The Missing Middle in Developing Countries Revisited

Kunal Dasgupta

University of Toronto

December 2010

Abstract

Evidence points to the existence of a "missing middle" in the size distribution of firms in developing countries. Furthermore, the "missing middle" seems to disappear as a country develops. In this paper, we develop a dynamic model to understand the evolution of the firm size distribution in developing countries. Unlike existing explanations, our model does not rely on frictions to generate the bi-modality in size distribution. Rather, the bi-modality arises due to agents optimally selecting into a traditional and a modern sector. The key parameter in our model is the mean knowledge of the newborns. For a low mean, the two sectors co-exist. As the mean rises, the size distribution converges from a bi-modal to an uni-modal distribution. We present some evidence to support our hypothesis.

1 Introduction

To be sure, for some people the best route out of poverty will be a bank loan. But for most, its going to be something much simpler: a regular paycheck.

James Surowiecki (2008)

A very large number of individuals in developing countries are engaged in the informal sector. According to the International Labour Office (ILO), “The streets of cities, towns, and villages in most developing countries .... are lined by barbers, cobblers, garbage collectors, waste recyclers, and vendors of vegetables, fruit, meat, fish, snackfoods, and a myriad of nonperishable items ranging from locks and keys to soaps and detergents, to clothing” (ILO, 2002, page 9). One half to three quarters of non-agricultural employment in developing countries is in the informal sector. The fraction of agents in the informal sector ranges from 51 percent in Latin America to almost 78 percent in Sub-Saharan Africa (ILO 2002).

The prevalence of the informal sector in developing countries has implications for the plant size distribution in these countries. For the U.S. in 1992, the share of employment rose gradually with plant size, a pattern that is true in other developed countries as well (see Cabral and Mata, 2003).[1] But it is in contrast to the experience of many developing countries, whose size distributions exhibit a “missing middle”. This can be seen most clearly in Table (1), which has been re-produced from Tybout (2000). According to the table, countries like Zambia or Thailand or Colombia have a very large proportion of the workforce employed in small establishments, but very few workers in medium-sized plants. This bi-modality, however, is not a feature of the stationary size distribution. The table suggests that the size distribution is evolving over time. In Taiwan and Indonesia for example, the “missing middle” seems to be disappearing slowly. What accounts for the presence of the “missing middle” in some economies and not in others? What explains the gradual disappearance of the “missing middle” over time? These are some of the questions that we shall try to answer in this paper.

We develop a model in which heterogeneous agents choose their occupations. The economy has two sectors - a traditional sector and a modern sector. In both sectors, production production is carried out in plants. Each plant is run by workers under the supervision of a manager/entrepreneur. The manager chooses the number of workers so as to maximize profits. There are two key differences between the two sectors. First, the modern sector has a larger “span of

[1]If the firm size distribution follows Zipf’s Law, the employment share of firms is monotonic increasing in the size of firms.

control". This implies that the profit-maximizing size of a plant would be greater for a manager, if he were in the modern sector. And second, a defining feature of the modern sector is the presence of knowledge spillovers. As a result, both workers and managers in the modern sector acquire more and more knowledge over time.

	Numbers of Workers										
	1–4		5–9		10–19		20–49		50–99		>99
United States, 1992[a]	1.3		2.6		4.6		10.4		11.6		69.4
Mexico, 1993[b]	13.8		4.5		5.0		8.6		9.0		59.1
Indonesia, 1986[c]	44.2			17.3					38.5		
S. Korea, 1973[d]	7.9				22.0					70.1	
S. Korea, 1988[e]			12					27			61
Taiwan, 1971[c]				29.1						70.8	
Taiwan, 1986[f]			20					29			51
India, 1971[g]		42				20				38	
Tanzania, 1967[g]		56				7				37	
Ghana, 1970[g]		84				1				15	
Kenya, 1969[g]		49				10				41	
Sierra Leone, 1974[g]		90				5				5	
Indonesia, 1977[g]		77				7				16	
Zambia, 1985[g]		83				1				16	
Honduras, 1979[g]		68				8				24	
Thailand, 1978[g]		58				11				31	
Philippines, 1974[g]		66				5				29	
Nigeria, 1972[g]		59				26				15	
Jamaica, 1978[g]		35				16				49	
Colombia, 1973[g]		52				13				35	
Korea, 1975[g]		40				7				53	

Table 1: *The distribution of employment shares across plant sizes (Source: Tybout, 2000)*

We show that in equilibrium, for certain values of the parameters, the least knowledgeable agents select into the traditional sector, while the most knowledgeable become managers in the modern sector. The ones with intermediate levels of knowledge work for wages in the modern sector. This generates a discontinuity in the plant size distribution. The bi-modality arises primarily because some agents, instead of running their own plants, choose to work for others.

We interpret development as an improvement in the educational attainment of the population. In the context of our model, this implies a fall in the cumulative density function of the underlying knowledge distribution. As the cumulative densities decline, two opposing forces come into play. On the one hand, there is an abundance of workers in the modern sector, which tends to pull down the wage. On the other hand, agents learn more on average, since they are surrounded by more knowledgeable people. This raises their continuation value. We show that the latter force dominates and the agent who was indifferent between the two sectors at time t, strictly prefers to be in the modern sector in $t+1$. Hence, over time, agents move from the traditional to the modern sector; the lower mode in the firm size distribution shrinks and eventually disappears.

Alternative explanations have been provided for the emergence of the bi-modal distribution. Rauch (1991) develops a model similar to ours in that agents are heterogeneous and plants use a Lucas span-of-control technology. In his model, prospective entrepreneurs can hire workers freely, as long as they pay the government stipulated minimum wage. Otherwise, they could pay a lower wage but be constrained by how many workers they can hire (remain informal). The "missing middle" arises due to the gap between the minimum wage and the informal sector wage. In fact, developing countries are plagued by regulatory and start-up costs. For example, according to the World Development Indicators, in 2002, it took an average of 149 days to start a business in the Democratic Republic of Congo compared to 6 days in the U.S.[2] Potential entrepreneurs enter the informal sector in order to avoid paying these costs. The size distribution in the developing countries could also be affected by high transportation costs, low levels of urbanization or even, Engel's law (See Tybout, 2000, for a discussion of these and related issues). Institutional factors like minimum wages, however, can not be the sole reason for this duality, as argued by Mazumdar (1983). According to Mazumdar (1983), there are many developing countries which exhibit duality, despite limited power of trade unions or government legislations.

One could argue that the size duality is a result of the choice between entrepreneurship and wage work, where the choices are dictated by other non-pecuniary considerations. Entrepreneurship, especially self-employment, could give greater autonomy to the agent - a chance to be "your own boss" (Hamilton, 2000). Consequently, individuals who prefer autonomy would select into becoming entrepreneurs. Non-pecuniary theories, however, can not explain cross-country differences in firm-size distribution or the evolution of size distributions over time since it would require systematic difference in preference across space and time.

The rest of the paper is organized as follows. The model is presented in Section 2. In Section 3, we define and characterize the equilibrium. The implications for the firm size distribution are explored in Section 4. Section 5 concludes.

2 The Model

A single, homogeneous good is produced in the economy. The good can be produced in either of two sectors : a knowledge-intensive modern sector (M) and a traditional sector (T) which uses a labor-intensive technology.[3] Agents are born every period and draw their knowledge k

[2]http://www.doingbusiness.org/ExploreTopics/StartingBusiness/

[3]According to Jorgenson (1967), "The special character of the theory of development of a dual economy is an asymmetry in the productive relations". Jorgenson, as well as, Lewis (1954) and Ranis and Fei (1961) assume that land and labor are used in the traditional sector while labor and capital are the factors of production in the modern

from the newborn distribution $\mathbb{N}(k)$ which is assumed to be Pareto with scale parameter $\underline{k}$ and shape parameter γ. Agents also die every period with an exogenous probability δ. Let us denote the endogenous knowledge distribution at time t by Φ_t, with ϕ_t being the corresponding density. Agents are assumed to be risk-neutral.

