Bringing Data to the Model: Firm-to-Firm Learning in a Structural Model*

Wyatt Brooks
University of Notre Dame

Kevin Donovan
University of Notre Dame

Terence Johnson
University of Notre Dame

February 2018

Abstract

We consider a general equilibrium model in which knowledge diffusion generates positive spillovers in the economy. With some probability, firms receive a random draw from the existing distribution of firm types, some portion of which is internalized in the firm's new productivity. The parameters governing the diffusion process – the likelihood of a draw and the extent to which higher productivity is internalized – are critical for the magnitude of these benefits. We prove that they can be identified with exogenous and random variation in the likelihood and productivity of matches. We then estimate these parameters using the results of a randomized controlled trial among Kenyan firms, in which firm owners from the left tail of the profit distribution are randomly matched one-to-one with owners from the right tail of the profit distribution. Our quantitative results imply an important role for knowledge diffusion. Removing a labor market distortion increases real income by 63 percent at our estimated parameters, compared to 36 percent in an identical model with no productivity transmission.

*Contact Information: Wyatt Brooks, wbrooks@nd.edu; Kevin Donovan, kdonova6@nd.edu; Terence Johnson, tjohns20@nd.edu

1 Introduction

A growing literature recognizes that knowledge or skill transfer can potentially play an important role in the development process.[1] However, almost by its very nature, 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, take seriously the quantitative implications of this knowledge transfer channel, and assess its role for understanding development and policy in the poorest parts of the world.

We do so by building a dynamic general equilibrium model in which agents have entrepreneurial productivity that is affected by the distribution of productivity among operating firms in the economy. In particular, every period agents receive a chance to learn from a randomly selected operating firm with probability η.[2] The productivity of this firm transmits to the agent with some intensity β. At the extremes, if $\beta = 0$ then there is no transmission of productivity and any innovation in productivity is purely random, while if $\beta = 1$, the firm's new productivity is equal to the greater of their own productivity or that of the firm from which they are learning. The model is similar to Jovanovic and Rob (1989), in that it implies a stationary equilibrium in which we characterize the evolution of productivities.

One can immediately see the identification problems. The importance of diffusion as a development-enhancing mechanism depends on the two parameters η and β. Moreover, an agent's skill is explicitly time varying, which eliminates fixed effect methods used in other work, such as Nix (2016) and Cornelissen et al. (2017). Instead we take a different tact. We prove that these parameters are uniquely identified in the presence of randomized changes in the likelihood and quality of matches. Intuitively,

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

[2] This passive learning differs from some of the literature, such as Lucas and Moll (2014).

β is identified by how the treatment effect varies with the profitability of a matched firm. Conditional on β, η is then identified from the overall treatment effect. These results show that the diffusion process can be identified, conditional on having the correct *type* of data.

Such well-suited data is available in Brooks et al. (forthcoming), which we use to parameterize the model. Brooks et al. (forthcoming) conduct a randomized controlled trial among Kenyan business owners in which low-profit owners were matched one-to-one with higher-profit owners, then compared to a control group. We found that profit was 23 percent higher in our treatment relative to control, suggesting an important role for knowledge diffusion.[3] Within the context of the model, the experiment provides us the relevant variation required to identify these parameters. First, it guarantees that the treated entrepreneur receives a draw (rather than only getting one with probability η). Second, we observe the profitability of both the treated firm and the firm it is matched to. These results imply $\beta = 0.15$.

We then use the model to quantify the gains from removing a labor market distortion. In the model, agents choose each period to be a worker or operate as an entrepreneur based on their own entrepreneurial productivity. We introduce a labor market wedge between the cost of labor that employers pay and the income that workers receive, which captures the fact that the rate of entrepreneurship is extremely high in developing countries.[4] When these distortions are high, fewer agents choose to be workers and a large number of them operate low productivity firms. This imposes a negative externality on other firms by changing the distribution of learning draws they receive each period. Higher β implies larger gains from removing the labor market distortion.

Our results imply a important quantitative role for knowledge diffusion. Income rises by 63 percent when this distortion is removed in the $\beta = 0.15$ case, compared to 36 percent in the no-diffusion case with $\beta = 0$. This amplification is monotonic in β so that, for instance, the gains from removing the distortion are 167 percent when β

[3] 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.

[4] This stands in for things like costly job search.

is increased to 0.5.

