University of Toronto
Department of Economics

Working Paper 410

Learning and Knowledge Diffusion in a Global Economy

By Kunal Dasgupta

September 30, 2010

Learning and Knowledge Diffusion in a Global Economy

Kunal Dasgupta *
University of Toronto

First draft : November 2008
This draft : September 2010

Abstract

I develop a dynamic general equilibrium model to understand how multinationals affect host countries through knowledge diffusion. Workers in the model learn from their managers and knowledge diffusion takes place through worker mobility. Unlike in a model without learning, I present a novel mechanism through which an integrated equilibrium represents a Pareto improvement for the host country. I go on to explore other dynamic consequences of integration. The entry of multinationals makes the lifetime earning profiles of host country workers steeper. At the same time, if agents learn fast enough, integration creates unequal opportunities, thereby widening inequality. The ex-workers of foreign multinationals also found new firms which are, on average, larger than the largest firms under autarky.

KEYWORDS : Multinationals, knowledge diffusion, learning, worker mobility, Pareto improvement, spin-offs.
JEL Code : F15, F23, F40.

*I am indebted to Esteban Rossi-Hansberg, without whose guidance and support this paper would never have been possible. I am also grateful to Gene Grossman and Marc Melitz for their valuable advice and encouragement. I would also like to thank Pol Antràs, Thomas Chaney, Arnaud Costinot, Gilles Duranton, Cecilia Fieler, Luis Garicano, Stefania Garetto, Hugo Hopenhayn, Marc Muendler, Jordi Mondria, Peter Morrow, Xianwen Shi, Colin Stewart, Daniel Trefler, Vinayak Tripathi, Jonathan Vogel, and especially Robert McMillan, Alexander Monge-Naranjo and Stephen Redding for their helpful comments and suggestions. Financial support from the International Economics Section, Princeton University is greatly appreciated.
E-mail : *kunal.dasgupta@utoronto.ca*

1 Introduction

One of the most common arguments in favor of multinational (MNE) activity in developing countries is that knowledge can thereby diffuse from Northern MNEs to Southern firms. This belief is reflected in the widespread use of investment incentives by many host country governments to attract prospective MNE investors (Oman, 2000). Yet, how does this knowledge diffuse to domestic firms? And what are the consequences of such diffusion for the domestic economy? In an era of unprecedented globalization, answers to these questions have taken on a great deal of policy relevance.

This paper presents a model that sheds light on the impact of MNE entry on welfare, wages and occupational choice in the presence of knowledge diffusion through worker mobility. Specifically, I develop a dynamic general equilibrium model, where agents have different levels of knowledge. Workers and managers get together in firms to produce output and knowledge. Complementarity between the worker's and manager's knowledge in the production and learning technologies leads to positive assortative matching (or PAM), whereby more knowledgeable workers team with more knowledgeable managers to produce and learn.[1] The equilibrium is characterized by a threshold level of knowledge, such that every agent below the threshold is a worker while those above are managers. The combination of PAM and learning, however, implies that every agent who starts his life as a worker works for better and better managers until he himself becomes a manager, provided that he survives long enough.

Globalization or integration allows managers to hire workers from other countries.[2] I consider a two country model, where the Foreign (or Northern) country has relatively more knowledgeable agents (in a sense to be made precise shortly) compared to the Home (or Southern) country. Following integration, new teams are formed as Foreign managers try to leverage their superior knowledge with respect to Home workers. In this setting, I identify two effects that determine Home wages. First, integration increases the competition for workers, which tends to raise wages. This is the labor demand effect. This alone would make some of the incumbent managers worse off. But there is also a new effect: the entry of MNEs creates the possibility for the workers to be matched with more knowledgeable managers. By working for the MNEs, workers can learn and earn more than under autarky. The result that MNEs hire more knowledgeable workers, however, implies that the less knowledgeable workers can expect to work for the MNEs in the future only if they learn from the less knowledgeable Home managers, and this creates a rent. Since learning is fully foreseen by the agents, the managers extract part of this rent by paying lower wages and thereby internalize the knowledge "spillover". This is the learning effect. If agents learn fast enough, this effect dominates and the wage schedule shifts down by enough to make the incumbent managers better off. The workers are better off too, because the increase in their continuation

[1] In their study of variation in management practices across firms and countries, Bloom and Van Reenen (2010) find that that the education of both workers and managers is strongly correlated with management scores; more educated workers are hired by more educated managers.

[2] In this paper, MNEs are synonymous with international production teams. I abstract from the issue related to the boundaries of international firms. For some recent papers which deal with this issue, see Antràs (2003), Antràs and Helpman (2004) and Grossman and Helpman (2003).

value outweighs the reduction in current wage. The above mechanism through which integration can lead to a Pareto improvement in the host country is new in the literature.

The model is explored further in the numerical section. The learning dynamics of the model imply that the lifetime earnings profiles of agents born as workers are upward sloping. The slope of the earnings profile depends on the matching function. By improving the matches, integration increases the amount of knowledge that agents can acquire in each period, thereby raising the gradient of the lifetime earnings profiles. This is a novel prediction that could be brought to data.

Since the distribution of knowledge is endogenous, the model also allows me to talk meaningfully about inequality and how it changes following integration. Unlike most of the existing models, which are confined to studying wage inequality, aggregate inequality in this model can be measured on the basis of per capita consumption. As is well known, inequality measures based on consumption or expenditure are more appropriate than measures based on income. More importantly, a change in inequality in this model reflects not only a change in the agents' earnings but also a change in the entire knowledge distribution following integration, a feature that is absent in static models of trade or FDI. I show that if agents learn fast enough, integration can increase aggregate inequality.[3] This increase in inequality is partly driven by the fact that individuals in an integrated economy differ in terms of the opportunities they face. Even if everyone gains from integration, the more knowledgeable workers gain relatively more as they have immediate access to Foreign managers. These workers become managers and experience a jump in earnings much sooner than workers who are born with low levels of knowledge.

Another novel prediction of the model concerns the spin-offs from the MNEs. Among the new managers entering the economy every period, a fraction consists of those who were previously working in other firms. I show that the biggest and most productive firms in the Home economy are, on average, run by Foreign managers. Combined with PAM, this implies that founders of spin-offs who have previous MNE experience are, on average, more knowledgeable than those who do not. Preliminary evidence from Ghana provided by Görg and Strobl (2005) seems to confirm this prediction. Furthermore, some of these spin-offs are larger and more productive than the largest firms under autarky.

That MNEs diffuse knowledge through labor turnover is well-documented. In one of the more formal studies, Poole (2006) provides evidence of knowledge spillovers through worker mobility in Brazil. The literature, however, abounds with case studies and anecdotal evidence. Giarratana et al. (2004) look at the spin-offs from MNEs that were created in India after the country liberalized in 1991 and find that the founders brought a high level of technological expertise from the MNEs to the new firms. In China, many potential managers perceive the MNEs as schools where they can train themselves; many of them leave to start their own business, once they have the required expertise.[4] Easterly (2001) discusses the Korean company Daewoo's decision to train the workers of a Bangladeshi textile firm in 1979; most of

[3] See Goldberg and Pavcnik (March 2007) for evidence on episodes of globalization, which include both trade and FDI liberalization, being accompanied by increasing inequality. IMF (2007) reports that inward FDI has exacerbated inequality in developing countries.

[4] See "China's people problem", *The Economist*, 14th April, 2005.

them left the parent firm during the 1980s to start their own garment export firms, laying the foundation for the $2 billion dollars Bangladeshi garment industry.[5] Despite being prevalent, knowledge diffusion from MNEs through worker mobility has received relatively little attention among economists. This paper tries to fill the gap.

