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

December 2018

Preliminary Draft – Comments Welcome

Abstract

We consider an increasingly utilized class of general equilibrium models in which knowledge diffusion generates positive spillovers across firms. Each period a firm "matches" with another firm, and can internalize some portion of the matched firm's productivity. Within this class of models, we prove that a small set of parameters characterizing the diffusion process are uniquely identified with exogenous and random variation in matches, and moreover, are independent of other parameter values and remaining model structure. We 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 diffusion parameters. We then embed the estimated process into a general equilibrium model to assess the quantitative importance of diffusion. The efficient level of learning spillovers increases income by 9 percent relative to a *laissez faire* policy. Moreover, diffusion benefits can overcome competitive pressure faced by incumbent firms when more productive new firms enter (e.g., via foreign firm entry, microenterprise training)

*Thanks to conference and seminar participants at Arizona State, Columbia, Illinois, Notre Dame, NYU, Stanford, UBC, and the Society for Economic Dynamics for helpful comments, especially Jesse Perla, Kevin Song, and Meredith Startz. *Contact Information:* Brooks, wbrooks@nd.edu; Donovan, kevin.donovan@yale.edu; Johnson, tjohns20@nd.edu

1 Introduction

A growing literature focuses on how knowledge or skill transfer among firms could play an important role in the development process (Lucas, 2009; Perla and Tonetti, 2014; Perla et al., 2015; Buera and Oberfield, 2017). 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 economic development. Relatedly, since knowledge is non-rival and generates positive spillovers, the magnitude of any policy depends crucially on the size of these externalities. Our goal in this paper is to take this class of models to the data, identify key diffusion parameters, and quantitatively assess the magnitude of these diffusion externalities.

Specifically, we consider an economy in which an agent with productivity z_t at time t has future productivity z_{t+1} given by the law of motion

$$\log(z_{t+1}) = c + \rho \max\{z_t, \hat{z}_t^{\beta} z_t^{1-\beta}\} + \varepsilon_t,$$

where c is a constant growth rate, ρ is the persistence of productivity, and ε is an exogenous shock. Diffusion occurs via the term $\hat{z}$. Each period, an agent gets the opportunity to "imitate" another agent in the economy, with that draw $\hat{z}$ from some matching function $\widehat{M}$.[1] If sufficiently beneficial, she can adopt some portion β of that new productivity, or disregard it.

A quantitative assessment of diffusion requires estimation of this process. Specifically, it requires the parameter β, which governs the elasticity with which one can internalize a match's productivity, and ρ, which governs how long the gains from a good match will persist. Moreover, the matching function $\widehat{M}$ may have relevant features that impact the quantitative importance of diffusion. For example, if agents can easily direct their search toward high productivity matches, diffusion may occur more rapidly. We allow for this by indexing $\widehat{M}$ with a parameter θ that governs such possibilities.[2]

[1] While left explicitly vague for now, $\widehat{M}$ is generally an equilibrium object. For example, the literature cited above assumes that $\widehat{M}$ generates uniformly random draws from the existing firm productivity distribution, though there are many other possibilities.

[2] As we discuss later, this is not the only interpretation of θ. We consider other examples in which agents are required to expend effort to improve their draws, and cases in which diffusion depends on strategic interaction among agents.

One can immediately see the identification problems inherent in such an exercise. First, the notion of productivity is broad. Unlike some recent work tracking the diffusion of specific technologies or pieces of information, this more "macro" measure encompasses a wide variety of different skills and information that can potentially impact another individual.[3] A corollary of that fact is that we cannot observe matches in the economy. Thus, the diffusion process is governed by functions whose inputs have no direct empirical counterparts. One potential solution to these issues is to identify individual-level productivities using fixed effect methods, but the explicit time varying productivities induced by diffusion generally make such techniques insufficient in this class of models.

In this paper we therefore 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 randomly select a subset of the equilibrium population and systematically vary the set of individuals from which they draw their matches, differences in profit identify the key diffusion parameters.

The identification results come from the two levels of randomization built into this (hypothetical) RCT. While the first step randomizes into treatment and control groups, the second step assigns specific – but randomly chosen – matches to each treatment firm. This allows us to use within-treatment variation to identify the intensity of diffusion, denoted above by β. Intuitively, if treated firms with high productivity matches have similar *ex post* profit to those who do not, it must be because productive matches do not translate into own-productivity. That is, β must be low. Thus, this parameter is identified by a straightforward measure of heterogeneity in the treatment effect. Working backwards, the first randomization into treatment and control groups allow us to identify θ. Of course, we cannot observe individual control matches. However, since our treatment guarantees a high quality match, small *ex post* differences in profit between treatment and control suggest that control firms are already quite

[3] Conley and Udry (2010) study the introduction pineapples in Ghana, Banerjee et al. (2013) study the diffusion of microfinance in India, and BenYishay and Mobarak (forthcoming) study the diffusion of pit planting in Malawi agriculture. In contrast, studies in the vein of Jovanovic and Rob (1989), such as Perla and Tonetti (2014) and Buera and Oberfield (2017), study a much broader notion of productivity that is not related to a specific technological innovation.

good at "directing" their search toward to the right tail of the distribution. Thus, θ is identified directly from the average treatment effect.[4]

An important implication of our theoretical results is that parameter estimates we obtain are independent of much of the remaining model structure. This is in large part due to the power of using randomized variation to identify the parameters. For example, our identification of the elasticity of match productivity to own productivity (β) uses only within-treatment variation. Thus, the details of who usually matches with who, and why matches occur, are irrelevant. Our estimates can therefore be embedded into a variety of models with different assumptions on the remaining structure of the economy. We emphasize this by specifying only the required assumptions for the identification results in Section 2, leaving the details of the remaining model structure for Section 4, when we require it for the quantitative results.

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 which we introduce the required variation. We identify a set of high productivity business owners, then randomly match them one-to-one with treatment firms, and compare these treated firms to a control group. 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 specific topics to discuss. The 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. (2018). Profits were 19 percent higher in the treatment group relative to the control, and increasing in the relative profit gap between the two firms entered into a match.[5] 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 build our results into a general equilibrium model with occupational choice, in which the match function (denoted $\widehat{M}$ above) is an equilibrium object. We conduct a series of model experiments to measure the importance of the diffusion chan-

[4]Note that we have not discussed the identification of persistence, ρ, despite the fact that it plays an important role. Intuitively, any procedure that requires observation after a match occurs will depend on ρ, since that parameter controls the decay of productivity. We save the details of that discussion for the main body of the paper.

[5]In principal, a change in profit could be driven by any number of channels, including profit sharing, loans, bulk discounts, etc. While interesting in their own right, these channels would not necessarily be consistent with the diffusion channel we highlight here. We consider a number of these possibilities in Brooks et al. (2018) and reject them.

nel given our estimated parameters. First, we consider the efficient level of learning externalities. Low productivity firms impose a cost on the economy, as their existence decreases the likelihood a firm matches with a high productivity firm. The efficient allocation increases average income increases by 9 percent, relative to an economy with no intervention. The magnitude is in large part generated by our diffusion parameters. The quantitative magnitude is primarily driven by the complementarity in the diffusion parameters, suggesting the measurement of these values is critical to understanding the importance of diffusion.

Second, we consider a counterfactual in which a pool of high productivity firms is added to the initial steady state set of firms.[6] These new firms increase labor demand, which increases the equilibrium wage. All else equal, this lowers profits from incumbents. However, it also allows these same incumbents 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 4 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.

1.1 Related Literature

This paper joins a relatively new literature that uses causal empirical estimates to identify critical model parameters in macroeconomic models. Kaboski and Townsend (2011) use exogenous variation in micro-loans to structurally estimate a model of borrowing constraints and entrepreneurship. Brooks and Donovan (2017) use exogenous variation in infrastructure placement to identify a model of risky farm investment. Our paper shares a similar style but focuses on knowledge diffusion. Closest in this dimension is Lagakos et al. (2018), who use the results from a randomized controlled

[6]This exercise is motivated by recent work in which opening to trade can have negative equilibrium effects on incumbents (Asturias et al., forthcoming) but also generates learning opportunities (Perla et al., 2015; Buera and Oberfield, 2017). In a more micro-development context, recent work on firm training studies the extent to which the gains from trained firms comes at the expensive of untrained firms (McKenzie and Puerto, 2017).

trial to, in part, identify the utility cost associated with migration. We share a similar goal of using a randomized control trial to identify parameters not directly observable in data.