1.1 Related Literature

This paper joins a small literature using randomized controlled trials to identify critical model parameters and applied to macroeconomic phenomena. Lagakos et al. (2018), for example, use a randomized controlled trial to identify and quantify the benefits of rural-urban migration in Bangladesh. Our focus is on the role of knowledge diffusion. There is a growing literature studying the effects in the context of growth 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 (2015), where agents imitate with only some elasticity. We also do not require agents 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 extend these models to consider international trade in promoting diffusion within and across countries, such as Buera and Oberfield (2015), and Perla et al. (2015). We do not consider this channel here.

2 Model

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 output 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. Defining π^* as the optimal profit after choosing l, the value of having entrepreneurial skill w 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

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

We denote M to be the cumulative density function of z in the mass of agents at any time. The density of productivity conditional on being an entrepreneur is therefore:

$$\widehat{M}(z) = \begin{cases} 0, & \text{if } z < \underline{z} \\ \frac{M(z)-M(\underline{z})}{1-M(\underline{z})}, & \text{if } z \geq \underline{z} \end{cases} \tag{2.4}$$

The wage is determined by the following labor market clearing condition:

$$\int_{\underline{z}}^{\infty} l(z,w)dM(z) = \int_0^{\underline{z}} dM(z) \implies \left(\frac{1-\alpha}{w}p\right)^{\frac{1}{\alpha}} \int_{\underline{z}}^{\infty} z dM(z) = M(\underline{z}) \tag{2.5}$$

Rearranging implies that the market clearing wage is:

$$w = p(1-\alpha)\left(\frac{\int_{\underline{z}}^{\infty} z dM(z)}{M(\underline{z})}\right)^{\alpha}. \tag{2.6}$$

We next discuss the formation and evolution of this distribution M.

2.1 The Evolution of Productivity

Each agent – entrepreneur or not – has productivity z that evolves over time. New agents draw productivity from an exogenous distribution G. Productivity of existing agents evolves as a result of idiosyncratic shocks and from the transmission of productivity across agents. The idiosyncratic shock is a draw ε from an exogenous, time-invariant distribution F. In addition, each period she has a probability η of receiving a draw $\hat{z}$ from the productivities of operating entrepreneurs $\widehat{M}$. That is

$$\hat{z} = \begin{cases} x \sim \widehat{M}, & \text{with probability } \eta \\ z, & \text{with probability } 1-\eta \end{cases}$$

Each agent's productivity in the next period depends on their productivity this period, their idiosyncratic shock ε, and their draw of $\hat{z}$ according to the function

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

Taken together, the distribution M evolves according to the following law of motion:

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

3 Identification of Productivity Transmission

Clearly, the key to the model is the transmission of productivity between firms. We first show that these parameters can be identified theoretically using specific types of variation. We then introduce and discuss a randomized controlled trial in which we induced this type of variation.

3.1 Identification of Model Parameters in Theory

Consider a "treatment" group in which we induce two changes: first, we guarantee a draw this period, so that these firms receive a $\hat{z}$ draw with probability 1 instead of probability η. Second, the draw is from the right tail of $\widehat{M}$ with probability one. We will show that this variation is sufficient to identify the diffusion parameters η and β when compared to a "control" group in which draws are taken from $\widehat{M}$ at rate η. That is, from the economy laid out in the previous section. We show that identification of η comes from the difference in the average profit of treated and untreated firms, while identification of β comes from the difference in treatment effects between those matched matched with highly profitable firms compared to those matched with lower profit firms.

In particular, assuming that we observe the distribution of productivity in the treated group, H, and in the group of matches $\widehat{H}$, then we identify η and β using two treatment effects: the treatment effect of matching random draws from H with random draws from $\widehat{H}$, and the treatment effect of matching random draws from H with a draw from the upper tail of $\widehat{H}$. The second moment identifies β based on the difference in those two treatment effects, then η is identified, conditional on β, from the magnitude of the first treatment effect.

To begin, consider the average profit of members of a distribution H receiving guaranteed $\hat{z}$ matches ($\eta = 1$) from a known distribution $\widehat{H_i}$. Average profit is given by:

$$E[\pi_i] = \alpha p^{\frac{1}{\alpha}} \left(\frac{1-\alpha}{w}\right)^{\frac{1-\alpha}{\alpha}} E\left[ce^{\varepsilon}\max[z, \hat{z}^{\beta}z^{1-\beta}]\right] = \alpha p^{\frac{1}{\alpha}} \left(\frac{1-\alpha}{w}\right)^{\frac{1-\alpha}{\alpha}} cE[e^{\varepsilon}]E\left[\max[z, \hat{z}^{\beta}z^{1-\beta}]\right], \tag{3.1}$$

where the second equality follows from the independence of ε and $\hat{z}$ draws.