My paper combines two strands of the existing literature. Kremer and Maskin (2006) and Antràs et al. (2006) use matching models to analyze how international team formation affects earnings and welfare. Those models, however, are static in nature, with the distribution of knowledge being given exogenously. I allow workers to learn from their managers, thereby extending the above mentioned models to a dynamic environment. By imposing quite weak restrictions on the production and learning technologies, I am able to characterize the equilibrium of both the closed as well as the integrated economies. In the process, I not only confirm some of their results in a more general setting, but generate new insights as to the nature of welfare gains, inequality and worker mobility.[6]

Monge-Naranjo (2007) and Beaudry and Francois (2010) develop dynamic general equilibrium models involving multinationals and on-the-job learning. Those models do not involve matching however. A key feature of Monge-Naranjo is costly skill acquisition by individuals. By assuming learning by observing within firms, I instead focus on matching and endogenous occupational choice. Therefore, my paper is complementary to his.

In terms of the firm's problem, my paper is closer to Beaudry and Francois. In their model, skills are transferred from managers to workers through learning by observing. As in my model, firms offer a bundle of wage and skill to workers, who trade current wage with continuation value. The authors are primarily concerned, however, with analyzing why managers from developed countries may not relocate to developing countries, despite a shortage of skills in the latter. This is different from my paper in that I focus on the impact of MNEs on domestic skill accumulation and its consequences, assuming that there are sufficient incentives for foreign managers to relocate.

Despite similarities, my model differs from the above mentioned papers in one key respect. In Monge-Naranjo (2007), workers get only one chance to learn (since he has a two period overlapping generations model), while in Beaudry and Francois (2010), workers immediately become managers upon learning.[7] The combination of long-lived agents, PAM and stochastic learning implies that workers in my model can learn throughout their lives and two workers with identical initial conditions can end up with very different life-time earnings profiles. Besides generating richer dynamics at the individual level, this makes it possible for even the less knowledgeable domestic workers to work in MNEs in the future,

[5]Evidence on founders of spin-offs inheriting knowledge from their parents in the U.S. economy is provided by Klepper (2002) for the automobile industry, Klepper and Sleeper (2005) for the laser industry, and Filson and Franco (2006) for the rigid disk drive industry.

[6]For quantitative models that compute static welfare gains associated with multinational production see Ramondo (2008), Garetto (2008) and Burstein and Monge-Naranjo (2009). Rodríguez-Clare (2007) develops a model of trade and diffusion where growth is caused by technological progress. Unlike my model, however, diffusion of ideas is an exogenous process.

[7]Chari and Hopenhayn (1991) develop of model of technology diffusion involving learning by doing at the firm level. After acquiring technology-specific skills, agents can choose whether they want to work as skilled or unskilled workers. Unlike my model, however, agents in their model live for two periods.

thereby opening up a new channel for welfare gain.

The remainder of the paper proceeds as follows. Section 2 describes the model while in Section 3, I study the properties of a stationary equilibrium. In Section 4, I analyze how integration affects matches, output and welfare in the host country. I study a numerical example in Section 5 and use it to further characterize the equilibrium. Section 6 concludes. All the proofs are in the Appendix.

2 The Model

My model introduces learning and dynamics to a framework that is similar to Antràs, Garicano and Rossi-Hansberg (2006) - henceforth AGR. Time is discrete.

2.1 Preferences and Endowments

There is a continuum of heterogeneous agents with different levels of knowledge. Knowledge is embodied in an agent, but can be acquired through interactions, i.e., an agent can learn from others. One can think of knowledge as some composite of different attributes that affects an agent's productive capability.[8][9] A newborn agent draws his knowledge from an exogenously given distribution $\Phi(k)$ with support $[\underline{k}, \overline{k}]$. I assume that $\Phi(k)$ is continuous, with $\phi(k)$ being the corresponding density. Agents also die every period with a constant probability δ and are replaced by newborns such that the population is constant. The actual distribution of knowledge at time t is denoted by $\Psi_t(k)$, with $\psi_t(k)$ being the corresponding density. Agents are risk neutral. Since the size of the population plays no role in the analysis, it is normalized to 1.

2.2 Production

Firms produce a single, non-storable good. A firm comprises of a manager and production workers. Workers do routine jobs and each worker combines with the manager to produce $f(y)$ units of output, where y is the knowledge of the manager. Thus, "f(y) captures the indivisibility of management-type decisions and implies a scale economy because it improves productivity of all the workers in the firm, irrespective of their numbers" as in Rosen (1982) p. 314. Notice that the productivity of workers in a firm run by a manager with knowledge y is simply $f(y)$. The manager pays wages to the workers and is the residual claimant on the output.[10]

There is a technological restriction to the number of workers a manager can hire. The span of control of a manager depends only on the knowledge of the workers he hires. This span of control is given by

[8] In the standard Mincerian wage equation, the right-hand side consists of education, as well as, experience. In this paper, knowledge encompasses both, along with other unobservables.

[9] For a trade model where agents have two attributes, see Ohnsorge and Trefler (2007).

[10] Here, as in Monge-Naranjo (2007), I assume that there is no difference between the managers and entrepreneurs. For a model which makes this distinction, see Holmes and Schmitz (1990).

$\tilde{n}(\min_i[x_i])$, where x_i is the knowledge of the i-th worker hired by a manager.[11] It is easy to see that given such a technology, the manager will hire only one type of worker in equilibrium. I denote the equilibrium span-of-control by $n(x)$ and use this notation in the rest of the paper.[12] Total output of a firm is then given by

$$q = f(y)n(x) \tag{1}$$

I make the following assumptions regarding the production technology:

ASSUMPTION 1a : *f is continuous, strictly increasing and weakly convex in y; n is continuous, strictly increasing and strictly concave in x. Furthermore, $\frac{\partial}{\partial y}[\frac{f'(y)}{f(y)}] \leq 0$.*

ASSUMPTION 1b : $\frac{f'(\bar{k})}{f(\bar{k})} > \frac{n'(\underline{k})}{n(\underline{k})}$.

Assumptions 1a and 1b together imply that output elasticity of the managerial task is greater than that of the production task at any knowledge level. As we shall see, this results in the more knowledgeable agents becoming managers in equilibrium. [13] Assumption 1b also says that for a given knowledge distribution, there should be sufficient asymmetry between the manager and the worker's contribution to output. This is a technical condition required for the existence and uniqueness of equilibrium.

The assumption that labor is the only factor of production is without loss of generality. Capital can be easily introduced in the model. The cost of capital has three components - sunk costs, fixed costs and variable costs. In the absence of uncertainty in production and credit market imperfections, the first two do not really have any effect; so I just normalize those to zero. As for variable capital requirement, think of it as being subsumed in $f(y)$.

2.3 Learning

Agents also learn in firms. Since the seminal work of Gary Becker (Becker, 1962), economists have been studying on-the-job training. In this paper, I abstract from formal training provided by firms and instead focus on the knowledge that workers acquire while producing. I follow Jovanovic and Rob (1989) in defining the learning technology. Specifically, within each firm, a worker learns from the manager.[14] Learning is stochastic and depends both on the knowledge of the manager and the worker. The random-

[11] This is similar to the O-ring production function (Kremer, 1993), whereby the workers in a firm are as good as their least skilled counterpart.

[12] For an alternative micro-foundation of such a technology, see Garicano (2000).

[13] To see this, note that the output elasticity of the manager's knowledge is $\frac{\partial q}{\partial y}\frac{y}{q} = \frac{f'(y)y}{f(y)}$, while that of the worker is $\frac{n'(x)x}{n(x)}$. Assumption 1b says that $\frac{f'(y)}{f(y)}$ is non-increasing in y while Assumption 1a implies that the same is true for $\frac{n'(x)}{n(x)}$. Moreover, $\frac{f'(\bar{k})}{f(\bar{k})} > \frac{f'(\bar{k})}{f(\bar{k})}\frac{f'(\underline{k})}{f'(\underline{k})} > \frac{n'(\underline{k})}{n(\underline{k})}$. This inequality, combined with the previous observation, implies that $\frac{f'(k)k}{f(k)} > \frac{n'(k)k}{n(k)}$ $\forall k$.