2.1 Traditional Sector

In the traditional sector, production is carried out within firms. Each firm consists of a manager and workers. Output depends on the knowledge of the manager, whereas the workers perform routine tasks. This is the familiar Lucas span-of-control type production function. At time t, the manager pays a wage w_t^T to all his workers (since workers perform routine tasks, their earnings are the same, irrespective of who they work for) and is the residual claimant of the output. A manager with knowledge k hires workers to maximize his profit given by

$$\pi_t^T(k) = \max_n \{kn^\alpha - w_t^T n\} \tag{1}$$

where $\alpha < 1$ is the span of control. A value of α less than one implies that there are diminishing returns in the number of workers. Smaller is α, fewer is the number of workers that a manager would hire. In the limiting case, as $\alpha \to 0$, we get self-employment. The value of α is determined by technology, and as we shall see, varies across sectors. There is no learning in the traditional sector. Therefore, in the steady-state, an agent choosing to be in the traditional sector has the same level of knowledge throughout his life. Solving equation (1) gives us $\pi_t^T(k) = A(w_t^T)^{\frac{-\alpha}{1-\alpha}} k^{\frac{1}{1-\alpha}}$, where $A = \alpha^{\frac{1}{1-\alpha}}(\alpha^{-1} - 1)$.

2.2 Modern Sector

The modern sector has a production technology that is similar to the traditional sector, the only difference being the span of control. In particular, managers in the modern sector have a bigger span of control. [4] At time t, the modern sector managers pay a wage of w_t^M to the workers. A manager with knowledge k hires workers to maximize his profit given by

$$\pi_t^M(k) = \max_n \{kn^\beta - w_t^M n\} \tag{2}$$

sector. Here, although labor is the only factor, the asymmetry arises through the way production is organized in the two sectors.

[4] This could be because the cost of supervising a larger workforce is lower in the modern sector.

where $\alpha < \beta < 1$. For technical purposes, we impose the following condition on the parameters :

Assumption 1 : $\delta\gamma - \frac{1}{1-\beta} > 0$

The above condition ensures that the equilibrium is well-defined. Every agent in the modern sector learns while working.[5] This is one of the defining features of the modern sector. An agent with knowledge k_t draws k_{t+1} from a distribution that depends on k_t. Let us denote this distribution by $\mathcal{L}(k_{t+1}; k_t)$. The dependence of $\mathcal{L}$ on k_t captures the assumption that agents who are more knowledgeable today are likely to be more knowledgeable tomorrow. An agent does not pay a price in order to learn; he acquires knowledge from those who surround him. Learning is a pure externality. [6] We also assume that the support of $\mathcal{L}(k_{t+1}; k_t)$ is $[k_t, \infty)$ - knowledge does not depreciate.

Finally, as in the traditional sector, the current profit of a modern sector manager is given by $\pi_t^M(k) = B(w_t^M)^{\frac{-\beta}{1-\beta}} k^{\frac{1}{1-\beta}}$, where $B = \beta^{\frac{1}{1-\beta}}(\beta^{-1} - 1)$.

2.3 Labor market clearing

Each sector has its own labor market. Both the labor markets are potentially dynamic, because of learning. Therefore, an agent who is a worker in the modern sector in period t, could become a manager in the traditional sector in period $t+1$. Let us denote the set of workers in the traditional and the modern sector by S_{WO}^T and S_{WO}^M respectively. Similarly, let us denote the set of managers in the traditional and the modern sector by S_{MA}^T and S_{MA}^M respectively. For any k, we must have $k \in \{S_{WO}^T, S_{WO}^M, S_{MA}^T, S_{MA}^M\}$. The labor market clearing condition in the economy is then given by

$$\int\limits_{S_{WO}^T} \phi(s)ds = \int\limits_{S_{MA}^T} n^T(s)\phi(s)ds \tag{3}$$

and

[5] This assumption is different from Dasgupta (2008) where only workers learned from their managers. However, in Dasgupta (2008), learning is match-specific; the same agent learns differently if working for different managers. The assumption that how much an agent learns is independent of who he is matched with makes the problem analytically tractable.

[6] For long, economists have realized the importance of learning on-the-job. It is well-known that an agent learns from his co-workers. We assume that an agent also learns from workers and managers in other organizations. One can rationalize this by assuming that within the modern sector, agents meet other agents at random. Upon meeting, knowledge is transferred from the more knowledgeable to the less knowledgeable.

$$\int_{S_{WO}^{M}} \phi(s)ds = \int_{S_{MA}^{M}} n^{M}(s)\phi(s)ds \tag{4}$$

where n^T and n^M solve equations (1) and (2) respectively. The left-hand sides of the above two equations represent labor supply, while the right-hand sides represent labor demand. Note that if equations (3) and (4) are satisfied, then the economy-wide demand for workers equals the supply of workers.

2.4 Evolution of the knowledge distribution

Apart from birth and death, agents move within the distribution due to learning. For a given k, let $E_t(k)$ denote the measure of agents who have knowledge less than k at the beginning of period t but more than k at the end of that period. Given the learning distribution $\mathbb{N}(k)$, the endogenous knowledge distribution Φ_t evolves according to the following law of motion :

$$\Phi_{t+1}(k) = (1-\delta)(\Phi_t(k) - E_t(k)) + \delta\mathbb{N}(k) \tag{5}$$

From the definition of $\Phi_t(k)$ and $E_t(k)$, it follows that $\Phi_t(k) - E_t(k)$ in the above equation denotes the measure of agents who, despite learning, have knowledge less than k at the end of period t. Since a fraction δ of these agents die at the beginning of period $t+1$, the first term on the right-hand side of equation (5) denotes the measure of agents with knowledge less than k who survive from period t. In the absence of population growth, a measure δ of agents who die every period is replaced by δ of newborns. Out of these, a fraction $\mathbb{N}(k)$ have knowledge less than k. Hence, the second term on the right-hand side of equation (5) denotes the measure of new agents with knowledge less than k at the beginning of period $t+1$.

3 Steady-state Equilibrium

Let $\{w_t^T, w_t^M\}_{t=o}^{\infty}$ be the sequence of equilibrium market-clearing wages. In the absence of aggregate uncertainty, the entire sequence is known by every agent in the economy. If an agent chooses to be a worker, he earns a wage while future earnings depends on his draw from the learning distribution. In particular, denoting the value function of a worker by $V_{WO}^i(k)$ where $i = T, M$, we have

$$V_{WO}^i(k_t) = w_t^i + (1-\delta)\int_{k_t}^{\infty} \max[V_{WO}^T(k_{t+1}), V_{MA}^T(k_{t+1}), V_{WO}^M(k_{t+1}), V_{MA}^M(k_{t+1})]d\mathcal{L}(k_{t+1};k_t)$$

where $V_{MA}^i(k), i = T, M$ is the value function of a manager. The term within the integral captures the fact that the agent optimizes in the next period based on how much knowledge he acquires. Similarly, if the agent chooses to be a manager, he solves

$$V_{MA}^i(k_t) = \pi_t^i(k_t) + (1-\delta)\int_{k_t}^{\infty} \max[V_{WO}^T(k_{t+1}), V_{MA}^T(k_{t+1}), V_{WO}^M(k_{t+1}), V_{MA}^M(k_{t+1})]d\mathcal{L}(k_{t+1};k_t)$$

The problem of the agent at time t is to choose the occupation that maximizes his welfare, i.e., $\max[V_W^T(k_t), V_M^T(k_t), V_W^M(k_t), V_M^M(k_t)]$. Henceforth, we shall focus on the steady-state of the system. Therefore, we shall drop all time subscripts.