Suppose that we observe the average profit from firms treated with draws from two different $\hat{z}$ distributions, $\widehat{H_1}$ and $\widehat{H_2}$. Define R as the ratio of average profits realized from those two treatments. Then:

$$R \equiv \frac{E[\pi_2]}{E[\pi_1]} = \frac{\int\int \max[z, \hat{z}^{\beta}z^{1-\beta}]^{\rho} d\widehat{H_2}(\hat{z})dH(z)}{\int\int \max[z, \hat{z}^{\beta}z^{1-\beta}]^{\rho} d\widehat{H_1}(\hat{z})dH(z)} \tag{3.2}$$

Then the ratio of those treatment effects identifies β for given values of R and ρ, and for known distributions H, $\widehat{H_1}$, and $\widehat{H_2}$. Because profit is proportional to productivity, and because β is identified from a ratio of profits, the distinction between profit and productivity is irrelevant in this model.

Proposition 1. *Suppose H, $\widehat{H_1}$ and $\widehat{H_2}$ are known, and $\widehat{H_2}$ first order stochastic dominates $\widehat{H_1}$. Then for a given $R \geq 1$ and ρ, the value of β that solves (3.2) is unique.*

Proof. Define:

$$\Upsilon(\beta) = \frac{\int\int \max[z, \hat{z}^{\beta}z^{1-\beta}]^{\rho} d\widehat{H_2}(\hat{z})dH(z)}{\int\int \max[z, \hat{z}^{\beta}z^{1-\beta}]^{\rho} d\widehat{H_1}(\hat{z})dH(z)} \tag{3.3}$$

Clearly $\Upsilon(0) = 1$. Therefore, to demonstrate uniqueness, because $R \geq 1$, we need only show that Υ is monotonically increasing. To do so, we show that $\log(\Upsilon(\beta))$ is strictly increasing:

$$\frac{\partial}{\partial\beta}\left[\log(\Upsilon(\beta))\right] = \rho\frac{\int\int_z^{\infty} \log\left(\frac{\hat{z}}{z}\right)[\hat{z}^{\beta}z^{1-\beta}]^{\rho} d\widehat{H_2}(\hat{z})dH(z)}{\int\int \max[z, \hat{z}^{\beta}z^{1-\beta}]^{\rho} d\widehat{H_2}(\hat{z})dH(z)} - \rho\frac{\int\int_z^{\infty} \log\left(\frac{\hat{z}}{z}\right)[\hat{z}^{\beta}z^{1-\beta}]^{\rho} d\widehat{H_1}(\hat{z})dH(z)}{\int\int \max[z, \hat{z}^{\beta}z^{1-\beta}]^{\rho} d\widehat{H_1}(\hat{z})dH(z)} \tag{3.4}$$

We can define new probability measures over $(z, \hat{z})$ for $\widehat{H_1}$ and $\widehat{H_2}$ (with corresponding probability density functions $\widehat{h_1}$ and $\widehat{h_2}$) respectively as:

$$\mu_i(z, \hat{z}) = \frac{\max[z, \hat{z}^\beta z^{1-\beta}]^\rho \widehat{h_i}(\hat{z}) h(z)}{\int\int \max[z, \hat{z}^\beta z^{1-\beta}]^\rho d\widehat{H_i}(\hat{z}) dH(z)} \tag{3.5}$$

where h is the probability density function for H. Then (3.4) can be rewritten as:

$$\frac{\partial}{\partial\beta}\left[\log(\Upsilon(\beta))\right] = \rho \int\int_z^\infty \log\left(\frac{\hat{z}}{z}\right)\mu_2(z, \hat{z}) d\hat{z} dz - \rho \int\int_z^\infty \log\left(\frac{\hat{z}}{z}\right)\mu_1(z, \hat{z}) d\hat{z} dz \tag{3.6}$$

Because $\widehat{H_2}$ first order stochastic dominates $\widehat{H_1}$, μ_2 puts greater probability mass on greater values of $\hat{z}$ than μ_1. Then, since the first and second terms integrate the same function over the same range, and $\log(\hat{z}/z)$ is increasing in $\hat{z}$, then the integral with the measure that puts greater mass on greater values of $\hat{z}$ is larger. Hence, the derivative is strictly positive and we conclude that $\log(\Upsilon(\beta))$ is strictly increasing. Therefore, the β that solves $R = \Upsilon(\beta)$ is unique. ■

In this paper we make use of Proposition 1 by comparing average profit in the overall treatment group to average profit within the subset of the treatment group receiving a guaranteed match from the right tail of the mentor distribution. The overall treatment group corresponds to $\widehat{H_1}$ and the subset of the treatment group receiving a highly productive match corresponds to $\widehat{H_2}$. Although there are many ways to define comparison groups within the set of those receiving mentors, this procedure ensures that first order stochastic dominance is satisfied.