[14] Unlike Jovanovic and Rob (1989), learning is one-sided. Assuming that managers also learn from workers could be an interesting extension and would be one channel through which growth can be introduced in this model.

ness in learning does not necessarily reflect any randomness in the knowledge transfer process but rather, is a simple way of modeling the heterogeneous capacity to absorb knowledge. A worker with knowledge x at time t has knowledge x' at time $t+1$. The learning distribution is given by $L(x'|x,y)$, $x' \in [x,y]$. For all $h(x')$ increasing in x', I make the following assumptions about the learning technology:

ASSUMPTION 2a : $\frac{\partial}{\partial x}\int h(x')dL(x'|x,y) > 0$.

ASSUMPTION 2b : $\frac{\partial}{\partial x}\int h(x')dL(x'|x,y) > 0$.

ASSUMPTION 2c : $\frac{\partial^2}{\partial x \partial y}\int h(x')dL(x'|x,y) \geq 0$.

The first two conditions are the familiar ones for first-order stochastic dominance. These conditions imply that expected learning is increasing in the knowledge of both the workers and the managers. The third assumption says that there is complementarity in learning. Although the learning technology is taken as given, it can be derived from a micro-founded model where learning requires effort and workers optimally choose how much effort to allocate (see Appendix).

3 Equilibrium

Agents are price-takers. There are two prices in the economy. First, the managers hire workers and pay a price for their marginal product. Second, the workers learn from the managers and pay a price for the acquired knowledge. It is inconsequential whether there are two transactions within the firm or whether the managers simply pay the wage net of the rent (Rosen, 1972). What matters is the net payment to workers $w_t(k)$; let us call this wage. Note that the wage is a function of the knowledge of the worker, but *not* the manager. I shall return to this issue shortly.

The absence of aggregate uncertainty in the model, combined with a large number of agents, implies that the evolution of the knowledge distribution is deterministic. Since in equilibrium $w_t(k)$ is a function of only the knowledge distribution, the evolution of $w_t(k)$ is also deterministic. Therefore, given an initial distribution of knowledge $\Psi_0(k)$, a competitive equilibrium is characterized by a deterministic sequence $\{\Psi_t(k), w_t(k)\}$. Given a sequence of wage functions $\{w_t\}_{t=0}^{\infty}$, the manager's problem is then defined recursively as

$$V_M(y,w_t) = \max_x f(y)n(x) - w_t(x)n(x) + (1-\delta)\max[V_W(y,w_{t+1}), V_M(y,w_{t+1})]. \qquad (2)$$

where $V_W(y,w_t)$ is the value function of an agent with knowledge y, if he chooses to be a worker while $V_M(y,w_t)$ is the value function if, instead, he chooses to be a manager.[15] The value of a manager

[15]Notice the absence of time discounting in the above formulation. This is because a positive probability of death acts as a

depends on the current distribution $\Psi_t(k)$ through the net wage schedule w_t, where the dependence of wage on k is understood. The second term on the right allows for the possibility that an agent, who is a manager at time t, might choose to be a worker at time $t+1$. $V_W(y, w_t)$ is given by

$$V_W(x, w_t) = w_t(x) + (1-\delta) \int_x^{m_t(x)} \max[V_W(x', w_{t+1}), V_M(x', w_{t+1})] dL(x'|x, m_t(x)). \quad (3)$$

where $m_t(x)$ is the knowledge of the manager who hires a worker with knowledge x in equilibrium. The term within the integral denotes the expected value of the worker if he works for a manager with knowledge $m_t(x)$. Depending on how much he learns, the worker might become a manager or continue as a worker at $t+1$. As before, this decision will depend not only on the worker's own knowledge but also on the wage function at $t+1$. Let the disconnected sets of workers be denoted by W_i, $i = 1, ...S$, where S is determined in equilibrium, and let $\overline{W} = \bigcup_{i=1}^{S} W_i$. Similarly, let the disconnected sets of managers be denoted by M_i, with $\overline{M} = \bigcup_{i=1}^{S} M_i$.[16] Then the labor market-clearing condition can be written as

$$\int_{\overline{W}} \psi_t(k) dk = \int_{\overline{M}} n(m^{-1}(k)) \psi_t(k) dk \quad (4)$$

The left-hand side denotes the total supply of workers. The right-hand side denotes the total demand for workers, where $n(m^{-1}(k))$ is the number of workers demanded by a manager with knowledge k.

Over time, workers' knowledge increases due to learning. But agents also die every period with probability δ and are replaced by newborns who draw knowledge from the exogenous distribution $\Phi(k)$. Birth, learning and death implies a rule for the evolution of the knowledge distribution $\Psi_t(k)$:

$$\Psi_{t+1}(k) = \delta\Phi(k) + (1-\delta) \int_{\underline{k}}^{k} \int_{s}^{k} dL(s'|s, m_t(s)) d\Psi_t(s) \text{ for all } k \in [\underline{k}, \overline{k}] \quad (5)$$

The first term on the right-hand side denotes the fraction of agents who are born in period $t+1$ with knowledge less than k. The second term denotes the agents who remain below k in period $t+1$, despite learning from their managers in period t. Equation (5) implies that Ψ_{t+1} is determined by how individuals acquire knowledge in period t, and the acquisition of knowledge by individuals is determined only by who they match with at time t, which in turn depends only on Ψ_t. Therefore, Ψ_{t+1} is a function of Ψ_t.

Definition 1. *A competitive equilibrium of this economy consists of the following objects:*

discount factor.

[16] For example, we could have the following allocation : $W_1 = [\underline{k}, k_1]$, $M_1 = [k_1, k_2]$, $W_2 = [k_2, k_3]$, $M_2 = [k_3, \overline{k}]$. Then, $\overline{W} = [\underline{k}, k_1] \bigcup [k_2, k_3]$ and $\overline{M} = [k_1, k_2] \bigcup [k_3, \overline{k}]$.

(i) Value functions, $V_W(k) : \overline{W} \to \mathbb{R}$ *and* $V_M(k) : \overline{M} \to \mathbb{R}$*;*
(ii) Current earnings, $w_t(k) : \overline{W} \to \mathbb{R}$ *and* $\pi_t(k) : \overline{M} \to \mathbb{R}$*;*
(iii) Matching function, $m_t(k) : \overline{W} \to \overline{M}$*;*
(iv) Occupational structure, $\overline{W}$ *and* $\overline{M}$*, such that*
(a) $V_W(k)$ *and* $V_M(k)$ *satisfy the worker's and manager's problems respectively;*
(b) $m_t(k)$ *is the corresponding policy function;*
(c) labor market clears;
(d) the knowledge distribution evolves according to equation (5).

It might seem natural to write the wage as $w_t(x, y)$ given that (a) the same manager can produce different levels of output by hiring different types of workers and (b) the same worker can acquire different levels of knowledge by working for different managers. Condition (a) suggests that the price of labor should be specific to a worker-manager pair while (b) suggests that the same should be true for the price of knowledge. In order to understand why $w_t(x)$ only has the knowledge of the worker as its argument, let us look at the underlying mechanism that determines the wage function.

In this economy, every agent with knowledge y, in the role of a manager, offers a "gross" wage schedule $\widetilde{w}_t(x, y)$ such that y *is indifferent between hiring any* x. $\widetilde{w}_t(x, y)$ is the wage offered by y if there were no learning and captures the worker's pay-off from production. The following lemma establishes some properties of $\widetilde{w}_t(x, y)$.

Lemma 1. $\widetilde{w}_t(x, y)$ *is increasing in* x *for all* y*, and* $\frac{\partial^2 \widetilde{w}}{\partial x \partial y} > 0$.

At the same time, each agent, in the role of a worker, offers a rent schedule $\widetilde{r}_t(x, y)$ such that he is indifferent across managers. $\widetilde{r}_t(x, y)$ is what the workers would pay to the managers if there were no production; it reflects the value of learning. The following lemma establishes a key property of the rent function.

Lemma 2. $\frac{\partial^2 \widetilde{r}}{\partial x \partial y} \geq 0$.