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

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

2 Identification of the Diffusion Process

We begin by specifying the class of diffusion processes we will study. The goal here is to lay out the required assumptions without the details of the full model in which we will eventually embed the diffusion process, as they are both cumbersome and unnecessary for our main identification results. Therefore, along the way, we will draw attention to the required assumptions so that is clear what is required for the results.

2.1 Setting Up the Problem

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

Each period, every agent receives two types of shocks to their productivity. First, they receive an idiosyncratic imitation shock $\hat{z}$. If their own productivity z is greater than $\hat{z}$, then the imitation opportunity is useless and it has no effect on the agent's future productivity. If $\hat{z} > z$, then the imitation opportunity contains some useful information that the agent can incorporate into their own future productivity. The intensity with which this imitation opportunity transmits to the agent's productivity in the subsequent period is governed by the parameter β. Second, firms receive random shocks ε that enter the next period's productivity multiplicatively. This shock is assumed to be uncorrelated with own productivity z or the imitation draw $\hat{z}$.[7] The functional form of the subsequent productivity z' is given by Assumption 1.

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

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

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

Notice that the max operator in the diffusion process rules out any productivity benefit accrued to a higher productivity firm from interaction with lower productivity firms (as in Jovanovic and Rob, 1989, for example). We address this issue directly in Section 3 and find no evidence that more productive firms gain profit from interaction with less productive firms.[8] Furthermore, it subsumes the perfect imitation case as $\beta = 1$, while allowing for possible imperfect adoption ($\beta < 1$).

Given the notion of productivity we consider here, we cannot observe it directly. Thus, we require a link between productivity and observable variables, in the case,

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

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

profit. The requirement is summarized in Assumption 2.

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

This assumption is satisfied by much of the literature on diffusion. A simple way to satisfy Assumption 2 is to assume $\pi_i = z_i$ as in Lucas (2009) and Perla and Tonetti (2014). A production function of the form $y = z^{\alpha}n^{1-\alpha}$ also satisfies Assumption 2.[9]

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

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

This assumption admits a variety of search and assignment processes. For example, one commonly used diffusion process is that agents draw randomly from the existing firms. Denoting M as the cdf of operating-firm productivity, this would imply $\widehat{M}(\hat{z}; z, \theta) = M(\hat{z})$. Even within the random search framework, Assumption 3 allows us to be somewhat broader, as agents may draw from better or worse distributions that the set of operating firms, where θ indexes how much the distribution of matches differs from the firm productivity distribution. We discuss this assumption in more detail in Section 2.5.

Based only on the assumptions presented in this subsection, there are two parameters crucial to understanding the diffusion process. First, the parameter β controls the intensity of diffusion from imitation opportunities to future productivity. Second, the parameter θ controls the distribution of draws of imitation opportunities. As θ

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

gets larger, agents are able to draw better imitation opportunities. Varying the values of β and θ allows for a range of possibilities. If both are very low, then diffusion plays very little role in firm dynamics. If β is high but θ is low, then agents could benefit greatly from useful imitation opportunities, but such opportunities are rare. If θ is high while β is low, then imitation opportunities are readily available but of limited usefulness. Finally, if both values are high, then diffusion plays a very important role in firm dynamics as firms are readily able to access useful imitation opportunities and those opportunities have large effects on firm productivity.

In the remaining parts of this section we discuss the variation needed to identify these parameters and then outline conditions under which we can obtain unique estimates of their values.

2.2 Variation Required to Identify Diffusion

Section 2.1 laid out a set of assumptions on the primitives of the model. We next turn to the variation in the data required to identify the parameters in the context of that framework. This is summarized in Assumption 4. After proving the identification results, Section 3 details a randomized controlled trial that satisfies these assumptions, thus allowing us to take the model to the data.

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

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

The first assumption imposes the usual exclusion restriction – that unobserved characteristics do not systematically vary across the two groups. The second formalizes the intuitive notion that we cannot observe control group matches, and they proceed

as defined by the $\widehat{M}$ function. That is, control group continues to match as defined by the underlying economy.[10] Finally, the third and fourth lay out what we require from our treatment. The third states that we can observe all matches, and those matches are drawn from some other distribution than the control group. Finally, the last assumption states a second exclusion restriction *within* the treatment group. We can break the treatment group into two arbitrary groups, and comparisons across these two groups are unbiased.

With these assumptions on the data generating process, we turn to identification. We first lay out an intuitive explanation of how the parameters are identified, then lay out the formal proofs of existence and uniqueness.

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

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

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

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

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

(Assumption 1) implies

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

Denote $\Gamma_1 \equiv \bar{z}_T^H / \bar{z}_T^L$. Then a bit of algebra on the above equation implies

$$\Gamma_1 = \frac{\int\int \max\left[z, \hat{z}^\beta z^{1-\beta}\right]^\rho d\widehat{H}_T^H(\hat{z}, z) dH_T^H(z)}{\int\int \max\left[z, \hat{z}^\beta z^{1-\beta}\right]^\rho d\widehat{H}_T^L(\hat{z}, z) dH_T^L(z)} \tag{2.2}$$

where the exogenous shocks ε cancel out via the within-treatment exclusion restriction in Assumption 4. While equation (2.2) seems somewhat complicated, our assumptions leave few unknowns. First, Γ_1 is directly observed in the data, as Assumption 2 implies we can read this ratio directly from the profit ratio of agents. Second, all terms on the right hand side are directly observable except β and ρ. This follows from the power of Assumption 4 and the fact that we are only using variation among treated agents. Thus, equation (2.2) *almost* gives us a moment to identify β. Instead, it gives us a moment to identify β conditional on the value of ρ.

This argument provides the fundamental logic of how β is identified, but is incomplete for two important reasons. First, even given a value of ρ, equation (2.2) is highly non-linear. Thus, neither existence or uniqueness of β is guaranteed. Second, the explanation above shows that we are also required to estimate ρ along with β.[12] After laying out a similar intuition for identification of θ, we return to the formal statement of our identification results.

"Directedness" of diffusion θ Next, we discuss the identification of θ, which controls the distribution of imitation draws. Here, we run into the issue that control firm draws are not known. The critical insight here, however, is that we can observe the average draw of both control and treatment groups, and can use these to identify θ. Take, for example, the case in which treatment draws are guaranteed to be high productivity. If we observe small differences in average *ex post* profit, it follows that control firms must also be drawing from a distribution with substantial mass on high productivity matches. That is, $\widehat{M}$ must be indexed by a high θ. Similarly, large differences in

[12] The simple intuition for why identification of β depends on ρ is that the moment defined in equation (2.2) relies on measurement of productivity *after* the match occurs. Since ρ controls the decay in productivity, any measurement with a time dimension will depend on ρ.

average profit suggest a low θ. This average difference in profit therefore allows us to infer what θ is, despite not observing the underlying matches in the control group.

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

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

As before, defining $\Gamma_2 \equiv \bar{z}_T/\bar{z}_C$ and applying the same logic to the exogenous shocks ε, we can re-write the equation as

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

Given the values of β and ρ have already been identified (proven formally in the next subsection), then all other parts of this equation come directly from the data, except for the parameter θ.[13] The assumed monotonicity of $\widehat{M}$ in θ is sufficient to prove that any θ that solves this equation is unique. We discuss the requirements for existence below.

This subsection laid out an intuitive procedure to identify the diffusion parameters β and θ, along with persistence ρ. The next two subsections formalize the existence and uniqueness conditions first for (β, ρ) in Section 2.3 then for θ in Section 2.4.

2.3 Existence and Uniqueness of (β, ρ)

We now formalize the argument made above on how to identify the diffusion intensity parameter β. As demonstrated previously, identification of β requires identification of ρ. It remains to be shown how that is achieved, and that the (β, ρ) pair exists and is unique.