Identification of η, the probability of a $\hat{z}$ draw, comes from comparing treated and untreated firms. In particular, an untreated firm receives a draw with probability η while a treated firm receives one with certainty. Denoting profits of untreated firms as π^U and of treated firms as π^T, the average profit of untreated firms is given by:

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

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

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

Rearranging this equation identifies η as summarized in the following 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 η is given in closed form by:*

$$\eta = \frac{\frac{E[\pi^U]}{E[\pi^T]}\int\int \max[z, \hat{z}^\beta z^{1-\beta}]^\rho d\widehat{H}(\hat{z})dH(z) - \int z^\rho dH(z)}{\int\int \max[z, \hat{z}^\beta z^{1-\beta}]^\rho d\widehat{M}(\hat{z})dH(z) - \int z^\rho dH(z)} \tag{3.9}$$

3.2 Identification of Model Parameters in Practice

The previous section shows that we can identify diffusion parameters using variation in the data. Here, we summarize a randomize controlled trial in which such variation exists. A complete description of the program and results is given in Brooks et al. (forthcoming). The goal of this study was to determine if managerial capital could be transmitted between entrepreneurs. To do this, our experimental design randomly matched older, profitable entrepreneurs to serve as mentors to 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.

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 approximately 3300 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. We use this baseline survey to draw our sample of young entrepreneurs that we target for treatment as well as our sample of older, profitable entrepreneurs that could serve as mentors. In addition, this baseline sample gives us estimates of moments of the population of operating firms.

Mentors were selected based on age, experience and profitability. From the initial baseline sample we kept all firms with owners over 40 years old with at least 5 years of experience. Then within this group we sorted based on profitability, and recruited mentors starting with the highest profit firms and moving downward. Of those contacted to serve as a mentor, 95% accepted. We reached a sufficient number of mentors at the 51st percentile of our recruitment frame. Therefore, the set of mentors can be viewed as representing the upper half of profitable firms operated by older, experienced entrepreneurs.

Mentors and mentees met weekly in one-on-one meetings to discuss business. The content of their discussion was self-directed, and we provided very little guidance. Instead, mentors were encouraged to provide whatever advice they thought would be most valuable. The official treatment period was only four weeks wherein mentors and mentees met regularly. However, the majority of mentor-mentee matches persisted much longer than this period. After a year, approximately half of pairs were still meeting.

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. The effects on profit 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% of the control group's mean profit. Meanwhile, the group that received business training classes realized no significant increase in profit. The point estimate of the effect of classes is equal to 4% of control group mean profit.

In the second column of Table 1, we break out these results by the profitability of the mentor. 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

Profit	(1)	(2)
Mentee:	414.46	
	(133.07)**	
Mentee:		356.03
mentor in $(0, 25)$ pctile		(180.17)**
Mentee:		449.11
mentor in $(25, 75)$ pctile		(172.01)***
Mentee:		514.54
mentor in $(75, 100)$ pctile		(270.54)*
Class	75.25	74.03
	(147.34)	(142.49)
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, though results are robust to other (or no) trimming procedures. Statistical significance at 0.10, 0.05, and 0.01 is denoted by *, **, and, ***.

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.

Finally, we used our selection procedure for mentors to look for effects on the mentors from engaging in the program. We surveyed both the mentors and a group of firms just below the 51st percentile cutoff that we used to select mentors. Then we employed a regression discontinuity design to look for a difference in profits around the 51st percentile cutoff generated by our intervention. We found no evidence of such an increase in profit, which shows that mentor profit did not change as a result of their participation in this program. This is consistent with the model described in Section 2 where higher productivity firms receive no benefit from interaction with lower productivity firms.