Definition 1. *A steady-state equilibrium of the model consists of (i) an occupational structure (workers and managers in the traditional sector, workers and managers in the modern sector), (ii) wages and profits in the modern sector such that (a) workers maximize utility, (b) managers maximize profits, (c) labor markets clear in both sectors and (d) the knowledge distribution evolves according to equation (5).*

We claim that the equilibrium is characterized by three thresholds k^T, k^*, k^M, such that agents with $k \in [\underline{k}, k^M]$ choose to work in the traditional sector, those with $k \in [k^M, k^*]$ choose to be managers in the traditional sector, those with $k \in [k^*, k^T]$ choose to be workers in the modern sector and finally the ones with $k \in [k^T, \infty)$ are managers in the modern sector. Given this equilibrium structure, the expressions for the present value of earnings under different occupation are greatly simplified. In particular, the value function for traditional sector workers becomes

$$V_{WO}^T(k) = \frac{w^T}{\delta} \tag{6}$$

while the value function for traditional sector managers becomes

$$V_{MA}^T(k) = \frac{\pi^T(k)}{\delta} \tag{7}$$

Notice that in the absence of learning, agents who start their lives in the traditional sector remain in that sector forever. These agents essentially solve a static problem, unlike the ones who start their lives in the modern sector. Under the proposed equilibrium, the value function for agents in the modern sector becomes

$$V_{WO}^{M}(k) = w^{T} + (1-\delta)\int_{k}^{k^{M}} V_{WO}^{M}(k')d\mathcal{L}(k') + (1-\delta)\int_{k^{M}}^{\infty} V_{MA}^{M}(k')d\mathcal{L}(k') \tag{8}$$

and

$$V_{MA}^{M}(k) = \pi^{M}(k) + (1-\delta)\int_{k}^{\infty} V_{MA}^{M}(k')d\mathcal{L}(k') \tag{9}$$

In the steady-state, every agent who starts his life as a worker in the modern sector, becomes a manager with a positive probability. Those who start their lives as managers, become more knowledgeable managers over time. For the agents in the manufacturing sector, the term $E_t(k)$ in equation (5) could be a very complex expression. Even if $\mathbb{N}(k)$ and $\mathcal{L}(k_{t+1})$ belong to well-known family of distributions, it is generally not possible to get a closed form solution for the stationary distribution $\Phi(k)$. There is, however, an exception as shown by the following lemma.

Lemma 1. *Suppose the learning distribution of an agent with knowledge k is Pareto with scale k and shape γ. Then the stationary distribution with support $[k^*, \infty)$ is also Pareto.*

Proof. We shall prove the lemma using the guess and verify approach. Suppose the stationary distribution is Pareto with scale k^* and shape ϵ, where ϵ is to be determined. From equation (5), the stationary distribution $\Phi(k)$ should satisfy

$$\Phi(k) = (1-\delta)\Phi(k) + \delta\mathbb{N}(k) - (1-\delta)E(k)$$

Re-arranging, we have

$$\mathbb{N}(k) - \Phi(k) = \frac{1-\delta}{\delta}E(k)$$

Using the fact that $\Phi(k) = 1 - (\frac{k^*}{k})^{\alpha}$ and $\mathbb{N}(k) = 1 - (\frac{k^*}{k})^{\gamma}$, the above equation can be re-written as

$$(\frac{k^*}{k})^{\alpha} - (\frac{k^*}{k})^{\gamma} = \frac{1-\delta}{\delta}E(k) \tag{*}$$

The probability that an agent with knowledge $x < k$ at time t exits the interval $[k^*, k]$ at $t+1$, is the probability that the agent draws a knowledge greater than k, that is

$$P(k_{t+1} \notin [k^*, k] | k_t = x) = P(k_{t+1} > k | k_t = x) = \int_k^\infty d\mathcal{L}(s; x)$$

The above probability corresponds to a specific knowledge level x. In order to compute $E(k)$, we need to integrate over all the different levels of knowledge in $[k^*, k]$, that is

$$E(k) = \int_{k_m}^{k} P(k_{t+1} > k | k_t = x)\phi(x)dx$$

Since the learning distribution is also Pareto with parameters x and γ, it follows that

$$P(k_{t+1} > k | k_t = x) = (\frac{x}{k})^\gamma$$

This, along with the fact that $\phi(x) = \frac{\epsilon(k^*)^\epsilon}{x^{\epsilon+1}}$, allows us to re-write $E(k)$ as

$$\begin{aligned} E(k) &= \int_{k^*}^{k} (\frac{x}{k})^\gamma \frac{\alpha(k^*)^\epsilon}{x^{\epsilon+1}} dx \\ &= \frac{\epsilon}{\gamma - \epsilon} \frac{(k^*)^\epsilon}{k^\gamma} (k^{\gamma-\epsilon} - (k^*)^{\gamma-\epsilon}) \\ &= \frac{\epsilon}{\gamma - \epsilon} [(\frac{k^*}{k})^\epsilon - (\frac{k^*}{k})^\gamma] \end{aligned}$$

Replacing the expression for $E(k)$ in equation (*) and canceling terms yields

$$\begin{aligned} 1 &= \frac{1-\delta}{\delta} \frac{\epsilon}{\gamma - \epsilon} \\ \implies \epsilon &= \delta\gamma \end{aligned}$$

The stationary distribution with support $[k^*, \infty)$ is Pareto with shape parameter $\delta\gamma$. □

The above proposition presents a very convenient property that follows from having both the newborn, as well as, the learning distribution to be Pareto. This assumption is not totally unreasonable. Since both knowledge and how agents learn are unobserved, we cannot hope to accurately measure these distributions. It is generally believed, however, that the distribution of consumption is log-normal (Hall, 1988). In Dasgupta (2008), we show that if the newborn distribution is truncated exponential, the distribution of consumption is uni-modal with a long upper-tail - characteristics of a log-normal distribution. In fact, this result holds for any newborn distribution whose density is a decreasing function of knowledge. Since this is true for the Pareto distribution as well, our assumption regarding $\mathbb{N}(k)$ is justified. At the same time, a Pareto learning distribution with scale k displays the following features: (a) It is bounded below by k and (b) The mean of the distribution is increasing in k. Both of these are desirable properties for a learning distribution. Since agents in the traditional sector do not learn, $E(k)$ is simply 0, and the part of the stationary distribution below k^* coincides with $\mathbb{N}(k)$. The part of the distribution in the manufacturing sector, $[k^*, \infty)$, is flatter than $\mathbb{N}(k)$. Learning leads to the agents in the modern sector acquiring knowledge over time, thereby shifting mass towards the upper tail.