Unlike $\widetilde{w}_t(x, y)$, which is the solution to a static problem, workers compute $\widetilde{r}_t(x, y)$ by taking into account the entire expected earnings profile. As Lemma 1 and 2 shows, the gross wage, as well as, the rent functions have positive cross partials. Each of these forces individually would lead to positive assortative matching (or PAM), whereby more knowledgeable workers work for more knowledgeable managers.[17] Having both of them acting in the same direction re-enforces the result. This is stated formally in the next lemma.

Lemma 3. $m_t(x)$ *is strictly increasing in x.*

Although PAM imposes some structure on the equilibrium allocation, the occupational structure is still too general (recall that S, the number of sets of workers and managers, could potentially take any

[17] This is standard feature of worker-manager relations that exhibit complementarity (Becker, 1973).

value). I claim that in equilibrium, there exists a threshold k_t^* such that all agents with knowledge less than k_t^* are workers, while those with knowledge above k_t^* are managers. That is, I claim that $S = 1$. Under threshold matching, equation (4) reduces to

$$\int_{\underline{k}}^{k} d\Psi_t(s) = \int_{k_t^*}^{m_t(k)} n(m_t^{-1}(s)) d\Psi_t(s) \quad \forall k \leq k_t^* \tag{6}$$

Note that the labor market-clearing condition is not standard. The left hand side denotes the supply of workers in the interval $[\underline{k}, k]$, while the right hand side denotes the demand for workers coming from managers in the interval $[k_t^*, m_t(k)]$. Measure consistency requires that these two values be equal *for every* k. This follows from lemma 3, since the workers hired by managers with knowledge in $[k_t^*, m_t(k)]$, must have knowledge in $[\underline{k}, k]$. Differentiating equation (6) with respect to k yields

$$m_t'(k) = \frac{\psi_t(k)}{n(k)\psi_t(m_t(k))} \tag{7}$$

The above differential equation, along with the boundary conditions $m_t(\underline{k}) = k_t^*$ and $m_t(k_t^*) = \overline{k}$, allows us to solve for the matching function. As the following lemma shows, given $\Psi_t(k)$, the threshold k_t^* and consequently the matching function is uniquely determined.

Lemma 4. *For a given $\Psi_t(k)$, k_t^* exists and is unique.*

A convenient property of this equilibrium is its block recursive structure - matches can be determined completely once we know the knowledge distribution. One does not need to know the wage schedule in order to determine the matches; rather, once the matches are determined, wages adjust so as to support the matches that emerge. Of course, this does not mean that the way agents match does not depend on wages. In this economy, wages (and profits) not only determine the remuneration of the agents but they also play an allocative role (Sattinger (1993)). But for the purpose of solving the model, the matching function can be derived without any information on the wage function.

In this model, a worker with knowledge x working for a manager with knowledge y, receives $\widetilde{w}_t(x, y)$ but has to pay back $\widetilde{r}_t(x, y)$. Accordingly, he earns a wage of $\widetilde{w}_t(x, y) - \widetilde{r}_t(x, y)$. In equilibrium, the worker earns $\widetilde{w}_t(x, m_t(x)) - \widetilde{r}_t(x, m_t(x)) = w_t(x)$. Therefore, *in equilibrium*, the wage depends only on the worker's knowledge. PAM also allows me to derive some properties of the value functions.

Lemma 5. *$V_W(k)$ and $V_M(k)$ exist, are continuous and increasing in k.*

This completes the characterization of the competitive equilibrium. The following lemma provides for the existence and uniqueness of the equilibrium.

Lemma 6. *There exists a δ^*, such that $\forall \delta > \delta^*$, a threshold equilibrium exists and it is unique.*

Finally, recall that equation (5) defines Ψ_{t+1} as a function of Ψ_t. We seek a fixed point of Ψ_t, i.e., an invariant knowledge distribution Ψ^*. As the following lemma shows, such a fixed point exists and is unique.

Lemma 7. *A unique, invariant knowledge distribution Ψ^* exists and any initial distribution Ψ_0 weakly converges to Ψ^*.*

Therefore, in the long run, the knowledge distribution converges to Ψ^*, with threshold k^*. Agents who are born with knowledge above k^* become managers instantaneously. Since managers do not learn, these agents are stuck with the level of knowledge they are born with. On the other hand, agents who are born with knowledge below k^* start their lives as workers. These agents learn every period and move up, until they eventually cross the threshold and become managers themselves. For these agents, the lifetime earning profiles are positively sloped. For the remainder of the paper, I shall restrict most of the attention to the stationary equilibrium of the model, although transitional dynamics are briefly considered in Section 5.

4 Analytical Results

In this section, I present some analytical results of my model. To simplify exposition, I make the following assumption about the learning technology:

ASSUMPTION 2d : *If, in period t, a worker with knowledge x works for a manager with knowledge y, then in period $t+1$, the worker's knowledge x' could be x with probability θ and y with probability $1-\theta$.*

Thus, learning is an all-or-nothing proposition for the worker. In the next section, I relax this assumption and work with a more general learning technology. Notice that, in spite of the learning distribution having just two points, it still satisfies Assumptions 2a, 2b and 2c.[18]

4.1 Autarky

I begin by examining the equilibrium under autarky. Recall that the density function for the newborn distribution is given by $\phi(k)$. The learning technology, along with the newborn distribution, implicitly defines the invariant distribution and allows me to solve for the threshold k^*.

Proposition 1. *k^* is defined implicitly by the following equation*

[18] To see this, note that for any $h(.), h' > 0$, the expected value of $h(x')$ is $\theta h(x) + (1-\theta)h(y)$. Therefore, $\frac{\partial E[h(x')]}{\partial x} = \theta h'(x)$ and $\frac{\partial E[h(x')]}{\partial y} = (1-\theta)h'(y)$. Furthermore, $\frac{\partial^2 E[h(x')]}{\partial x \partial y} = 0$.

$$\int_{\underline{k}}^{k^*} \frac{\phi(k)}{n(k)} dk = \int_{k^*}^{\overline{k}} \phi(k) dk + (1-\theta)(\frac{1-\delta}{\delta}).$$

Moreover, k^* has the following properties :

$$\frac{\partial k^*}{\partial \theta} < 0; \frac{\partial k^*}{\partial \delta} < 0.$$

Proposition 1 sheds light on how the distribution changes as the rate of learning increases. δ, being the probability of death in a period, proxies for the length of a time period. A lower δ, holding θ unchanged, implies that agents are acquiring the same expected knowledge over a smaller interval of time. On the other hand, a lower θ, holding δ unchanged, implies that agents are acquiring more expected knowledge over the same interval of time. Both these cases translate into faster learning for the agents. An increase in the rate of learning makes the knowledge distribution negatively-skewed, as more and more mass shifts to the upper tail. Consequently, labor market-clearing requires that the threshold shift to the right.

Recall that a worker with knowledge k produces $f(m(k))$ units of output. Hence, total output produced in this economy is given by

$$Y = \int_{\underline{k}}^{k^*} f(m(k)) d\Psi(k) \tag{8}$$

Total welfare is given by

$$W = \int_{\underline{k}}^{k^*} V_W(k) d\Psi(k) + \int_{k^*}^{\overline{k}} V_M(k) d\Psi(k) \tag{9}$$

In this model, individual welfare equals the present value of consumption (or income, since the good is non-storable) because agents are risk-neutral.

4.2 Integration

Integration, in the context of my model, means that managers from one country can hire workers in another country, i.e., integration leads to the creation of MNEs. The managerial input is rival and as a result, managers cannot operate plants in both countries.[19] The motive behind the formation of MNEs is exploiting differences in factor prices.[20] In this paper, I focus on full integration, i.e., I assume that

[19] Whether managers travel from the source-country to the host-country or not, however, is irrelevant.

[20] This motive for establishing subsidiaries in other countries is the same as in Helpman (1984).

MNEs are formed costlessly. In particular, I assume away any cost that might be associated with opening a plant in another country. I do acknowledge that these costs are important, but the introduction of such costs increases the complexity of the model without any gain in insight.