4 Quantitative Exercise

Our quantitative exercise considers a distortion to the choice to become a worker or entrepreneur. As emphasized in Gollin (2008), in developing countries the fraction of the labor force that is self-employed is vastly higher than in developed countries. 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.[5] Now the cutoff value of $\underline{z}$ is determined by an indifference condition that takes the wedge into account:

$$(1-\tau)w + T = \pi(\underline{z}, w) + T \implies \underline{z} = (1-\tau)\left(\frac{w}{p}\right)^{\frac{1}{\alpha}} \frac{(1-\alpha)^{\frac{\alpha-1}{\alpha}}}{\alpha} \tag{4.1}$$

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 externality 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

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

operates as entrepreneurs. We then measure the effect of removing the distortion by comparing total income in the stationary equilibrium with high τ to total income in the stationary equilibrium where $\tau = 0$.

4.1 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.

Identification of the parameters of the productivity transmission process, η and β, is discussed extensively in the previous section. In particular, here these parameters are identified from the overall treatment effect from mentors (equal to 23.0% of control group mean profit) and the treatment effect from receiving a mentor from the top quartile of the mentor profit distribution (equal to 29.9% of control group mean profit).

We make use of the results in Propositions 1 and 2 as follows. For the purposes of identifying η and β, we assume that firm productivity is Pareto-distributed with shape parameter θ and scale parameter x_0. We choose θ 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. We also assume that $\rho = 0.8$. Applying Propositions 1 and 2 in this way implies that $\beta = 0.147$ and $\eta = 0.774$.[6]

Next, we parameterize 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

[6]Notice that this procedure has the useful property that the only other parameter value needed to identify β and η is ρ. In this sense, the calibrated values of the productivity transmission process are not dependent on the calibration strategy used to derive the other parameters.

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 $\alpha = 0.583$.

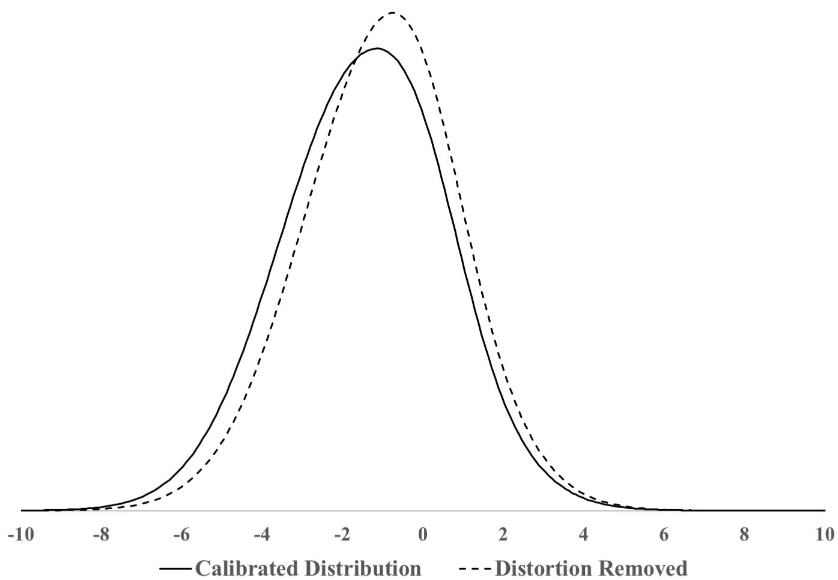
We choose the price of output p to be numeraire. Our remaining parameters are σ, σ_0, c and τ. We jointly match these four parameters to the following four moments: the variance of log-profit in the overall population of operating firms (1.399), the variance of log-profit among new entrants (0.961), 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%).[7] These moments are matched when $\sigma = 1.389$, $\sigma_0 = 0.985$, $c = 1.499$ and $\tau = 0.983$.

4.2 Gains from Removing Distortion

Our main exercise is to compute the stationary equilibrium under the calibrated value of τ, and the stationary equilibrium with $\tau = 0$ and compare total real income in the two equilibria. When β is at its calibrated value, we find that real income increases by 62.55% from completely removing the distortion. These gains come from two sources. First, reducing the distortion makes more labor input available for the high