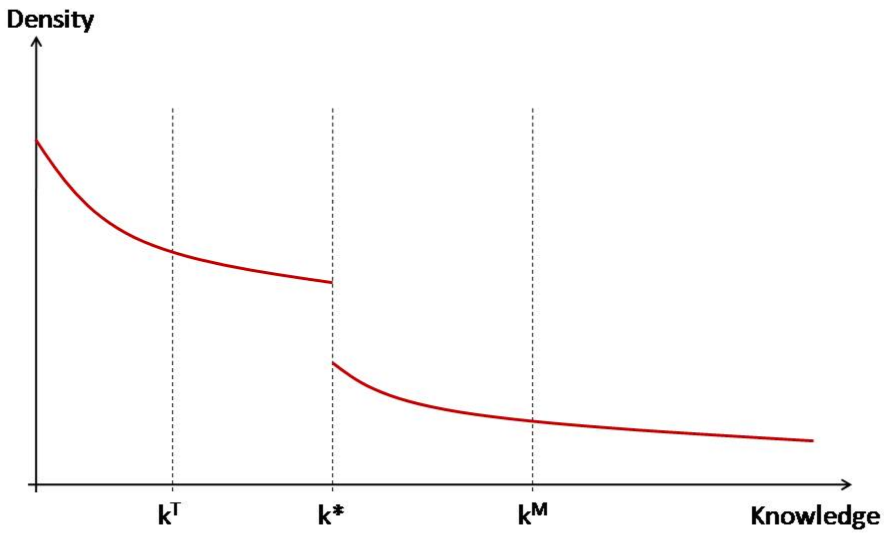

Figure 1: *Steady-state densities*

A possible equilibrium configuration is shown in Figures (1) and (2). Note the discontinuity in the knowledge distribution at k^* in Figure (1) that arises due to learning. In a small interval just to the left of k^*, there is no dynamics. Whatever knowledge an agent is born with, he stays at that knowledge level till he dies. To a small interval to the right of k^*, every period, some agents learn and move out, resulting in a discrete fall in density. The welfare, or present value of income, of agents is plotted in Figure (2). Note that in equilibrium, the value function must be continuous throughout the domain of k. The same, however, is not true of current earnings. In particular, there is a discontinuity in earnings at k^*; the profit that k^* earns if he is a manager in the traditional sector, $\pi^T(k^*)$, is strictly greater than the wage that he earns if he is a worker in the modern sector, w^M. To see why this is the case, suppose that $\pi^T(k^*) = w^M$. Since a traditional sector manager does not learn, he would continue to earn $\pi^T(k^*)$ in every period. A worker in the modern sector,

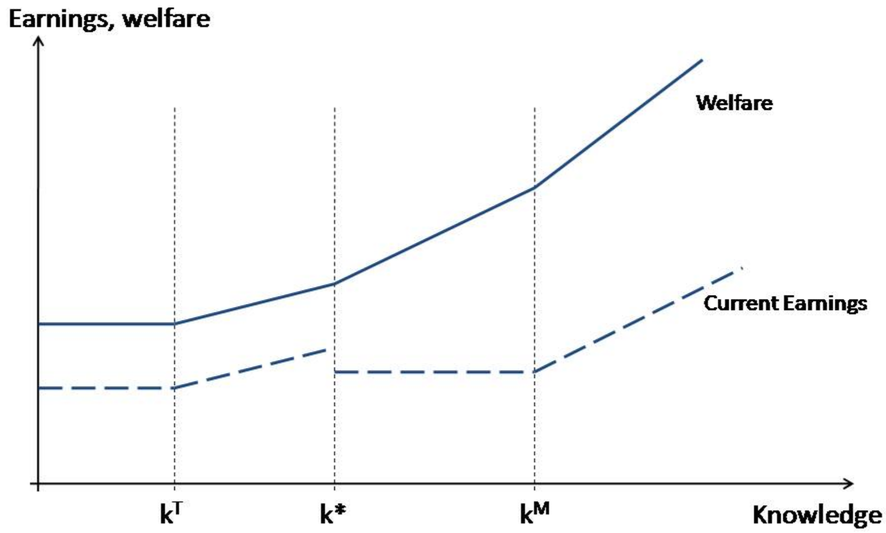

Figure 2: *Steady-state current earnings and welfare*

however, learns almost surely, and will earn strictly more than w^M from the next period. Hence, we will have $V^T_{MA}(k^*) < V^M_{WO}(k^*)$, which cannot be an equilibrium. The same argument shows that a situation with $\pi^T(k^*) < w^M$ cannot be an equilibrium either. It follows that the current earnings of not only the modern sector workers *but also* the least knowledgeable modern sector managers are less than those of the most knowledgeable traditional sector managers, as shown in Figure (2).

Normally, the value functions in the modern sector, $V^M_{WO}(k)$ and $V^M_{MA}(k)$, cannot be solved analytically. But a Pareto learning distribution allows us to get closed-form solutions for both the worker's, as well as, the manager's value functions, as illustrated in the following proposition.

Proposition 1. *In the modern sector, the value function of the manager takes the form* $V^M_{MA}(k) = Ck^{\frac{1}{1-\beta}}$ *while the value function of the worker takes the form* $V^M_{WO}(k) = \frac{w^M}{\delta} + Dk^{\delta\gamma}$*, where*

$$C = \frac{\gamma - \frac{1}{1-\beta}}{\delta\gamma - \frac{1}{1-\beta}} B(w^M)^{\frac{-\beta}{1-\beta}}$$

$$D = (1-\delta)\left(\frac{\gamma}{\gamma - \frac{1}{1-\beta}} C(k^M)^{\frac{1}{1-\beta}} - \frac{w^M}{\delta}\right)(k^M)^{-\delta\gamma}$$

Proof. In the appendix. □

The above proposition suggests that the value functions of both the workers and the managers depends on two endogenous variables - w^M and k^M. In order to solve for these two objects, we need two equations which are readily supplied by the indifference condition at the threshold k^M and the market clearing condition. At k^M, the agent must be indifferent between being a worker and a manager. Accordingly,

$$w^M = \pi^M(k^M)$$

Note that for the agent at k^M to be indifferent, it must be the case that $V^M_{WO}(k^M) = V^M_{MA}(k^M)$. Since both the worker and the manager with knowledge k^M learn the same, however, their continuation value is the same. Hence, it suffices to look at their current earnings. Replacing the value of $\pi(k^M)$ and solving for w^M yields

$$w^M = B^{1-\beta} k^M \tag{10}$$

The labor market clearing condition in the modern sector is given by

$$\int_{k^*}^{k^M} \phi(s)ds = \int_{k^M}^{\infty} n^M(s)\phi(s)ds$$

Replacing the values of $\phi(s)$ and $n^M(s)$ in the above equation yields

$$w^M = \beta[\frac{(k^M)^{\frac{1}{1-\beta}-\delta\gamma}}{(1-\frac{1}{\delta\gamma(1-\beta)})((k^*)^{-\delta\gamma}-(k^M)^{-\delta\gamma})}]^{1-\beta} \tag{11}$$

Figure (3) plots equations (10) and (11) in the (w^M, k^M) plane, for a given value of k^*. The upward-sloping line is the locus of w^M and k^M that makes the agent with knowledge k^M indifferent between being a worker or a manager. Since profits increase in own knowledge, a manager with higher knowledge can be indifferent only if the wage is also higher. The downward-sloping line is the locus of (w^M, k^M) that clears the labor market. *Ceteris paribus*, an increase in k^M implies an excess supply of workers. Market-clearing would then require a fall in wages.