Identification of ρ requires a second moment from the data. The strategy we employ is similar to that used in standard models of firm dynamics with AR(1) processes, but

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

with one important difference: in the presence of diffusion, the imitation opportunity itself must be accounted for. To see this, note that we can write the law of motion for diffusion as

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

If one were then to run a standard lagged regression

$$\log(z') = \eta + \gamma \log(z) + \nu, \tag{2.4}$$

the estimate $\widehat{\gamma} \neq \rho$ because of the omitted variable bias induced by the fact that the error $\nu = \beta\rho \log\left(\max\{\hat{z}/z, 1\}\right) + \varepsilon$. Thus, to measure ρ, we need to control for this bias. The ability to observe all matches in our treatment group allows us to do exactly this. Denoting $\Gamma_3 \equiv 1 + \frac{Var(z)}{\mathbb{E}[z]\mathbb{E}[z']}\widehat{\gamma}$ as our empirical moment, where $\widehat{\gamma}$ is the coefficient from (2.4), some algebra on the law of motion for productivity implies that we can write the theoretical counterpart to Γ_3 as

$$\Gamma_3 = \frac{\int\int z^{1+\rho} \max\left[1, \frac{\hat{z}}{z}\right]^{\rho\beta} d\widehat{H}_T(\hat{z}, z) dH_T(z)}{\int z dH_T(z) \cdot \int\int z^{\rho} \max\left[1, \frac{\hat{z}}{z}\right]^{\rho\beta} d\widehat{H}_T(\hat{z}, z) dH_T(z)}. \tag{2.5}$$

Note that the matches enter explicitly into (2.5) to take into account the bias induced diffusion. Thus, the observability of matches in the treatment implies that (2.5) identifies ρ up to the value of β. Taken together, we now have two moment conditions and two unknown parameters. It remains to be shown that they have a solution and that it is unique. This is done in Proposition 1.

Proposition 1. *Suppose Assumptions 1 and 4 are satisfied. If the following two conditions hold, then there exists a unique pair* $(\beta^*, \rho^*) \in (0,1)^2$ *that solve equations (2.2) and (2.5). Those conditions are:*

$$\Gamma_1 \in \left(1, \frac{\int\int \hat{z} d\widehat{H}_T^H(\hat{z}, z) dH_T^H(z)}{\int\int \hat{z} d\widehat{H}_T^L(\hat{z}, z) dH_T^L(z)}\right) \tag{2.6}$$

$$\Gamma_3 \in \left(1, 1 + CV(z)^2\right) \tag{2.7}$$

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

The proof is relegated to Appendix D. Notice that the addition of Assumption 2

implies the immediate corollary that the conditions in Proposition 1 may be stated in terms of profit instead of productivity.

This demonstrates two important things about the identification of β and ρ. First, not all possible values of Γ_1 or Γ_3 can be rationalized by the model. For instance, if better imitation opportunities led to worse performance, no parameterization could account for that. Second, these two parameters are identified without the need to specify broader elements of the model. No assumption have been made about price determination, the origin of the set of operating firms or if the economy is in a steady state. The identification of these parameters does not depend on those issues. Therefore, any model satisfying Assumption 1 can make use of estimates that arise from the strategy in this paper, and any data satisfying Assumption 4 can be used to implement that strategy.

2.4 Existence and Uniqueness of θ

The last part of the identification is the uniqueness of θ. Like (β, ρ), not every possible

Here we formalize the argument made following the derivation of equation (2.3). First, note that the model cannot rationalize all possible values of Γ_2. We can compute the bounds of the set of moment values that can be rationalized in the model explicitly as

$$\Gamma_2^{min} = \inf_{\theta} \frac{\int\int \max\left[z, \hat{z}^{\beta} z^{1-\beta}\right]^{\rho} d\widehat{H}_T(\hat{z}, z) dH_T(z)}{\int\int \max\left[z, \hat{z}^{\beta} z^{1-\beta}\right]^{\rho} d\widehat{M}(\hat{z}; z, \theta) dH_C(z)} \tag{2.8}$$

$$\Gamma_2^{max} = \sup_{\theta} \frac{\int\int \max\left[z, \hat{z}^{\beta} z^{1-\beta}\right]^{\rho} d\widehat{H}_T(\hat{z}, z) dH_T(z)}{\int\int \max\left[z, \hat{z}^{\beta} z^{1-\beta}\right]^{\rho} d\widehat{M}(\hat{z}; z, \theta) dH_C(z)}. \tag{2.9}$$

Under Assumption 4, the densities $\widehat{H}_T$, H_T, and H_C are observed. Moreover, Proposition 1 (under Assumption 1) implies values of β and ρ. Finally, Assumption 3 says that $\widehat{M}$ is a known function. Hence, both Γ_2^{min} and Γ_2^{max} can be solved explicitly.[14] Then identification of θ follows immediately by the following proposition.

Proposition 2. *Under Assumptions 2, 3 and 4, if $\Gamma_2 \in (\Gamma_2^{min}, \Gamma_2^{max})$, then a solution to equation (2.3) exists and is unique.*

[14] Notice the strong interdependence between the values of β and ρ, and the bounds of the identifiable set of θ. At the extreme, if $\beta = 0$, then $\Gamma_2^{min} = \Gamma_2^{max}$ and θ is not identified as strict monotonicity fails in the proof of Proposition 2.

The proof is Appendix D, but follows almost immediately from first order stochastic dominance assumption embedded in Assumption 3. With a slight abuse of notation, Assumption 3 implies that there is a unique function $\theta(\beta, \rho)$ that solves (2.3). Combining that with Proposition 1 then guarantees the value θ is unique. Taken together, Propositions 1 and 2 show that under our assumptions, the variation induced here uniquely identify the diffusion parameters (β, ρ, θ). After a brief discussion of the procedure below, we will turn to a randomized controlled trial that induces such variation.

2.5 Discussion

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

First, note that the empirical moments required for the estimation are are easily obtained from data. We require only a measure of treatment heterogeneity (Γ_1 to identify β), the (normalized) coefficient from a lagged profit regression (Γ_3 to identify ρ) and the average treatment effect (Γ_2 to identify θ). Thus, the moments allow for a relatively straightforward movement between model and data.

A second question is the extent to which our identification results are robust to other economies, or put differently, where our required assumptions fail. First, the identification of β and ρ holds in nearly all economies. The key here is the power of the (hypothetical) design. That the matches are randomized – and observable – within the treatment group implies only treatment firm data are required to identify (β, ρ). Thus, the details of who searches, or why, is irrelevant for the estimation (conditional on Assumptions 1 and 2).

A more subtle restriction is built into our assumption on $\widehat{M}$ in Assumption 3. Here, we require that the draw of a match $\hat{z}$ depends only on z and a parameter θ. This assumption nests as a special case work by Jovanovic and Rob (1989) and Lucas (2009), and Buera and Oberfield (2017), who assume uniform draws from the existing distribution of operating firms. In that case, if M is the cdf of operating firm productivity, $\widehat{M}(\hat{z}; z, \theta) \equiv M(\hat{z})$.[15] Lucas and Moll (2014) and Perla and Tonetti

[15]Of course, M is generally an equilibrium object. That fact is irrelevant for the identification, but matters more

(2014) make the same assumption, but extend these models by endogenizing the cost of receiving a draw. These models generally fail Assumption 3, because the decision to search depends on the remaining details of the model and equilibrium. In this case, we lose the independence of θ from the remaining model structure.[16]

We emphasize two features of our strategy however. First, even in these cases, the identification of (β, ρ) is still possible. Thus, if one was willing to accept a fixed θ (as in Lucas and Moll, 2014 and Perla and Tonetti, 2014), the remaining identification would still hold. Anticipating the results, we show that the value of θ has little quantitative impact and thus fixing it at an arbitrary value *ex ante* has little impact. Second, Assumption 3 still allows for a wide set of underlying matching processes. For example, in addition to random search, the assumption admits pure assignment models as well. We detail a number of different underlying models that satisfy our assumptions in Appendix A.

3 Data and Estimation

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

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

when we embed our estimates in a general equilibrium model in Section 4.