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

Figure 1: Log Productivity Distribution, before and after distortion removed



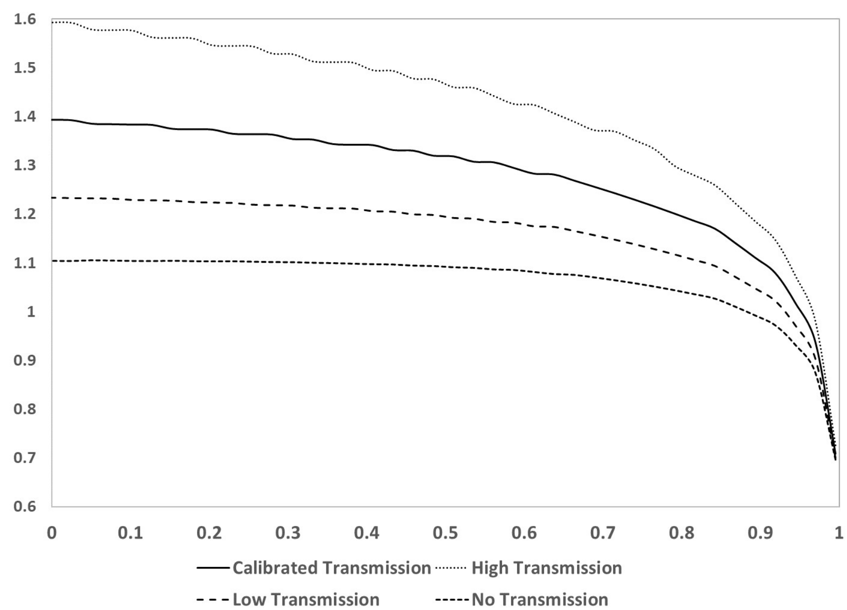
productivity entrepreneurs both by increasing the supply of workers, and by reducing labor demand from low productivity entrepreneurs. This allows higher productivity entrepreneurs to increase their scale and increase total output. Second, because of the transmission of productivity, the distribution 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 1.

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 35.63% 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 (62.55% versus 35.63%), 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 1. 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.

4.3 Varying Transmission Parameter

Figure 2: Aggregate Income Varying the Distortion τ, by β



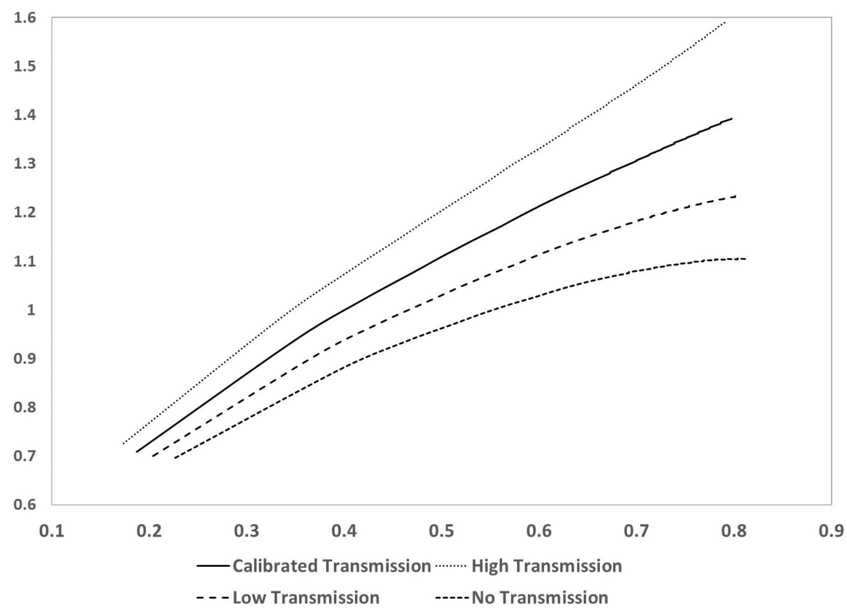
Next we compare versions of the previous exercise as we vary τ and consider four values of β. We consider the calibrated value, $\beta = 0.147$, 50% less than the calibrated value, $\beta = 0.074$, 50% more than the calibrated value, $\beta = 0.221$, and no transmission, $\beta = 0$. We then vary τ from 0 to 0.995.[8] The results from this are illustrated in Figure 2.

Figure 2 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

[8] 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.

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 3. 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 3: Aggregate Income Varying Share of Agents that are Workers, by β



5 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.

Here we consider a particular type of distortion that directly affects the choice to be a worker or entrepreneur. However, many other types of distortions also affect that margin indirectly, such as variable cost distortions or size-dependent policies. That is, when highly productive firms are distorted below their optimal labor choice, equilibrium wages are reduced and more agents choose self-employment. In future work, we will also consider these types of distortions.

References

Brooks, Wyatt, Kevin Donovan, and Terence Johnson, "Mentors or Teachers? Microenterprise Training in Kenya," *American Economic Journal: Applied Economics*, forthcoming. forthcoming.

Buera, Francisco J. and Ezra Oberfield, "The Global Diffusion of Ideas," June 2015. Chicago Fed Working Paper.

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

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.

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

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.

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.