Equating (10) and (11), we can get a closed-form solution for k^M :

$$k^M = [\frac{\delta\gamma - 1}{\delta\gamma(1-\beta) - 1}]^{\frac{1}{\delta\gamma}} k^* \tag{12}$$

Both k^M and w^M are proportional to k^*. Let us re-write equation (12) as $k^M = \zeta(\beta, \gamma, \delta)k^*$. The function ζ is decreasing in γ and δ, but increasing in β. Recall that the invariant distribution $\Phi(k)$ is Pareto with shape $\delta\gamma$ in $[k^*, \infty)$. A smaller γ implies fatter tail for the newborn distri-

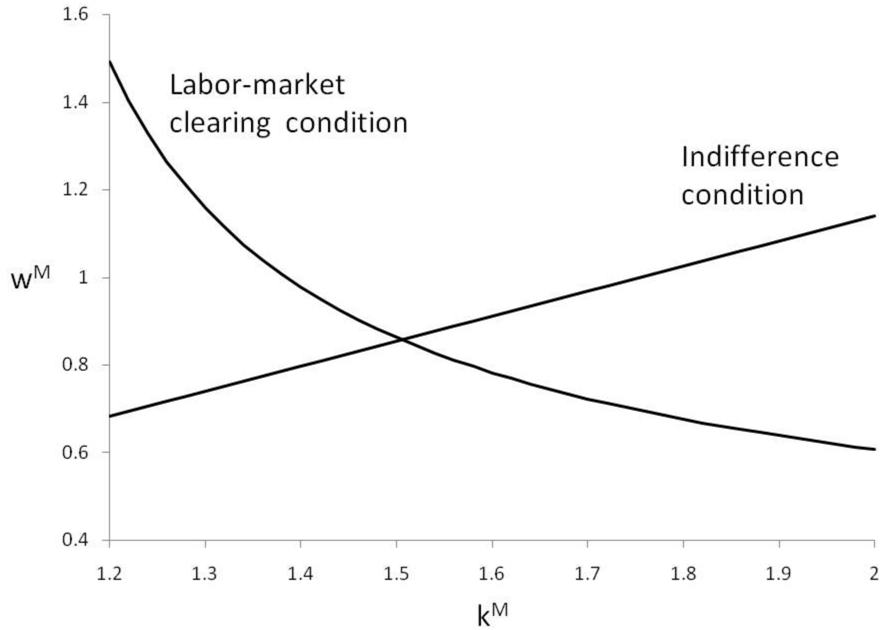

Figure 3: *Determination of* w^M *and* k^M

bution $\mathbb{N}(k)$; that translates into a fatter tail for the invariant distribution. For a given k^*, labor market clearing then requires that k^M must increase. The effect is similar for a decrease in δ. A smaller δ implies that fewer agents are dying every period, resulting in more people moving up along the knowledge distribution. As a result, mass shifts to the upper tail. A fall in β, on the other hand, has an opposite effect. β, by controlling the span of control, determines the demand for labor. A fall in β reduces the aggregate demand for labor in the Modern sector and *ceteris paribus* creates an excess supply of workers. Naturally, k^M has to shift down to clear the market.

The analysis of the traditional sector proceeds in a similar way. The equilibrium wage w^T and the threshold k^T are determined by the indifference condition of the marginal agent and the labor market clearing condition:

$$w^T = A^{1-\alpha} k^T \tag{13}$$

$$\int_{\underline{k}}^{k^T} \phi(s) ds = \int_{k^T}^{k^*} n^T(s)\phi(s) ds \tag{14}$$

The invariant distribution in $[\underline{k}, k^*]$ coincides with the newborn distribution, and is therefore

Pareto with scale parameter γ. Given k^*, equations (13) and (14) can be solved to yield the following implicit function for k^T :

$$\underline{k}^{-\gamma} - (k^T)^{-\gamma} = \frac{\alpha\gamma}{\gamma(1-\alpha)-1}((k^T)^{-\gamma} - (k^T)^{-\frac{1}{1-\alpha}}(k^*)^{\frac{1}{1-\alpha}-\gamma}) \tag{15}$$

Once we solve for k^T, equation (13) gives us the value of w^T that clears the labor market in the traditional sector. Plugging this value into the profit function for the traditional sector manager with knowledge k^*, we get $V^T_{MA}(k^*)$. As discussed earlier, the equilibrium for the entire economy is obtained when the agent with knowledge k^* is indifferent between the two sectors, i.e., $V^T_{MA}(k^*) = V^M_{WO}(k^*)$.

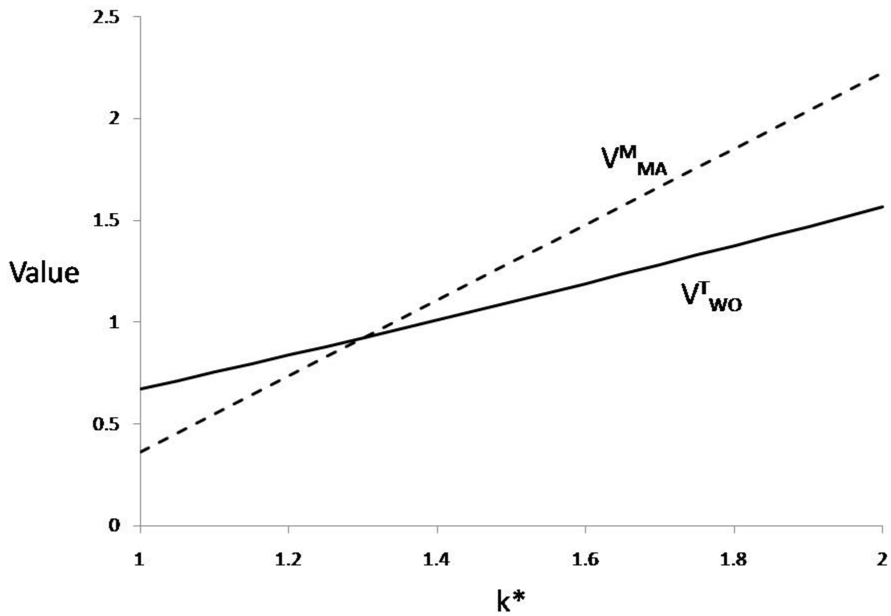

Figure 4: *Determination of k**

Figure (4) shows how the values of $V^T_{MA}(k^*)$ and $V^M_{WO}(k^*)$ change with k^*. Although both the curves are upward-sloping, $V^T_{MA}(k^*)$ is flatter. This is a consequence of the assumption that diminishing returns to labor set in faster in the traditional sector relative to the modern sector ($\alpha < \beta$). For the traditional sector to exist, we must have $V^M_{WO}(\underline{k}) < V^T_{MA}(\underline{k})$, that is, the agent with knowledge $\underline{k}$ must strictly prefer being in the traditional sector. The following proposition formally states the conditions under which the two sectors co-exist:

Proposition 2. *As long as $\underline{k} < k^*$, where k^* solves $V^T_{MA}(k^*) - V^M_{WO}(k^*) = 0$, both the traditional* *and modern sectors co-exist. For a given newborn distribution, co-existence is more likely (i) higher is α and (ii) lower is β.*

A rise in α is like an improvement in the traditional sector technology; it raises the profits of the traditional sector managers and causes the agent with knowledge k^* to switch sectors. Similarly, a higher β results in higher present value of the modern sector workers not only through higher wages but also higher future profits.

4 Firm size distribution

Until now we have been analyzing the optimization problem of agents and its effect on the allocation of agents across different sectors and occupations. In this section, we analyze the implications for the size distribution of firms.

4.1 The Missing Middle

The result that the invariant knowledge distribution in sector M is Pareto allows us to characterize the size distribution of firms in the modern sector, as shown by the following lemma.

Lemma 2. *If the invariant knowledge distribution $\Phi(k)$ is Pareto, then the size distribution of firms is also Pareto.*

Proof. In the appendix. □

As long as the condition for the co-existence of both the sectors, as stated in Proposition (2)are satisfied, Lemma (2) and the assumption that the optimal scale in the two sectors (captured by α and β) are different generate a firm-size distribution where the middle managers are missing. In Table (2), we show two hypothetical firm size distributions generated by our model.