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

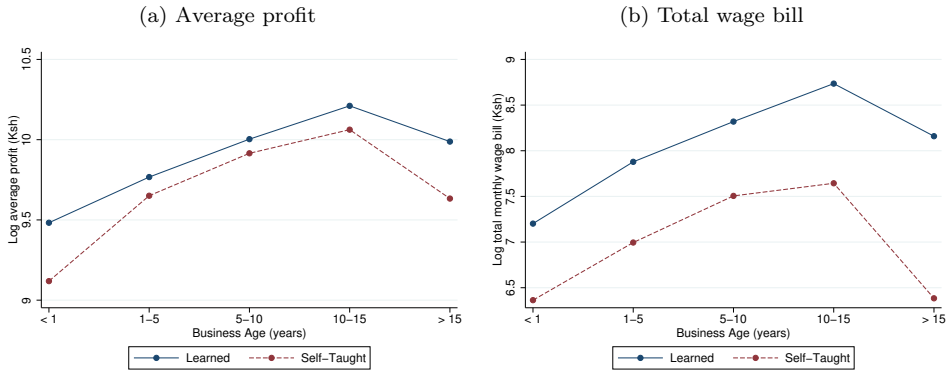
[17] In Brooks et al. (2018), we further randomize another group into formal business training classes. While we do not utilize this classroom training treatment arm here, it is interesting to note that the results differ substantially across these treatment arms. We show that this to the fact that matching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

3.1 Details of RCT and Data Collection

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

Qualitative Evidence on the Importance of Learning To begin, Figure 1 plots business scale measures based on self-reported learning methods from the baseline survey. Fifty-five percent of all firms claimed they were self-taught, while the rest claimed to learn either from another business operator, in school, or through an apprenticeship. Figure 1a shows that the self-taught earn less profit at any point over the lifecycle. The average profit of a self-taught firm is 18 percent less than firms that learn from others, while Figure 1b show that self-taught firms pay a smaller total wage bill.

Figure 1: Self-Reported Learning Methods and Business Scale



Selection and Randomization We start from a sample of female business owners who have been in operation for less than 5 years.[18] We then randomly select a subset of these business owners to randomly match with an older, more experienced owner. In

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

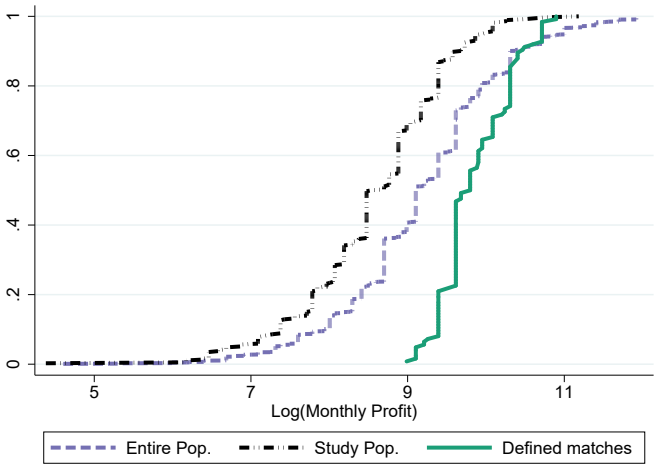
this way, we guarantee a high quality match for these business owners (in an intent-to-treat sense). Thus, the randomization allows us to compare these young owners chosen into the treatment against those other young business owners who were not. These individuals were then surveyed 7 times over 17 months, at $t = 1, 2, 3, 4, 7, 12, 17$.[19,20]

The older business owners who entered into a match were selected from those businesses with owners over 40 years old and at least 5 years of experience. This hopefully minimized the importance of "luck" in baseline profit realizations to allow us to focus on truly productive business owners.

We then recruited business owners with the highest profit until we had a sufficient number for matches. Of those contacted to serve as a mentor, 95 percent accepted. We reached a sufficient number of mentors at the 51st percentile of our recruitment frame. These matches with treatment firms were random conditional on industry.

To summarize, Figure 2 plots the cumulative distribution function of baseline profit for the entire sample, the population we study, and the selected matches. One can see that our study population is somewhat poorer than the entire population, while the matches are drawn from the far right tail of the baseline profit distribution.

Figure 2: Baseline Profit Distributions



[19]Note that this procedure satisfies all the requirements in Assumption 4. We do not assume we can observe control matches (part 2 of Assumption 4), but the randomization immediately satisfies the exclusion restriction (part 1). Our selection the treatment matches satisfi1s the final aspect of Assumption 4 when combined with Assumption 2.

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

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.[21]

The program was pitched to both sides of the match as a mentee-mentor relationship, and thus was explicitly focused on business success. The older, more successful business owners were the "mentors," while the younger owners were the "mentees," consistent with both their profitability and time engaged in business. The mentors were told they could potentially help other business owners learn the requisite skills required to operate in Nairobi. However, we provided no topics to discuss, instead preferring that the content of any discussions was self-directed. After signing up mentors we simply provided the mentees with the mentor's phone number and told them that a prominent business owner in Dandora was willing to discuss business questions with them. Whether they contacted the mentor, or ever met, was their decision. However, all matches met at least once in the official month-long treatment period.

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

3.2 RCT Results

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

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

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

We found that entrepreneurs receiving a mentor realized a statistically significant

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

increase in profit. Moreover, the increase is economically significant, representing 23 percent of the control group's mean profit. In the second column of Table 1, we break out these results by the profitability of the more profitable member of the match. Here we divide the set of mentors by their percentile ranking within the whole set of mentors. We find that the point estimates are ordered by the profitability of each group, so that the most profitable mentors generated the largest treatment effects for their mentees.

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

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

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

As discussed in detail in Brooks et al. (2018), we found that the mechanism for this increase in profit was reduced costs, such as lower prices for inputs. This is important because greater profits driven by greater demand may mean that the mentee group took sales away from the control group, which would bias these results upward. However, we found no significant increase in revenue, which rules out that concern. Instead, mentees learned how to achieve the same scale at lower cost, consistent with higher productivity in the sense of the model. Moreover, we are able to rule out a number of other alternatives that would be inconsistent with knowledge diffusion, such

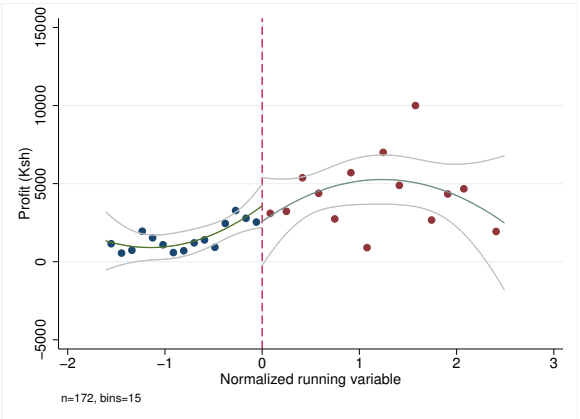
as mentors giving loans to mentees.

Impact on Higher Profit Business Owner Finally, we consider the impact on more productive members of the match. The diffusion process in Section 2 assumes that there should be no gains to these individuals, which is implicitly assumed through the use of the max function in law of motion for productivity (Assumption 1). These individuals were not randomly selected relevant to their peers, and thus cannot be directly compared to a control group as the others were. However, our design allows us to use the selection procedure to identify the causal impact of being chosen. Specifically, we surveyed both those chosen for the program and those just below the cutoff for selection, then employed a regression discontinuity design to study the impact of being chosen into the program.

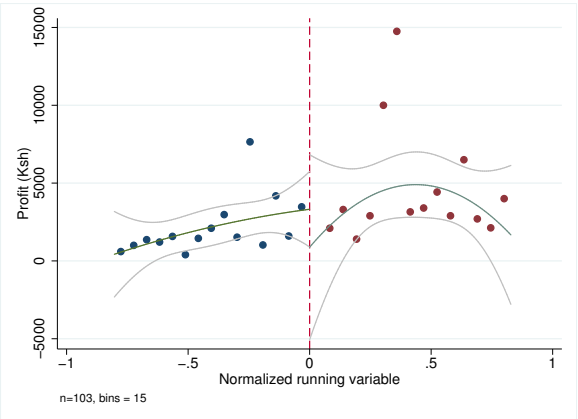
Figure 3 plots profit along with a fitted quadratic and its 95 percent confidence interval. Figure 3a uses the entire sample, while Figure 3b uses the Imbens and Kalyanaraman (2012) procedure to choose the optimal bandwidth. Both use 15 bins on either side of the cutoff. Figure 3 suggests no statistically discernible discontinuity around the cutoff.

