceed 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.

A.2 Proposition 2: θ in baseline model

Proof. Follows immediately from the fact that $\theta_1 > \theta_2$ implies that $\widehat{M}(\cdot;\theta_1)$ first order stochastic dominates $\widehat{M}(\cdot;\theta_2)$. ■