Size of firm	1-3	4-9	10-19	20-49	50-99	>100
$\underline{k} > k^*$						
Employment share	3.1	4.0	4.5	9.2	10.1	69.3
$\underline{k} < k^*$						
Employment share	33.8	8.2	6.1	10.4	8.8	32.7

Table 2: *Employment shares across firms of different sizes*

The first case refers to a situation where the entire economy has been modernized. In this case, the distribution of employment shares across different firms is well-behaved. Notice the similarity with the size distribution of the United States in 1992, as shown in Table (1). The second case is one where the least knowledgeable agents in the economy select into the traditional sector, resulting in the rise of the "missing middle".

4.2 Discussion

A consequence of the indifference condition of the agent at k^* is that w^M is strictly less than $\pi^T(k^*)$. If an agent switches from the traditional to the modern sector, he has to take a cut in current earnings. The modern sector, however, provides the agent with the opportunity to learn which raises his future earnings. Since learning is a function of own knowledge, for the least knowledgeable agents, the higher expected earnings is not enough to compensate for the current drop in earnings. Accordingly, these agents choose to enter the traditional sector. Two related points need to be mentioned at this time. Firstly, a popular view among some academics is that individuals in many developing countries voluntarily choose to be in the informal sector, presumably because earnings in the informal sector are higher than the average formal sector wage.[7] According to the ILO (1981), "a majority of the entrepreneurs (in the informal sector) earned an income substantially higher than that implied by the legal minimum wage" (ILO, page 195). It goes on to conclude that "there is little evidence to show that the participants in this sector hold out for better jobs in the formal sector. In other words, virtually all of them seem to view the informal sector as a permanent source of employment and income" (ILO, page 198). The above analysis suggests that such a view is erroneous.

In our model, the gap between current earnings in the traditional and the modern sector in favor of the former arises naturally in equilibrium. This outcome, however, is not because the traditional sector is "better" in some sense, but because of learning in the modern sector. Although, some of the agents who select into the traditional sector necessarily earn more than the modern sector workers, they are worse-off in the long-run. This brings us to our second point which deals with the welfare properties of the equilibrium. The equilibrium in our model is first-best efficient due to the absence of market distortions.[8] The traditional sector is not the "last resort"

[7] Portes and Sassen-Koob (1987) mention studies in Colombia and Brazil in the mid-70s that found the income of self-employed to be significantly higher than the average wage in the modern sector. The authors go on to conclude that there are "significant income variation among the informally employed including income levels that exceed, at times, the average among formal workers"(page 37).

[8] Even in the absence of knowledge spillovers, the equilibrium is only first-best efficient due to the overlapping generations framework. The presence of spillovers does not change this property because learning depends on the newborn distribution and not the stationary distribution.

in the sense that agents, who fail to get a modern sector job, fall back on self-employment. The least knowledgeable *optimally* select into the traditional sector. Consequently, the role for policy in eliminating this dual economic structure is limited in this model. This is in contrast to models/explanations that rely on distortions to generate the dual economic structure.

4.3 Development and Convergence

Development, in the context of our model, is a reduction in the cumulative density of the newborn distribution. We interpret the newborn distribution as arising from the schooling choices made by individuals when they are young. The newborn distribution is, in fact, the distribution of knowledge across agents when they enter the job market. We assume that as a country develops, the educational attainment of its population goes up, leading to higher mean knowledge. Looking across countries, the relation between economic development and educational attainment is unmistakable. This relation is displayed in Figure (5), which plots per capita GDP and average years of schooling for a sample of countries. The GDP data is from the Penn World Tables while the schooling data is the one collected by Barro and Lee (1993). In spite of a lot of variation, on average the population in richer countries seem to have higher average years of education.

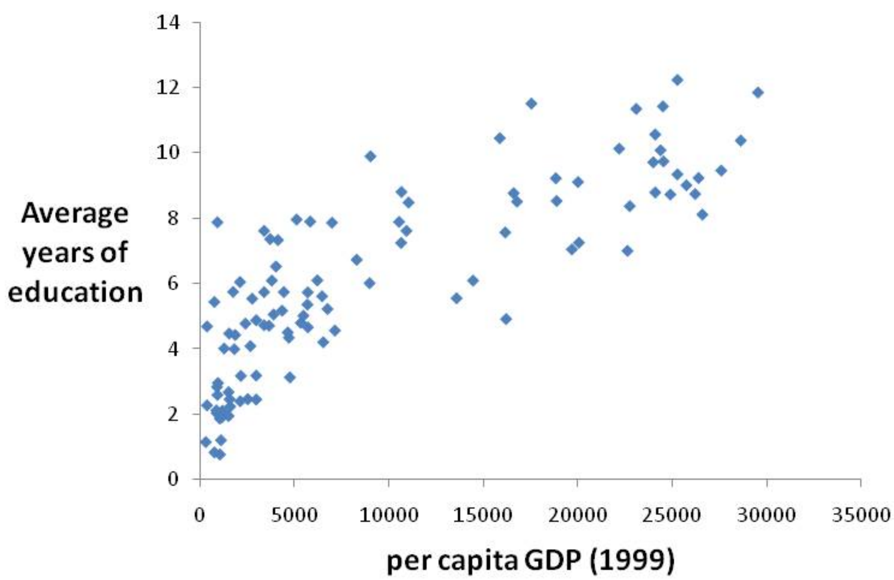

Figure 5: *The relation between average years of education and per capita GDP for a sample of countries in the year 1999*

With a Pareto newborn distribution, the cumulative density depends on both scale $\underline{k}$ and shape γ. We hold $\underline{k}$ constant and interpret development as a reduction in γ. Consequently, $\mathbb{N}_{t+1}(k)$

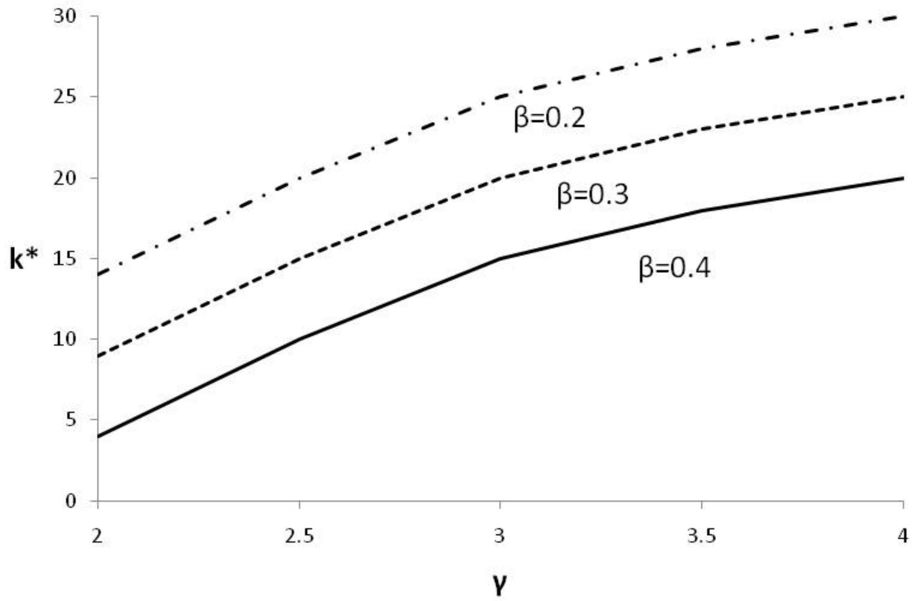

Figure 6: *How k^* changes with γ and β*

dominates $\mathbb{N}_t(k)$ in the first-order stochastic dominance sense. Recall from Proposition (2), that the condition for the existence of the traditional sector is $\underline{k} < k^*$. As k^* falls, the traditional sector contracts until it disappears completely. From our discussion in Section (3) we know that k^* depends on γ in a complex way. As a result, it is difficult to sign $\frac{\partial k_m}{\partial \gamma}$ analytically. Figure (6) shows that for a given value of β, a reduction in γ leads to a reduction in k^*. As γ falls and the mean knowledge in the economy increases, two opposing forces come into play. On the one hand, there is an abundance of workers in the modern sector, which tends to pull down the wage. On the other hand, agents learn more on average, since they are surrounded by more knowledgeable people. This raises their continuation value. The latter force dominates and the value function of the worker in the modern sector shifts up. Consequently, the knowledge at which an agent is indifferent between the two sectors goes down. The mass of agents in the traditional sector declines both because the support of the distribution under the traditional sector shrinks and the distribution becomes flatter.