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

(a) Full sample



(b) IK Optimal Bandwidth



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

mentors, we run the regression

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

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 (equation 2.1), which is assumed here and in much of the existing literature.

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

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

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

We emphasize that while this is consistent with the model described in Section 2, where higher productivity firms receive no benefit from interaction with lower productivity firms, there is nothing in the experimental design that guarantees this outcome.[22]

[22]For example, the relationship may be consistent with a "collaboration" model in which both sides gain from interacting with the other. On the other hand, if the time requirement is high enough, there could be negative effects

3.3 Diffusion Parameter Estimates

We use these results to estimate the parameters of the diffusion process. Applying Propositions 1 implies that $\beta = 0.31$ and $\rho = 0.43$. Estimating θ requires taking a stand on the functional form of $\widehat{M}$. In our quantitative results below, we assume

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

where M is the distribution of productivity in the economy and $\underline{z}$ is the the lowest productivity of operating firms. If $\theta = 0$, firms are draw directly the existing distribution of firms, while $\theta > 0$ pushes more mass into the right tail. Estimating this function using Proposition 2 implies $\theta = 0.87$.

4 Equilibrium Impact of Diffusion

With the estimated diffusion parameters in hand, we now fill out the model and equilibrium details to study the quantitative importance of diffusion. As we have emphasized throughout, this is only one potential model in which one could deploy these results. Naturally, the quantitative magnitudes change depending on the exact model specification, but the estimated diffusion parameters do not.

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

Occupational Choice and Recursive Formulation 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.

As before, an entrepreneur's profit is

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

on mentors. However, the high take-up and persistence of matches suggests this second explanation is unlikely *a priori*.

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

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

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

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

which satisfies Assumption 2 on the proportionality of profit and productivity. This further implies that agents face a cutoff rule to determine their occupation. For a given wage w, there is a $\underline{z}(w)$ such that any agent with $z < \underline{z}$ becomes a worker, while agents with $z \geq \underline{z}$ become entrepreneurs.[23]

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

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

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

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

This gives a particular functional form of the distribution of draws. The parameter θ controls the part of the distribution from which draws arise, satisfying our Assumption 3. In particular, when $\theta = 0$, imitation draws are uniform from the set of operating firms. However, we also allow for the possibility that draws may be better or worse than that. As θ grows toward 1, the mass in $\widehat{M}$ goes deeper into the right tail of operating firms, so that imitation draws are much better than random.[24] Alternatively,

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

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

as θ goes to negative infinity, the probability mass converges to $\underline{z}$, so that no operating firm ever receives a useful imitation draw. This case captures the possibility that diffusion is completely absent.

Unlike our assumed $\widehat{M}$ in Section 2, however, $\widehat{M}$ now depends on the economy-wide productivity distribution M, an equilibrium object. It must therefore be consistent with the the diffusion process in the economy. The law of motion for M is

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

where G is the exogenous distribution from which new entrants draw productivity.[25]

4.1 Competitive Equilibrium

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

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

4.2 Calibration of Remaining Parameters

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 parametrization can be broken into three different parts that can be considered separately. First, as we showed previously, the diffusion parameters are independent of the remaining model parameters. Thus, we can simply impose our

[25] 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.

estimated parameters $\beta = 0.31$, $\rho = 0.43$, and $\theta = 0.87$.

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 $\alpha = 0.583$.

Our remaining parameters are σ, σ_0, c and τ. We jointly match these four parameters to the following four moments: the standard deviation of log-profit in the overall population of operating firms (1.02), the variance of log-profit among new entrants (1.04), and the ratio of the average profit of firms overall to the average profit of new entrants (1.52). These moments are matched when $\sigma = 2.45$, $\sigma_0 = 2.12$, and $c = 0.37$.

Table 3: Targets and Parameter Choices

Model Parameter	Description	Parameter Value	Target Moment	Source	Target Value	Model Value
Group 1						
θ	Directedness of search	0.87	Overall treatment effect (as %)	RCT results	19	19
β	Contribution of matched firm productivity to own	0.31	Treatment effect from mentor in top 25% (as %)	RCT results	52	52
ρ	Persistence of productivity	0.43	Normalized profit covariance among treatment group	RCT results	1.18	1.18
Group 2						
μ	Mean of exogenous productivity shock distribution	n.a.	$\mathbb{E}[e^{\varepsilon}] = 1$	n.a.	1	1
σ	St. dev. of exogenous productivity shock distribution	2.45	Variance of log profit in all firms	Baseline survey	1.02	0.42
σ_0	St. dev. of new entrant productivity distribution	2.12	Variance of log profit among new entrants	Baseline survey	1.04	0.961
c	Growth factor in productivity evolution	0.37	Ratio of average profit of all firms to new entrants	Baseline survey	1.41	1.560
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.

4.3 Equilibrium Outcomes

Figure 4 shows per capital income in the stationary equilibrium in the (β, ρ) space. Relative to a model with no diffusion, the implied equilibrium incomes can be substantially larger. At our estimated values $(\beta, \rho) = (0.31, 0.43)$, average income is 70 percent higher than a world in which diffusion is shut down (i.e., $\beta = 0$ holding $\rho = 0.43$). In principal, however, those differences can be even larger, as one can see from the complementarity in between high values of β and high values of ρ in Figure 4. Fixing $\rho = 0.8$, the difference in equilibrium income between $\beta = 0$ and $\beta = 0.8$ explodes to a factor of 2832. Thus, while the model indeed allows for diffusion to be an important factor in driving income, our estimates suggest a relatively more modest view. We will return to this when we discuss the magnitudes of counterfactual policy in the next section.

Figure 4: Equilibrium Income per Person

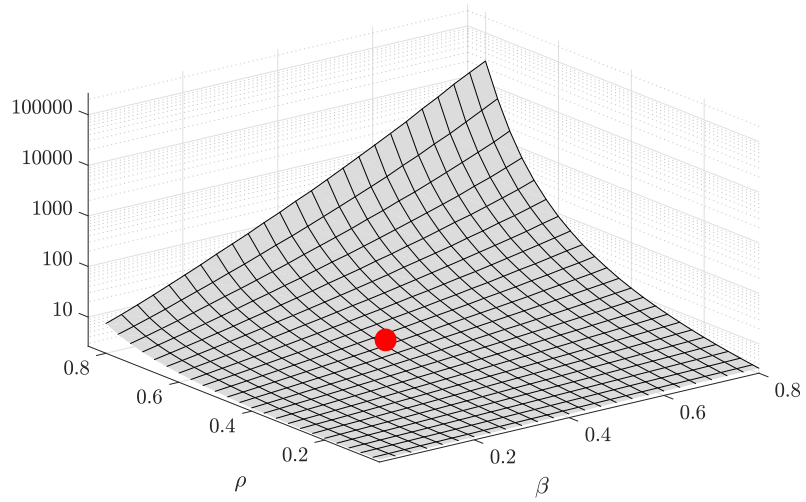


Figure notes: This figure plots average income for a variety of (β, ρ) combinations holding the remaining model parameters fixed. Our estimated values are indicated by the red dot. The income scale is measured in logs.

4.4 Quantitative Results

We consider two exercises in the model to study the impact of diffusion in our parameterized model. In Section 4.5, we consider gains from optimally subsidizing individuals to work for wages. In Section 4.6 we introduce a new set of agents who have higher

productivity than the existing firms in the stationary equilibrium. This experiment allows us to study a scenario in which existing firms may be negatively affected, as the new productive firms increase the equilibrium wage. The goal here is to see if diffusion can offset the potentially negative impact on existing firms.[26] We again find an important role for diffusion. In our model, existing firms increase profit by 0.1 percent. In the same model without diffusion, profit declines by 4 percent. Thus, the equilibrium impact of diffusion more than makes up for the negative equilibrium impact of higher wages.

4.5 Optimal Wage Subsidy

The equilibrium of this model is inefficient due to the negative externality generated by marginal firms. Because knowledge diffuses via random search (conditional on directedness parameter θ), marginal firms decrease the likelihood of an individual learning from the right tail of the knowledge distribution. The planner, therefore, wishes to subsidize wage labor, thus inducing marginal firms to exit and increasing the average match quality of an individual firm. To study this result, we solve for the optimal wage subsidy. That is, instead of choosing between working for profit $\pi(z)$ and working at wage w, they now have income equal to $\max\{\pi(z), (1+\tau)w\}$. The subsidy is financed via lump sump tax on all agents. We solve for the τ^* that maximizes average income in the economy.