Let us introduce some notation. Define the subscripts $i = \{A, I\}$, $j = \{H, F\}$, where A and I stand for autarky and integration respectively, while H and F stand for Home and Foreign respectively. The Home newborn distribution is denoted by $\Phi_H(k)$ with support $[\underline{k}, \overline{k}_H]$, while the Foreign newborn distribution is $\Phi_F(k)$, with $[\underline{k}, \overline{k}_F]$ being the corresponding support. I assume that $\overline{k}_F > \overline{k}_H$ and that $\Phi_F(k)$ first-order stochastic dominates $\Phi_H(k)$. The latter assumption reflects the relative abundance of more knowledgeable agents in the Foreign country.[21] The steady-state knowledge distributions are indexed by i and j. So, for example, $\Psi_{A,H}(k)$ is the Home steady-state knowledge distribution under autarky. I also assume that the two countries have the same population. With integration, the fundamental change is in the distribution of newborns, which is given by

$$\Phi_I(k) = \begin{cases} \frac{1}{2}\Phi_H(k) + \frac{1}{2}\Phi_F(k) & \text{for } k \in [\underline{k}, \overline{k}_H] \\ \frac{1}{2} + \frac{1}{2}\Phi_F(k) & \text{for } k \in [\overline{k}_H, \overline{k}_F] \end{cases} \tag{10}$$

$\Phi_I(k)$, combined with the learning technology, determines the integrated knowledge distribution $\Psi_I(k)$. The new threshold, k_I^*, would typically be different from k_A^*, the autarky threshold. Before deriving the relation between the thresholds under autarky and integration, let us state the following result.

Lemma 8. *If a knowledge distribution G first-order stochastic dominates another distribution H, then $k_G^* > k_H^*$, where k_G^* and k_H^* are the thresholds under G and H respectively.*

Equation (10), along with the assumption that $\Phi_F(k)$ first-order stochastic dominates $\Phi_H(k)$, implies that $\Phi_I(k)$ first-order stochastic dominates $\Phi_H(k)$. In the benchmark case of no-learning, the knowledge distributions in both the countries coincide with the newborn distributions. Consequently, under no-learning, $k_I^* > k_{A,H}^*$ (this follows directly from Lemma 6). With learning, however, the knowledge distributions are no longer exogenous. Still, one can derive a relation between $k_{A,H}^*$ and k_I^*, as shown in the following proposition.

Proposition 2. *$k_I^* > k_{A,H}^*$, where k_I^* is defined implicitly by the following equation*

$$\int_{\underline{k}}^{k_I^*} \frac{\phi_H(k) + \phi_F(k)}{2n(k)} dk = \frac{1}{2} \int_{k_I^*}^{\overline{k}_F} (\phi_H(k) + \phi_F(k)) dk + (1-\theta)(\frac{1-\delta}{\delta}).$$

[21] In this paper, I want to highlight the dynamic gains to workers in developing countries arising from the access to better managers from developed countries. As Bloom and Van Reenen (2010) point out, better managers (or better management practices, to be precise) are relatively more abundant in the U.S. compared to developing countries like India or China.

The range of knowledge for the Home workers expands under integration. The agents with knowledge in $[k^*_{A,H}, k^*_I]$ switch from being managers to workers. The entry of highly knowledgeable Foreign managers raises the opportunity cost of being a manager for a Home agent. This is not only due to (possibly) higher wages paid by the MNEs but also due to the better learning opportunities provided by the MNE managers. An incumbent Home manager weighs the cost of becoming a worker for a MNE (forgone current profits) against the benefit (higher expected profits in the future). For the managers in $[k^*_{A,H}, k^*_I]$, benefits outweigh costs and consequently they switch.

Although Proposition 2 indicates the direction of change for the threshold, it says nothing about its magnitude. In particular, the following two scenarios are possible:

Case I ($k^*_I > \overline{k}_H$) : In this case, every agent born in the Home country starts his life as a worker. The support of $\Psi_{I,H}(k)$ is $[\underline{k}, m(\overline{k}_H)]$, despite the fact that, the Home newborn distribution $\Phi_H(k)$ still has the smaller support.[22] Though theoretically an interesting case, this situation is quite extreme because it implies that integration results in the destruction of all incumbent firms (managers), who are replaced by a new class of bigger and more productive firms.

Case II ($k^*_I < \overline{k}_H$) : In this case, the support of $\Psi_{I,H}(k)$ is $[\underline{k}, \overline{k}_F]$. This case is characterized by the birth of a new class of Home firms (with knowledge in $[\overline{k}_H, \overline{k}_F]$), who are on par with the Foreign MNEs in terms of size and productivity. But unlike Case I, a set of incumbent Home managers with knowledge in $[k^*_I, \overline{k}_H]$ continues to operate in the integrated economy.

Whether we are in Case I or Case II depends on the parameters of the model. As long as $\overline{k}_F$ is not too different from $\overline{k}_H$, there will be some incumbent managers in the Home country.[23] Intuitively, a large gap between $\overline{k}_F$ and $\overline{k}_H$ implies that following integration, the Home agents have an opportunity to work for very knowledgeable managers. This is also true for every incumbent Home manager, who would rather work in Foreign MNEs, learn and become much better managers in the future than remain managers with low levels of knowledge.

Irrespective of which case we are in, integration affects the matching of agents. An immediate implication of Proposition 4 is that $m_I(\underline{k}) > m_{A,H}(\underline{k})$, where $m_{A,H}(.)$ and $m_I(.)$ are the matching functions under autarky and integration respectively.[24] Therefore, the least knowledgeable worker in the Home country, and by continuity, a set of less knowledgeable workers, is matched with better managers. This is formally stated in the following proposition:

Proposition 3. *A positive measure of Home workers have a better match in the integrated equilibrium compared to autarky.*

[22] It is not the case that every Home agent is a worker. There are Home managers in $[k^*_I, m(\overline{k}_H)]$. This, however, means that the Home managers in the integrated economy have knowledge greater than $\overline{k}_H$.

[23] Note that for a given $\overline{k}_H$, there exists a k' such that $\overline{k}_F < k'$ implies that $k^*_I < \overline{k}_H$. This follows from the result that k^*_I is monotone increasing in $\overline{k}_F$, and $k^*_I < \overline{k}_H$ when $\overline{k}_F = \overline{k}_H$.

[24] To see this, note that $m_A(\underline{k}) = k^*_A$ and $m_I(\underline{k}) = k^*_I$. Proposition 4 then gives the result.

On the other hand, some of the Home managers are now matched with less able workers. Since the output of firms depends positively on workers' knowledge, the output of some of the Home firms under integration are necessarily lower than in autarky.

Corollary 1. *Under integration, the output of a positive measure of Home firms goes down.*

Note that since the output produced by a worker depends only on the knowledge of the manager he is matched with, the productivity of a firm, as measured by the value-added per worker, does not change.

Total Home output (GDP) produced in the integrated equilibrium is given by

$$Y = \int_{\underline{k}}^{k_I^*} f(m_I(k)) d\Psi_{I,H}(k) P_H \tag{11}$$

This is different from Gross National Income (GNI), which is given by

$$GNI = \int_{\underline{k}}^{k_I^*} w_I(k) d\Psi_{I,H}(k) P_H + \int_{k_I^*}^{\overline{k}_F} \pi_I(k) d\Psi_{I,H}(k) P_H \tag{12}$$

The difference between the two arises because in an integrated equilibrium, a part of the Home output goes to the Foreign country as profits of Foreign MNEs, while some of the Home firms may become multinationals and earn profits from their operations in the Foreign country. Finally, aggregate welfare in the Home country is given by

$$W = \int_{\underline{k}}^{k^*} V_W(k) d\Psi_{I,H}(k) P_H + \int_{k^*}^{\overline{k}_F} V_M(k) d\Psi_{I,H}(k) P_H \tag{13}$$

To sum up, with integration, the threshold of the knowledge distribution shifts to the right. This necessarily means that some of the Home workers are hired by more knowledgeable managers. These workers also learn more compared to autarky. At the same time, some of the incumbent firms suffer a decline in output.