From lemma (2), we know that the optimal firm size of a manager is increasing in β. A higher β improves the ability of a manager to hire more workers. For a given wage, this increases the demand for labor; the market-clearing wage rises, thereby raising the present value of all the workers. Consequently, k^* shifts down.

How do we know that it is the improvement in educational attainments that is responsible for the change in the firm size distribution over time? Answering this question is beyond the scope of the paper, partly due to a lack of reliable data on firm size distributions in developing countries. But we present some evidence that is consistent with our proposed explanation. Table (1) suggests that both Indonesia and Taiwan saw their size distributions gradually evolving from a bi-modal to an uni-modal distribution during the 70s and 80s. As shown in Figure (7), the average schooling in both of these countries went up during this period. In Figure (8), we plot the evolution of the knowledge distribution in Indonesia and Taiwan. 0 on the horizontal axis refers to no schooling, 1 is primary, 2 is secondary and 3 is post-secondary. It is clear from the figures that in both countries, the mass of the distribution has been slowly shifting towards higher levels of knowledge.

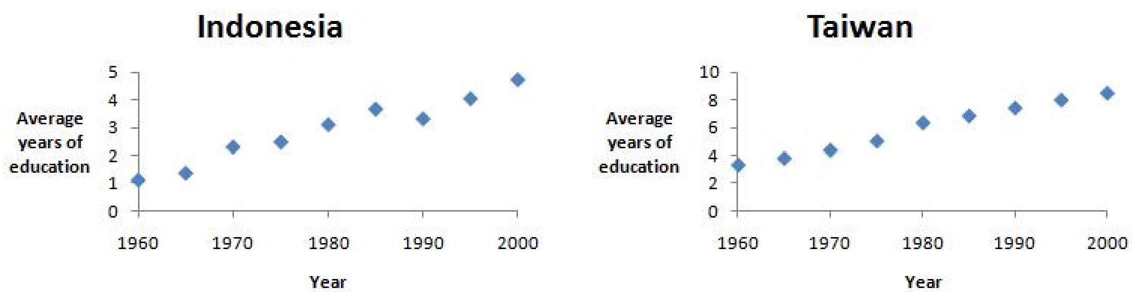

Figure 7: *Change in the average years of education*

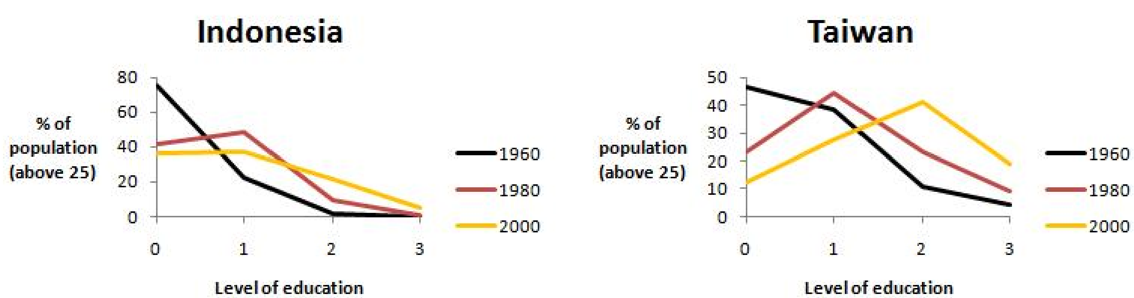

Figure 8: *Evolution of the distribution of education*

Our choice of changing γ while holding $\underline{k}$ constant follows from the observation that the support of the knowledge distribution does not seem to be changing over time. For example, Figure (8) shows that although the distributions in both countries has been changing over time, the lower bound has not changed. In fact, even the most developed countries of the world have some individuals who have no schooling (Barro and Lee, 1993). An implication of holding $\underline{k}$ constant over time is that the traditional sector disappears *in spite* of there being some agents with very low levels of knowledge in the economy. An increase in average knowledge, by raising the learning opportunities of agents in the modern sector, enables individuals hitherto in the traditional sector to move into the modern sector. At the risk of oversimplification, our model seems to suggest that modernization of the economy does not require educating each and every less knowledgeable agent in the economy; educating only a fraction suffices. It also suggests that individuals will choose a level of education that is inefficiently low.

5 Conclusion

The firm size distribution in many developing countries is bi-modal, with a large concentration of extremely small firms. Moreover, the distribution seems to be converging to an uni-modal distribution, at least in some countries. In this paper, we have developed a simple model in order to explain these two facts. In our model, heterogeneous agents choose between two sectors, traditional and modern, and two occupations, being a worker or a manager. We show that if the mean knowledge in the economy is not too high, an equilibrium allocation of agents emerges where the least knowledgeable agents select into the traditional sector and become self-employed, whereas the most knowledgeable become managers in the modern sector and hire those with intermediate levels of knowledge. The firm-size distribution exhibits a "missing middle" because some agents, instead of running their own firms, become workers in others' firms. The equilibrium is also characterized by an earnings gap between most knowledgeable managers in the traditional sector and the least knowledgeable workers in the modern sector. This does not, however, imply that the workers in the modern sector are worse-off. On the contrary, these workers learn and the present value of their earnings is actually higher than their counterparts in the traditional sector. Over time, as the mean knowledge of an economy rises, the traditional sector shrinks, till it eventually disappears.

We would like to conclude with two observations. Firstly, one could easily introduce a spatial dimension in the current model by assuming that the two sectors are located at different points in space - the city is characterized by modern form of production while the more traditional methods are practiced in the rural area. In this case, the model provides another motive for migration. Agents migrate from rural to urban areas because of knowledge spillovers in cities.[9] In the countryside, agents are spatially dispersed. On the other hand, agents are highly concentrated in cities. As a result, there are frequent interactions leading to a transfer of knowledge between agents, most of which are not priced (See Robert E. Lucas, 2004, for a similar explanation).

[9] See Jaffe et al. (1993) for evidence on knowledge spillovers in cities.

Secondly, the model has a strong policy implication. Providing income transfers to the least knowledgeable might raise their welfare, but this is balanced by the tax that the government has to impose on the more knowledgeable agents to finance these transfers. On the other hand, taxing the population and then spending the receipts on educating the newborns clearly raises aggregate welfare because the knowledge acquired by an agent not only raises his own productivity, but also the productivity of those around him. In this context, the opening quote of our paper becomes relevant.

References

Robert J. Barro and Jong-Wha Lee. International comparisons of educational attainment. *Journal of Monetary Economics*, 32(3):363–394, 1993.

Luis M. B. Cabral and Jose Mata. On the evolution of the firm size distribution: Facts and theory. *American Economic Review*, 93(4):1075–1090, 2003.

Kunal Dasgupta. Learning, knowledge diffusion and the gains from globalization. Working paper, University of Toronto, 2008.

Robert E Hall. Intertemporal substitution in consumption. *Journal of Political Economy*, 96(2): 339–57, April 1988.

Barton H. Hamilton. Does entrepreneurship pay? an empirical analysis of the returns to self-employment. *Journal of Political Economy*, 108(3):604–631, 2000.