At our estimated parameters, we find that the optimal subsidy increases income by 9.1 percent. However, this gain requires a heavy-handed policy as the optimal subsidy is 12.8 times the equilibrium wage. The next two subsections attempt to understand these two pieces of the policy – the impact on income and the seemingly high subsidy level required to attain it.

4.5.1 Understanding the Income Gains from Policy

We begin by investigating the various diffusion parameters and their impact on the gains from an optimal wage subsidy. Figure 5 begins by plotting the optimal subsidy level and relative income gain compared to the baseline economy for various levels of

[26] This arises in a variety of economic scenarios. For instance, opening to trade exposes existing firms to increased competition (Asturias et al., forthcoming). In the more micro-development literature, this has been studied recently in the context of spillovers from micro-enterprise training (McKenzie and Puerto, 2017).

ρ.

Figure 5: Gains from Optimal Wage Subsidy

(a) Optimal subsidy level, $1 + \tau^*$ (b) Relative income gain versus $\tau = 0$

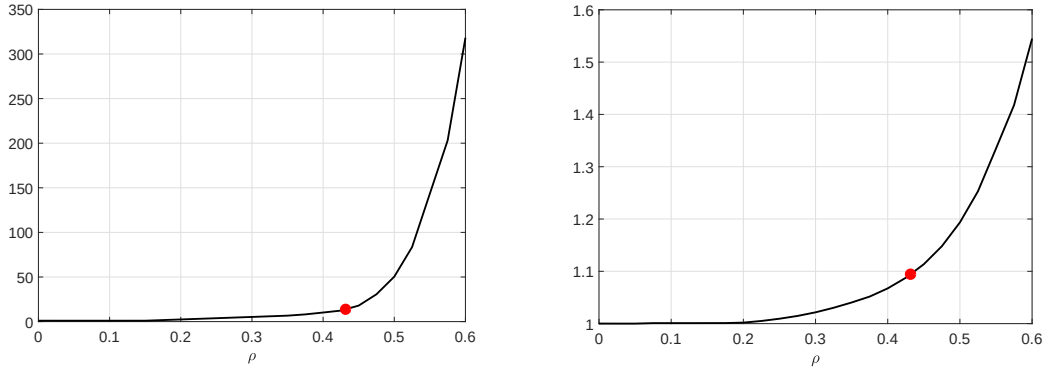


Figure notes: This figure plots the level and impact of the optimal wage subsidy for various levels of ρ, holding the remaining parametrization fixed. Our estimated results are indicated by the red dot.

There are two results one can see from Figure 5. First, in principal, high persistence ρ can substantially increase the impact of policy. If $\rho = 0.6$, the optimal subsidy increases income by 54 percent compared to the baseline economy with $\tau = 0$. Again, however, this implies an extremely large subsidy rate of 318 times the equilibrium wage. This implies an interesting potential role for policy – to the extent that income persistence varies across economies (countries, sectors, etc.), there is potentially much larger scope for policy gains in other economies than we find here.[27]

Figure 6 turns to the importance of θ (directedness) and β (intensity) in generating these income gains. Figure 6a shows that lower θ increases the gains from policy. This is because higher θ implies that individuals are more easily able to access high-z matches, limiting the impact of this policy designed to make marginal firms exit. Holding the rest of our parametrization fixed, the optimal subsidy increases income by 9.9 percent at $\theta = 0$ compared to 9.1 percent at our estimated $\theta = 0.8705$.

Figure 6b shows that differences in θ are dwarfed by differences in β. At our estimates parameters for θ and ρ, the optimal subsidy increases income by 0.7 percent (when $\beta = 0.10$), 9.1 percent ($\beta = 0.31$, our estimated value), and 23.8 percent ($\beta = 0.50$). As one might suspect, increasing β makes it easier for skill to diffuse,

[27] We note that our experiment did not empirically change the persistence of profit across treatment and control groups. Thus, one could potentially use non-RCT panel data to provide an indicator of where diffusion policy was likely to generate the largest gains.

thus providing a larger role for policy. However, as was the case when considering average income in the stationary equilibrium, the complementarity between β and ρ plays a critical role. At $\rho = 0.6$, the income gains are 3.0 percent ($\beta = 0.10$), 50.5 percent ($\beta = 0.31$), and 278 percent ($\beta = 0.5$). Thus, the impact of higher persistence increases at high levels of β relative to low levels of β.

Figure 6: Gains from Optimal Wage Subsidy

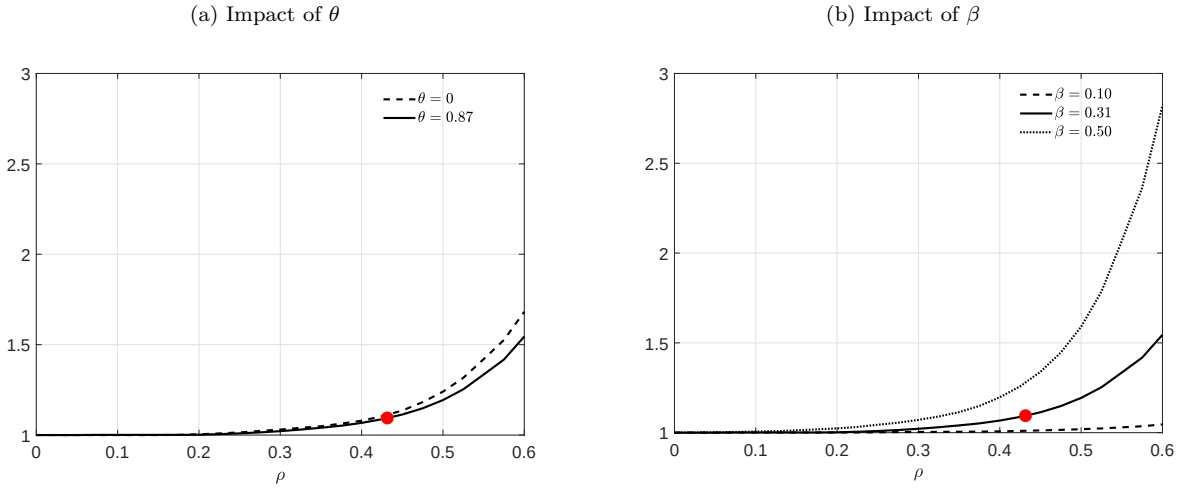


Figure notes: This figure plots the average income change in response to the optimal wage subsidy subsidy for various diffusion parameter combinations. Our estimated gains are indicated by the red dot on the solid line.

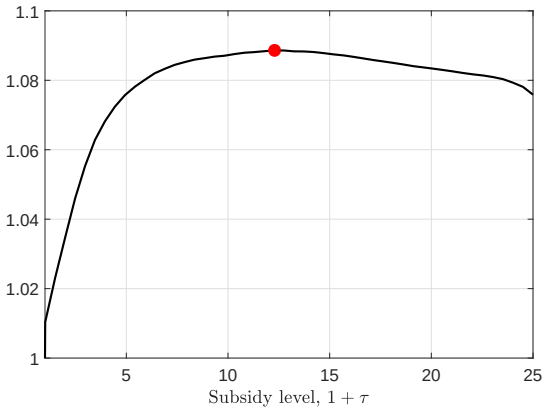
The intuition for this complementarity between β and ρ is straightforward. The ability to internalize productivity (high β) is worth relatively little if those gains die out quickly (low ρ). Thus, with low persistence, there is little to be gained from intervention regardless of β. Once the impact of a good match lasts for multiple periods, the planner has a larger role to play in managing the externalities created by diffusion. Taken together, the results show that the magnitude of the income gains we estimate are a function of the relative low levels of β and ρ in the $[0, 1] \times [0, 1]$ space combined with a strong complementarity between β and ρ implied by the model. The second issue is high subsidy required to generate the income gains we observe in these counterfacutals. We turn to this next.