4.3 Change in Welfare

Integration changes individual, as well as, aggregate welfare of the Home country. I focus my attention on the case where there are surviving Home managers, i.e., Case II.[25] In order to understand how learning affects welfare, first let us look at the benchmark case of no learning. This is similar to the static framework presented in AGR. A key result that emerges from AGR is that integration raises aggregate

[25]The reason for this is the following : If $\overline{k}_F$ is very different from $\overline{k}_H$, then irrespective of whether agents learn or not, every Home agent is better off working for the more knowledgeable Foreign managers. Thus we get Pareto improvement, but the Home firms disappear completely.

consumption, and with risk-neutral agents, the aggregate welfare of the Home economy. What about individual welfare? In the previous section, we showed that the output produced by the less knowledgeable Home managers goes down under integration.[26] The actual change in profits and welfare, however, depends on the wages they pay, which would be different from those under autarky. Of course, as wages change, the welfare of the workers change too.

Proposition 4. *In the absence of learning, an integrated steady-state equilibrium with incumbent Home firms can never represent a Pareto improvement relative to the autarky steady-state equilibrium in the Home country.*

In the absence of learning, integration creates winners and losers. The identity of the winners and losers, though, will depend on the specific parameter values. If we think of workers and managers as two separate factors of production, Proposition 4 essentially gives us a Heckscher-Ohlin like result.[27]

Does Proposition 4 continue to hold when we introduce learning? In order to prove otherwise, I have to show that every agent in the Home economy is strictly better off under integration. Corollary 1 implies that some of the Home managers earn lower revenue compared to autarky.[28] Hence, for these managers to be better-off under integration, the wage bill has to go down more than revenue.

In this model, there are two forces that determine wages. First, there is a *labor demand effect*. The entry of Foreign managers increases the demand for Home workers. At the same time, integration increases competition faced by the Home workers from their Foreign counterparts. As shown by AGR, (1) if the two countries are not too similar and, (2) if the span of control is not too small, the labor demand effect raises the wages of all Home workers.

Second, there is a *learning effect*. A worker, in this model, can be hired by any manager with a positive probability. Working for a more knowledgeable manager means higher expected learning and consequently, higher earnings. Hence, the entry of highly knowledgeable Foreign managers raises the continuation value of the Home workers. PAM, however, implies that the most knowledgeable managers hire only the most knowledgeable workers. Therefore, the less knowledgeable workers can work for the MNEs only if they learn and acquire enough knowledge from their current managers, some of whom are the incumbent Home managers. A positive value of learning implies that workers are willing to pay in order to learn. Thus learning creates a rent. This allows the managers to compress the wage. The workers accept this wage reduction because they expect to be compensated in the future. So the learning effect tends to lower the wage schedule.

The final impact on wages depends on the relative strengths of the two effects. The following Proposition shows the condition under which the Home economy realizes Pareto gains.

[26] This is true for both the learning and no-learning case.

[27] To be technically correct, we have infinitely many factors.

[28] Since there is only one good, output equals revenue.

Proposition 5. *The integrated steady-state equilibrium is a Pareto improvement for the Home country if*

$$(1-\theta)(\frac{1-\delta}{\delta}) > \frac{2f(\overline{k}_H)[n(\overline{k}_H)-n(\underline{k})]+n(\underline{k})[\mu(\overline{k}_H)f(\overline{k}_H)-\mu(\underline{k})f(\underline{k})]}{\mu(\underline{k})[f(\overline{k}_F)+f(\underline{k})]n(\underline{k})-\mu(\overline{k}_H)[n(\underline{k})+n(\overline{k}_H)]f(\overline{k}_H)}$$

where $\mu(\underline{k}) = \frac{n(\underline{k})}{1+n(\underline{k})}$ *and* $\mu(\overline{k}_H) = \frac{n(\overline{k}_H)}{1+n(\overline{k}_H)}$.

Let us denote by Ω, the set of pairs of θ and δ that satisfy the above condition. The left-hand side of the above expression is positive by definition. The numerator of the fraction on the right-hand side is positive too.[29] For Ω to be non-empty, the denominator has to be positive and large enough. I can show that the denominator is an increasing function of $\frac{f'(k)}{f(k)} - \frac{n'(k)}{n(k)}$, i.e., the degree of asymmetry between the manager's and the worker's contribution to output. Intuitively, the greater is this asymmetry, the greater is the increase in the worker's earning when he becomes a manager; and the greater is the wage cut that the worker is willing to accept in order to learn.

Proposition 5 also sheds light on how welfare changes as the rate of learning changes. Assuming that the right-hand side of the expression in Proposition 5 is positive, the inequality is not satisfied for high enough δ (or θ). In the limiting case of no-learning, $\delta = 1$ (or $\theta = 1$), the left-hand side is equal to zero. As δ (or θ) falls, the left-hand side starts to increase and at some point, exceeds the right-hand side. According to Corollary 1, some of the incumbent firms produce less under integration relative to autarky. For these firms to be better off, they must be paying lower wages to the workers. Corollary 2 follows naturally.

Corollary 2. *If all Home agents gain from integration, some of the Home workers must earn a lower wage compared to autarky.*

Pareto gains are important for political economy reasons. Gaining support for trade or FDI liberalization becomes much easier if every agent gains from integration. Notice, however, that proposition 5 compares welfare across two steady-states. It is possible that even though the new steady-state generates a Pareto improvement, the policy lowers the welfare of some agents relative to autarky. In the next section, however, we show that integration generates a Pareto improvement not only across steady-states, but also at the instant the policy is enacted.

4.4 Remarks

Few remarks are in order. First, the rent-sharing in the model takes place without any bargaining between the workers and the managers. Rather, the division is determined by labor market clearing. Two assumptions of the model are key for this result - fixed supply of agents and endogenous occupational choice. To see this, first consider a model without learning. In equilibrium, the agent with the threshold

[29] Since $\frac{n(k)}{1+n(k)}$ is increasing in k.

knowledge k^* earns the same wage and profit and accordingly, has the same value from both occupations. Now assume that agents can learn. Can we have a situation where the workers extract the entire rent from learning? No, because then the value from being a worker will be higher than the value from being a manager for the threshold agent; although the current wage and profit are the same, the continuation value of a worker will be higher than the continuation value of a manager because the worker, almost surely, will be a manager with knowledge greater than k^* in the next period. This, however, cannot be an equilibrium since the agent with knowledge k^*, and by continuity, agents with knowledge just above k^*, will strictly prefer to be workers. Given the fixed supply of agents, this will create an excess supply workers. For the labor market to clear, wages would have to be pushed down. The amount by which wage falls for each worker is the rent that he pays to the manager. A similar argument shows that managers cannot extract the entire learning rent either.

Second, if a MNE has more knowledge than an incumbent Home firm, PAM implies that the workers in the MNE are more knowledgeable than the ones in the Home firm. PAM also implies that after working for the (more knowledgeable) MNE, a worker never works for the (less knowledgeable) Home firm. Therefore, there is no flow of knowledge from the MNE to the Home firm. Despite this, the incumbent firm could be better off if learning is fast enough.[30] Of course, some of the former MNE workers set up their own firms and these managers directly benefit from the superior knowledge of MNEs.[31]