ILO. Women and men in the informal economy: A statistical picture, 2002.

Adam B Jaffe, Manuel Trajtenberg, and Rebecca Henderson. Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics*, 108(3): 577–98, 1993.

Dale W. Jorgenson. Surplus agricultural labour and the development of a dual economy. *Oxford Economic Papers*, 19(3):288–312, 1967.

Arthur W. Lewis. Economic development with unlimited supplies of labour. *Manchester School of Economic and Social Studies*, 22(2):139–191, 1954.

Dipak Mazumdar. Segmented labor markets in ldcs. *The American Economic Review*, 73(2): 254–259, 1983.

Alejandro Portes and Saskia Sassen-Koob. Making it underground: Comparative material on the informal sector in western market economies. *The American Journal of Sociology*, 93(1): 30–61, 1987.

Gustav Ranis and John C. H. Fei. A theory of economic development. *The American Economic Review*, 51(4):533–565, 1961.

James Rauch. Modelling the informal sector formally. *Journal of Development Economics*, 35 (1):33–47, 1991.

Jr. Robert E. Lucas. Life earnings and rural-urban migration. *Journal of Political Economy*, 112 (S1):S29–S59, 2004.

S.V. Sethuraman ed. The urban informal sector in developing countries (international labour office), 1981.

James R. Tybout. Manufacturing firms in developing countries: How well do they do, and why? *Journal of Economic Literature*, 38(1):11–44, 2000.

Appendix

Proof of Proposition 1 : The profit of a manager with knowledge k is given by $\pi(k) = kn^{\beta} - wn$. Replacing the optimum value of n, we have $\pi(k) = Bw^{\frac{-\beta}{1-\beta}}k^{\frac{1}{1-\beta}}$, where $B = \beta^{\frac{1}{1-\beta}}(\beta^{-1} - 1)$. Let us guess that $V_M(k) = C_1 + C_2 k^{\lambda_m}$.

Now,

$$
\begin{aligned}
\int_k^{\infty} V_M(k')d\Gamma(k') &= \int_k^{\infty}(C_1 + C_2 k^{\lambda_m})\frac{\gamma k^{\gamma}}{k'^{\gamma+1}}dk' \\
&= C_1\gamma k^{\gamma}\int_k^{\infty} k'^{-\gamma-1}dk' + C_2\gamma k^{\gamma}\int_k^{\infty} k'^{\lambda_m-\gamma-1}dk' \\
&= C_1 + \frac{C_2\gamma}{\gamma-\lambda_m}k^{\lambda_m}
\end{aligned}
$$

where the last line follows under the condition $\lambda_m - \gamma < 0$. As we shall, this is in fact true in our model. Therefore the RHS of equation (9) is

$$
Bw^{\frac{-\beta}{1-\beta}}k^{\frac{1}{1-\beta}} + (1-\delta)(C_1 + \frac{C_2\gamma}{\gamma-\lambda_m}k^{\lambda_m})
$$

This must be equal to $C_1 + C_2k^{\lambda_m}$. Matching the coefficients and after a bit of algebra, we have

$$
\begin{aligned}
\lambda_m &= \frac{1}{1-\beta} \\
C_1 &= 0 \\
C_2 &= \frac{\gamma-\lambda_m}{\delta\gamma-\lambda_m}Bw^{\frac{-\beta}{1-\beta}} = C
\end{aligned}
$$

Replacing the value of λ_m from above, $\delta\gamma - \lambda_m = \delta\gamma - \frac{1}{1-\beta}$. If Assumption 1 holds, $\delta\mu - \lambda_m > 0$. Then $\delta < 1$ implies that $\mu - \lambda_m > 0$, as claimed earlier.

For the worker's value function, we guess that $V_W(k) = D_1 + D_2k^{\lambda_w}$. Now,

$$
\begin{aligned}
\int_k^{k^*} V_W(k')d\Gamma(k') &= \int_k^{k^*}(D_1 + D_2k^{\lambda_w})\frac{\gamma k^{\gamma}}{k'^{\gamma+1}}dk' \\
&= D_1 - D_1k^{*-\gamma}k^{\gamma} + \frac{\gamma}{\gamma-\lambda_w}D_2k^{\lambda_w} - \frac{\gamma}{\gamma-\lambda_w}D_2k^{*\lambda_w-\gamma}k^{\gamma}
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\int_{k^*}^{\infty} V_M(k')d\Gamma(k') &= \int_{k^*}^{\infty} Ck'^{\lambda_m}\frac{\gamma k^\gamma}{k'^{\gamma+1}}dk' \\
&= \frac{\gamma}{\gamma-\lambda_m}Ck^{*\lambda_m-\gamma}k^\gamma
\end{aligned}
$$

Re-arranging all the terms, the RHS of equation (8) can be written as

$$
w + (1-\delta)D_1 + (1-\delta)\Xi k^\gamma + (1-\delta)\frac{\gamma}{\gamma-\lambda_w}D_2k^{\lambda_w}
$$

where $\Xi = -D_1k^{*-\gamma} - \frac{\gamma}{\gamma-\lambda_w}D_2k^{*\lambda_w-\gamma} + \frac{\gamma}{\gamma-\lambda_m}Ck^{*\lambda_m-\gamma}$.

The above expression must be equal to $D_1 + D_2k^{\lambda_w}$. Matching the coefficients, we have $D_1 = \frac{w}{\delta}$. Also, note that $\lambda_w \neq \gamma$. Otherwise, the above expression explodes. Therefore, equating the coefficients of k^{λ_w}, we have $\lambda_w = \delta\gamma$. Furthermore, Ξ must be equal to 0, that is

$$
D_1k^{*-\gamma} + \frac{\gamma}{\gamma-\lambda_w}D_2k^{*-\lambda_w-\gamma} = \frac{\gamma}{\gamma-\lambda_m}Ck^{*-\lambda_m-\gamma}
$$

Solving for D_2 yields

$$
D_2 = (1-\delta)(\frac{\gamma}{\gamma-\lambda_m}Ck^{*\lambda_m} - \frac{w}{\delta})k^{*-\lambda_w}
$$

This completes our proof.

Proof of Proposition 2 : The first-order condition from the manager's profit-maximization problem yields the following optimal value for n

$$
\begin{aligned}
n &= (\frac{\beta}{w})^{\frac{1}{1-\beta}}k^{\frac{1}{1-\beta}} \\
&= (\frac{\beta}{w})^{\frac{1}{1-\beta}}z
\end{aligned}
$$

where $k = z^{1-\beta} = h(z)$. Let k be distributed as Pareto with scale k_m and shape α. The density $\phi(k)$ is given by

$$
\phi(k) = \begin{cases} \frac{\alpha k_m^\alpha}{k^{\alpha+1}} & \text{if } k > k_m \\ 0 & \text{otherwise} \end{cases}
$$

Using the formula for transforming probability densities, the density of z can be obtained as

$$
\begin{aligned}
g(z) &= \phi(h(z))\frac{dh(z)}{dz} \\
&= \frac{\alpha k_m^{\alpha}}{z^{(1-\beta)(\alpha+1)}}(1-\beta)z^{-\beta} \\
&= \frac{\widetilde{\alpha} z_m^{\widetilde{\alpha}}}{z^{\widetilde{\alpha}+1}}
\end{aligned}
$$

where $\widetilde{\alpha} = \alpha(1-\beta)$ and $z_m = k_m^{\frac{1}{1-\beta}}$. Hence, z also follows a Pareto distribution with parameters z_m and $\widetilde{\alpha}$. Since n is a linear function of z, the size distribution of firms also follows a Pareto distribution.