4.5.2 Why are the subsidy rates so high?

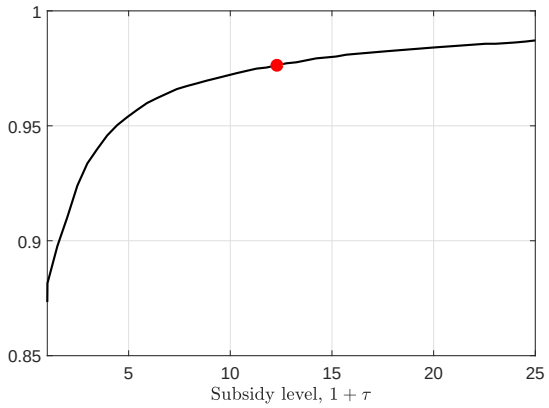
Figure 7 plots equilibrium outcomes as the subsidy level changes. The key role for the planner here is to increase expected match quality. This requires moving marginal firm owners to working.

Figure 7: Subsidy Variation on Equilibrium Outcomes

(a) Average Income



(b) Share of Population as Workers



(c) Average Firm Productivity

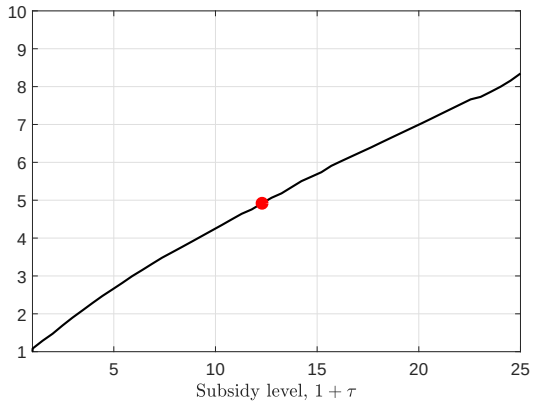


Figure notes: This figure plots variation in outcomes as the subsidy level changes in the baseline economy. Panels (a) and (c) are normalized by the no-subsidy baseline equilibrium. Optimal subsidy outcomes are marked by the red dot.

4.6 The Impact of New Productive Firms

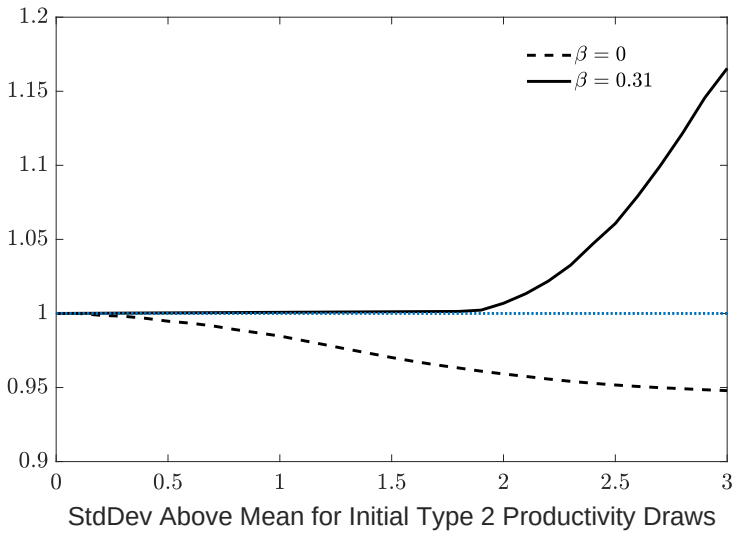
A second test of the importance of diffusion is the extent to which it can compensate incumbent firms from competitive pressured induced by the introduction of new, more

productive firms. This arises in variety of contexts, including theories of creative destruction (Aghion et al., 2014) and trade (Asturias et al., forthcoming; Buera and Oberfield, 2017). In perhaps a more useful context for our estimated economy, recent work on microenterprise training has studied the extent to which benefit to trained firm owners comes at the expense of untrained owners (McKenzie and Puerto, 2017).

To study this, 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) and a measure 0.1δ of new agents (type 2) draw from a distribution that has a higher average mean. 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. Throughout, we hold the rest of the parametrization fixed.

Figure 8 plots the average income changes for incumbent firms. The horizontal axis varies how many standard deviations above the type 1 mean productivity realization the type 2 draws are, while the y-axis is the average change in income relative to the baseline equilibrium with no new entrants.

Figure 8: Average Income Gain for Existing Firms after New Entrants



When we eliminate the diffusion channel by setting $\beta = 0$, the average incumbent firm is unambiguously hurt by these more productive new entrants. Because type 2 agents have higher average productivity, they demand more labor and cause the equilibrium wage to increase. However, once we return to our estimated $\beta = 0.31$,

the average incumbent benefits in all cases. Diffusion introduces a second effect of improving the $\widehat{M}$ distribution from which these firms draw. Our estimated model implies that this second effect more than offsets the first. For some perspective, a 2 standard deviation shift in the initial mean productivity draw for type 2 agents induces 4.1 percent decline in average incumbent income without diffusion, compared to a 0.1 percent increase in average income once we allow for diffusion. Figure 8 shows that the gap between the two cases widens as the new entrants are productive, as the diffusion benefits start to provide larger income gains to the incumbents.

Again, our estimated model predicts an important role for diffusion. Diffusion forces are sufficiently strong here to reverse the negative impact on incumbent firms. It is worth emphasizing that this is not a generic result in the model, but a function of the estimated parameter values. For sufficiently low combinations of β and ρ, the average incumbent firm can indeed be affected negatively by the entry of productive firms.

5 Discussion

5.1 Alternative Modeling Choices

As we highlighted in Section 2, the model laid out here is not the only one in which one could deploy our estimated diffusion parameters, nor the only diffusion process. We highlight the random search framework in the main body of the paper (net of the directedness parameter) because it allows us to link a specific framework to the broader identification results in Section 2 in a relatively transparent way. Moreover, it underlies a large portion of the literature.

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

In Appendix B we hold the diffusion process fixed, but vary the surrounding model structure. We consider two possibilities. First, we we allow for fixed, *ex ante* heterogeneity in the utility benefit from operating a firm. This accounts for the fact that individuals potentially have different incentives for starting a business that may be unrelated to productivity, such as flexible hours. This changes the planner's problem,

as not all individuals of the same productivity exit in response to a subsidy.

5.2 External Validity

External validity is a concern in any randomized controlled trial, and our results are certainty subject to the same criticism. While there is little we can do to directly address this, we note that our results suggest a number of important dimensions of heterogeneity one could test across sectors or countries.[28]

For example, the estimated persistence of productivity, ρ, is low relative to other estimates in more developed countries(see Brooks et al., 2018, for a longer discussion of this result). Moreover, the quantitative results show that this persistence parameter is an important margin for the quantitative importance of diffusion. To study how our results would change when applied to other economies with more persistent productivity, we provide a number of additional results in Appendix C. We begin by showing that the treatment did not change the persistence of profit between treatment and control firms. Thus, one could estimate persistence independent of available RCT data. Next, we ask what combinations of (β, θ) would rationalize the results for a fixed level of ρ and show that there are larger gains from policy at higher levels of ρ. Thus, at least from the perspective of our estimated parameters, our results suggest a lower bound.[29]

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

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

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

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

important questions in the future.

6 Conclusion

This paper uses evidence from a randomized controlled trial to identify a model of firm-to-firm transmission of productivity. Our results imply an important role for diffusion. The efficient level of the learning externality increases income by 9 percent, and the strength of diffusion is sufficient to overcome the negative pressure on incumbent firms due to the entry of more productive firms. We emphasize that this need not be true, as both the model and experimental design *ex ante* allow for the possibility that the impact of diffusion is small or nonexistent.

We view these results as an important first step that highlights the possibilities of linking aggregate diffusion models with causal identification. The next steps require a more detailed investigation of model and data. For example, one question that remains unanswered both in this paper and the broader literature is why individuals do not seek out the most productive business owners to learn from, given the seemingly large benefits and low costs we observe (Brooks et al., 2018). Beaman and Dillion (2017) point to frictions in the information market, while Fogli and Veldkamp (2016) point out that aggregate growth-reducing network structures can be an optimal response to the possibility of detrimental flows through the network (e.g., disease). Put differently, here we assume a diffusion process assumed by much of the recent literature (though generalized somewhat), and estimate it. The next steps require distinguishing across diffusion processes, including any potential distortions that arise. Different field experiments, designed with an eye toward aggregate theory, could provide more detailed information on the law of motion for productivity to help further refine such model choices.