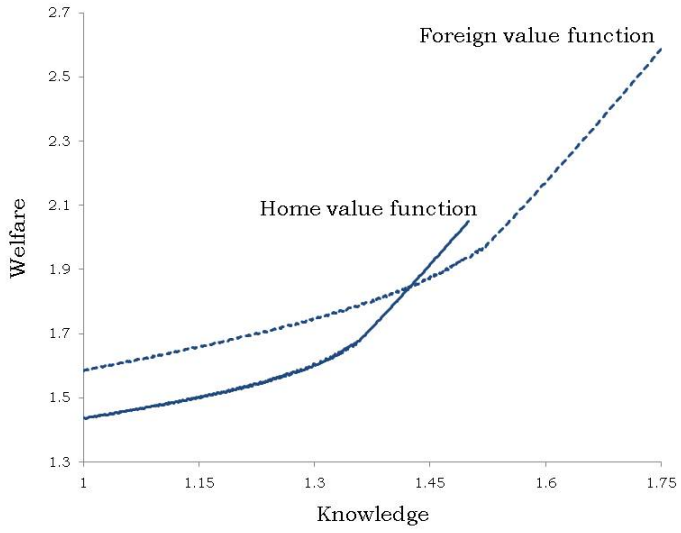
Third, knowledge in this model has only one dimension, i.e., it is completely general. I abstract from firm-specific knowledge along the lines of Hashimoto (1981) and Carmichael (1983) among others. Unlike Beaudry and Francois (2010), I also assume that the same knowledge can be used for both production and management. This assumption is necessary for a key feature of the model - workers can smoothly move from one firm to another, until they become managers themselves. This assumption, however, can be relaxed without sacrificing the tractability of the model. One can introduce a parameter σ, where $\sigma \leq 1$ measures the fraction of the knowledge that a worker acquires in a firm, that would be useful in other firms. In this case, all the qualitative results would go through, but with lower welfare gains. The scenario where all knowledge is firm-specific would correspond to $\sigma = 0$. In this case, there would be no movement of workers across firms just as in a static model.

And finally, Corollary 1 suggests that some of the incumbent Home firms will be producing less following integration. If agents learn fast enough, the current earnings of some of the Home workers is also lower (Corollary 2). Hence, simply looking at current wages or output might give the impression that workers and firms in the Home country are worse off following integration when, in fact, they could all be better-off. My model therefore suggests that one should interpret lower wages or output with caution, especially when drawing conclusions about welfare.

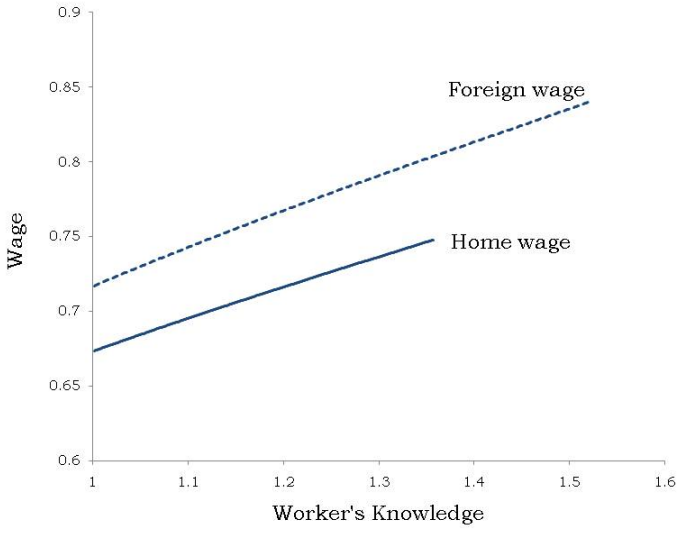
[30] The traditional view regarding knowledge spillover is that workers with experience in MNEs are hired by domestic managers. These workers bring with them knowledge regarding better technology and management practices and this raises the productivity of the domestic firms. See, for example, Barba Navaretti and Venables (2004).

[31] This effect is similar to Monge-Naranjo (2007) where the transfer of skills from MNEs materialize in a new sector of firms, not in the pre-existing sector of firms.

5 Numerical Results



(a) *Welfare under autarky*



(b) *Wages under autarky*

Figure 1: *Home and Foreign economy under autarky*

In this section, I parametrize the model and solve it numerically. The exercise serves two purpose.

First, it allows visual representation of the results from the previous section, thereby strengthening the intuition behind them. Second, it allows us to derive additional dynamic results which cannot be obtained analytically. Throughout, the focus is on the Home economy.

The only change from the last section is in the learning technology. I assume that a worker with knowledge x, and working for a manager with knowledge y, draws his knowledge in the next period from a distribution which is uniform on $[x, y]$. The production function is given by $f(y) = y^{\alpha}$, $n(x; \beta) = x^{\beta}$. Finally, I assume that the distribution of newborns is a truncated exponential in $[1, \overline{k}]$ with parameter λ. By setting $\underline{k} = 1$, the size of the smallest firm is implicitly set to two (one manager and one worker). The following figures are drawn for $\alpha = 1$, $\beta = 0.5$, $\lambda = 1$, $\overline{k}_H = 1.5$, $\overline{k}_F = 1.75$.

5.1 Earnings and Welfare

Figure 1 compares the Home and Foreign economies under autarky. Figure 1a shows the welfare of an agent *as a function of his knowledge at birth*. The only difference between the two countries is in the distribution of newborns. In particular, the Foreign country has a larger knowledge support. This translates into relatively greater endowment of more knowledgeable agents in the Foreign country.[32]

The less knowledgeable agents are relatively scarce in the Foreign country and hence, are better off compared to their Home counterparts. But the most knowledgeable agents at Home are better off compared to their foreign counterparts. The relative abundance of less knowledgeable agents at Home translates into a Home wage schedule that lies below the Foreign wage schedule. Thus, labor is cheap at Home and this motivates the formation of MNEs as the two countries integrate. In the previous section, we had compared the steady-state under two regimes: autarky and integration. To compute the welfare gains from the integration policy, however, one must compare the autarky steady-state equilibrium with the integrated equilibrium in the period just after the policy is put in place. This requires solving the entire transitional path.

5.1.1 Transition

With the policy in place, there is no immediate change in the knowledge distribution. Consequently, the matches do not change. As soon as the policy is implemented, however, the agents' expectations about the future knowledge distributions change. Figure 2 shows the transition of the integrated knowledge distribution from the time the Home country integrates until the new steady-state is reached. The initial distribution is simply the sum of the Home and Foreign steady-state distributions. As shown in Figure 2, there are three discontinuities in the initial distribution. The first one occurs at k_A^*.[33] The second one

[32] If both Φ_H and Φ_F are truncated exponentials with the same parameter and $\overline{k}_H < \overline{k}_F$, then Φ_F first-order stochastic dominates Φ_H.

[33] There is always a discontinuity at the threshold. In a small interval to the left of the threshold (where all agents are workers) there is both an inflow of workers and an outflow of workers. But in a small interval just to the right of the threshold, there is only inflow and no outflow (since managers do not learn).

occurs at $\overline{k}_H$, because no Home agents are born to the right of $\overline{k}_H$. The third discontinuity occurs at the Foreign threshold. In the new steady-state, the discontinuities occur at k_I^* and $\overline{k}_H$.