References

Aghion, Philippe, Ufuk Akcigit, and Peter Howitt, "What Do We Learn from Schumpeterian Growth Theory?," in Philippe Aghion and Steven N. Durlauf, eds., *Handbook of Economic Growth*, 2014, pp. 515–563.

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

Asturias, Jose, Manuel García-Santana, and Roberto Ramos, "Competition and the Welfare Gains from Transportation Infrastructure: Evidence from the Golden Quadrilateral of India," *Journal of the European Economic Associationn*, forthcoming.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

McKenzie, David and Susana Puerto, "Growing Markets through Business Training for Female Entrepreneurs: A Market-Level Randomized Experiment in Kenya," March 2017. World Bank Policy Research Working Paper 7993.

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

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

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

Table of Contents for Online Appendix

A Examples of Diffusion Processes **40**

A.1 Model of Multiple Draws 40

A.2 Model with Effort Choice 40

A.3 Innovations through Imitation 41

A.4 Model with Deterministic Assignment 42

B Other Models **43**

B.1 *Ex Ante* Heterogeneity in Utility Benefit from Entrepreneurship 43

C Additional Results **44**

C.1 Persistence Across Groups 44

C.2 Fixing ρ and Re-Estimating Diffusion Parameters 45

C.3 Allowing for Idiosyncratic Distortions 46

D Proofs **47**

D.1 Proof of Proposition 1 47

D.2 Proof of Proposition 2 48

A Examples of Diffusion Processes

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

A.1 Model of Multiple Draws

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

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

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

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

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

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

A.2 Model with Effort Choice

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

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

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

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

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

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

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

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

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

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

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

or following the notation more standard in the paper:

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

A.3 Innovations through Imitation

Here we show how the Buera and Oberfield (2017) environment maps into that considered in this paper. In their model (adapted to our notation), an agent with productivity z receives new arrivals of ideas that have two components: z_m that comes from a random match from another agent, and γ a random innovation on that idea.

Then $\hat{z} = \gamma^{1/\beta} z_m$. Here, ν is an exogenous random variable and z_m is a uniform draw from the distribution of productivities. Then if γ has a cumulative density function given by Γ, then:

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

A.4 Model with Deterministic Assignment

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

The utility of an agent with productivity $\hat{z}$ influencing an agent with productivity z is given by:

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

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

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

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

However, since every agent can only be influenced by one agent each period and they strictly prefer to be influenced by agents of higher productivity, it is possible that (even if the distribution is continuous) that the ideal agent for $\hat{z}$ is already matched to another influencer. Therefore, intuitively, the probability distribution over assignment between $\hat{z}$ and z is constructed by starting at the upper support of the distribution

M, allowing the highest productivity firms to choose their most preferred matches, then descending down through the distribution letting each firm choose to influence its preferred firm among those remaining. Note that not all firms need have another firm to influence if their utility from doing so be negative.

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

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

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

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

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

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

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

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

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

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

B Other Models

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

[to be completed]

C Additional Results

C.1 Persistence Across Groups

One possibility is that our treatment changed the persistence of income across firms. We test that here, by comparing the persistence of profit across treatment and control groups. To begin, we run the following regressions separately on the treatment and control groups

$$\begin{aligned} \pi_{i,t} &= \alpha + \rho\pi_{i,t-1} + X_i\eta + \nu_t + \xi_{it} \\ \pi_{i,12} &= \alpha + \rho\pi_{i,0} + X_i\eta + \xi_i. \end{aligned}$$

The first uses profit ($\pi_{i,t}$) data from all seven waves and asks how lagged profit predicts current profit, where ν_t is a wave fixed effect. To eliminate any risk that these results are contaminated by seasonality, the second uses only the baseline data and one year post. That is $\pi_{i,12}$ and $\pi_{i,0}$ are recorded a year apart. The results are in Figure 4, and shows little difference across control and treatment. Note that persistence does increase somewhat using only the annual data, suggesting there are quantitatively important seasonal shocks in the data.

Table 4: Persistence of Profitability

	All Periods		Annual	
	Control	Treatment	Control	Treatment
Lagged profit	0.161	0.185	0.270	0.267
	(0.040)***	(0.047)***	(0.093)***	(0.150)*
Wave F.E.	Y	Y	N	N
Controls	Y	Y	Y	Y

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

As a slightly more formal test we combine the control and treatment data and run

$$\begin{aligned} \pi_{i,t} &= \alpha + \rho\pi_{i,t-1} + \omega(\pi_{i,t-1} \times T_i) + \lambda T_i + X_i\eta + \nu_t + \xi_{it} \\ \pi_{i,12} &= \alpha + \rho\pi_{i,0} + \omega(\pi_{i,t-1} \times T_i) + \lambda T_i + X_i\eta + \xi_i. \end{aligned}$$

where $T_i = 1$ if an individual is in the treatment group. The relevant coefficient estimates $\widehat{\rho}$ and $\widehat{\omega}$ are in Table 5, and confirm the previous results.

Table 5: Persistence of Profitability

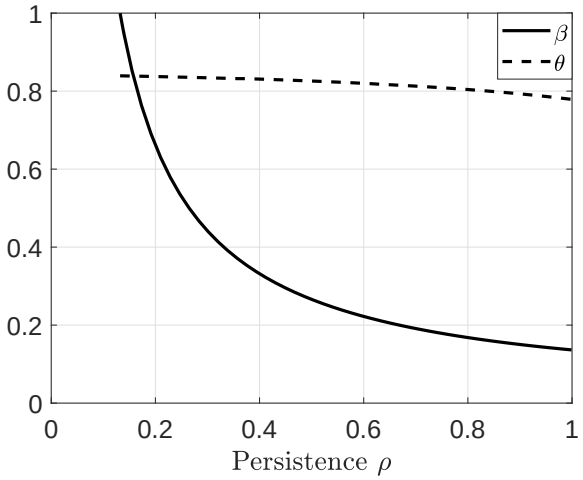
	All Periods	Annual
Lagged profit	0.191	0.269
	(0.045)***	(0.098)***
Lagged profit	0.003	-0.019
× Treated	(0.059)	(0.147)
Wave F.E.	Y	N
Controls	Y	Y

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

C.2 Fixing ρ and Re-Estimating Diffusion Parameters

In light of the previous results showing that our experiment did not induce changes in persistence of profit, we re-estimate the diffusion parameters β and θ, conditional on the value of ρ. To do so, we simply drop the second moment Γ_3 and choose (β, θ) to match (Γ_1, Γ_2). Figure 9 shows how the estimates vary with difference levels of ρ.

Figure 9: Re-Estimating β and θ for fixed ρ



First, θ varies little, as to be expected from its small quantitative importance in quantitative results of Section 4.4. Second, note that β and ρ co-move negatively.[30] Intuitively, low ρ dampens the theoretical impact of getting a good match. Thus, the estimation responds by requiring a larger intensity parameter β to match $\Gamma_1^{empirical}$. Generating both a high β and high ρ via our estimation procedure would have required

[30] At $\rho = 0.13$, $\beta = 1$. Thus, the parameter values are undefined at $\rho < 0.13$.

a massive impact of a more productive match relative to a less productive match within the treatment group.

C.3 Allowing for Idiosyncratic Distortions

[to be completed]

D Proofs

D.1 Proof of Proposition 1

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

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

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

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

Define:

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

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

Then define:

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

Last, define:

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

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.

D.2 Proof of Proposition 2

Proof. The right hand side of equation (2.3) is continuous in θ by the continuity of $\widehat{M}$ in Assumption 3. Given that $\Gamma_2 \in (\Gamma_2^{min}, \Gamma_2^{max})$, existence follows by the intermediate value theorem. Uniqueness follows from the strict monotonicity of the right hand side of equation (2.3) in θ. This follows from Assumption 3 in that $\theta_1 < \theta_2 \implies \widehat{M}(\cdot, \cdot, \theta_2)$ first order stochastic dominates $\widehat{M}(\cdot, \cdot, \theta_1)$. ■