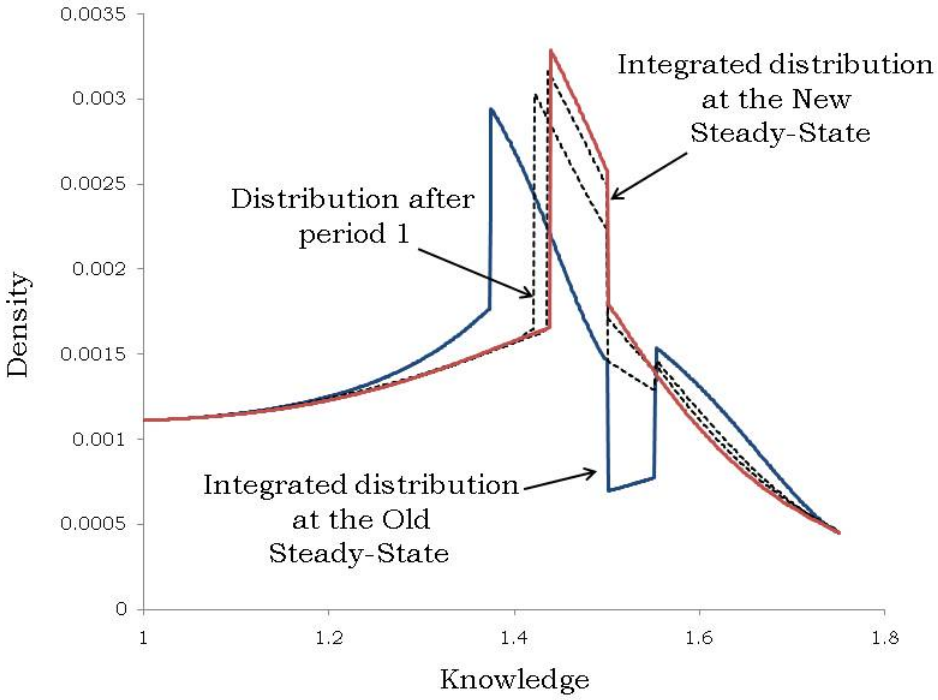
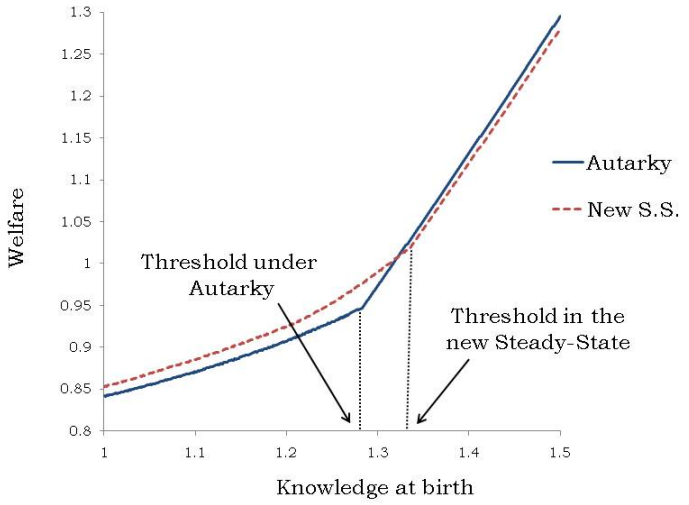


Figure 2: *Evolution of the knowledge distribution*

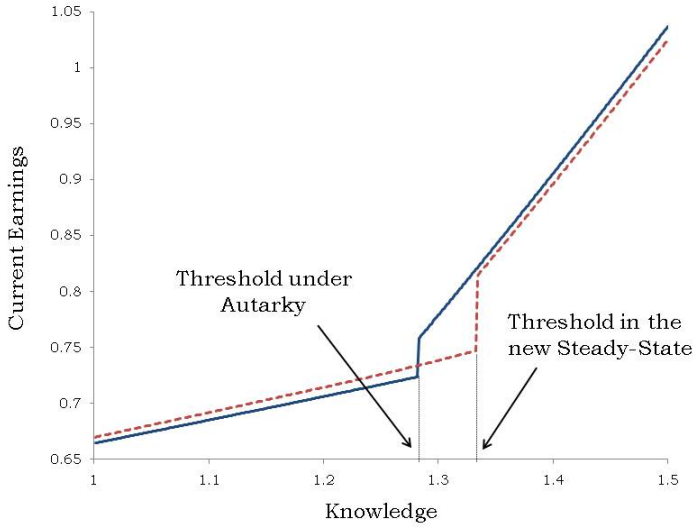
The agents have rational expectations and know exactly how the distribution will evolve. Accordingly, they know what the wages and profits will be at each period during the transition. This, in turn, allows them to compute their welfare in every period. Having solved the transition, I go on to study the integrated equilibrium for different learning rates. Recall from Section 4, that keeping the learning distribution unchanged, this implies choosing different values for δ.

5.1.2 Slow learning ($\delta = 0.8$)

In Figures 3 and 4, we compare the two steady-states: autarky and integration. In the steady-state under integration (New S.S.), the welfare of individuals who are born with less knowledge is higher, while the welfare of those born with high levels knowledge is lower, as shown in Figure 3a. Figure 4a indicates that the incumbent managers have a worse match; every incumbent manager produces less under integration. This is confirmed in Figure 4b. Therefore revenues are lower. But the effect on profits, which determines the managers' welfare, also depends on the wage bill.



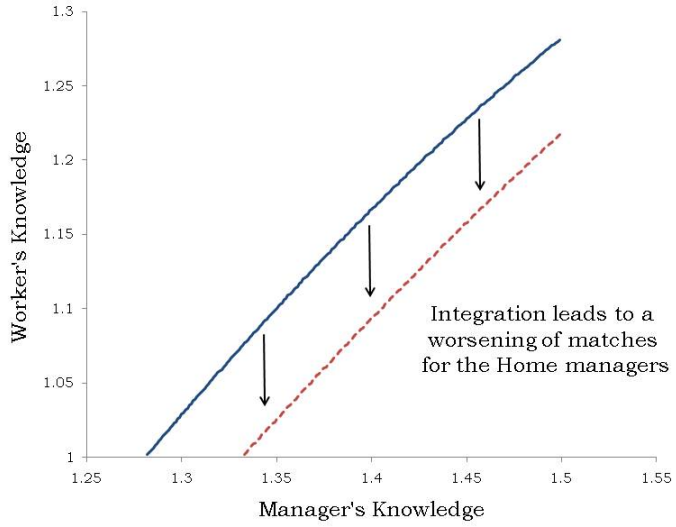
(a) *Welfare*



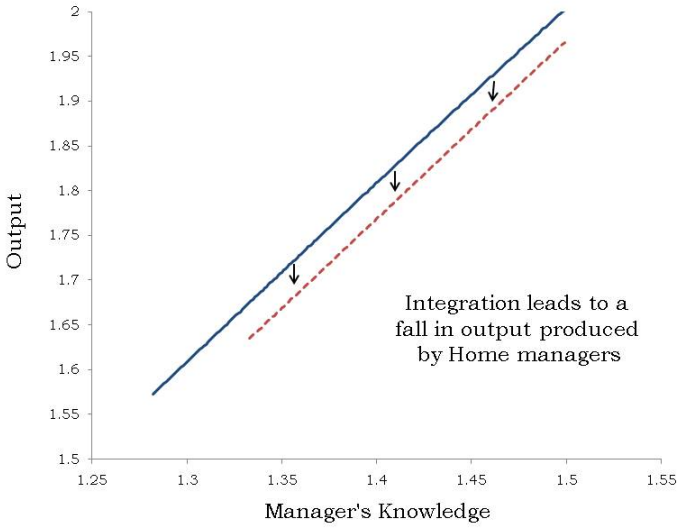
(b) *Current earnings*

Figure 3: *Effect of Integration on welfare and earnings when learning is slow*

The discussion in the previous section suggests that, the effect of integration on wages depends on the relative strength of the labor demand effect and the learning effect. When agents learn slowly, the former effect dominates and the wage schedule shifts up, thereby lowering the profits, and welfare, of incumbent



(a) *Inverse matching function*



(b) *Output*

Figure 4: *Effect of Integration on matches and output when learning is slow*

managers. This is shown in Figure 3b.[34] Notice that I restrict my attention to the agents who are born in $[\underline{k}, \overline{k}_H]$. Although, in the new steady state, there are Home agents with knowledge in $[\overline{k}_H, \overline{k}_F]$, but at the

[34]It can be shown that for $\delta = 0.8$, not only are there losers in the new steady-state, but the policy itself creates losers.

time of birth, these agents still draw their knowledge from the Home newborn distribution, which does not change with integration. Home agents attain knowledge in $[\overline{k}_H, \overline{k}_F]$ *through learning, not through birth*. Thus globalization creates a class of firms which are bigger and more productive than the "best" firms under autarky. [35]

5.1.3 Fast learning ($\delta = 0.5$)

Results are different when agents learn at a faster rate. This is displayed in Figure 5. Now, the learning effect dominates the labor demand effect, thereby lowering the wage schedule. Although the output of the incumbent Home managers fall due to a worsening of their matches, just like in the previous case, the wage bill decreases by so much, that it outweighs the fall in revenue, resulting in higher profits. This makes the home managers better-off. The less knowledgeable agents are better-off too, as the increase in their continuation value outweighs the decline in wages. This is true both in the new steady-state, as well as, in the period following the policy implementation.[36]

5.1.4 Discussion

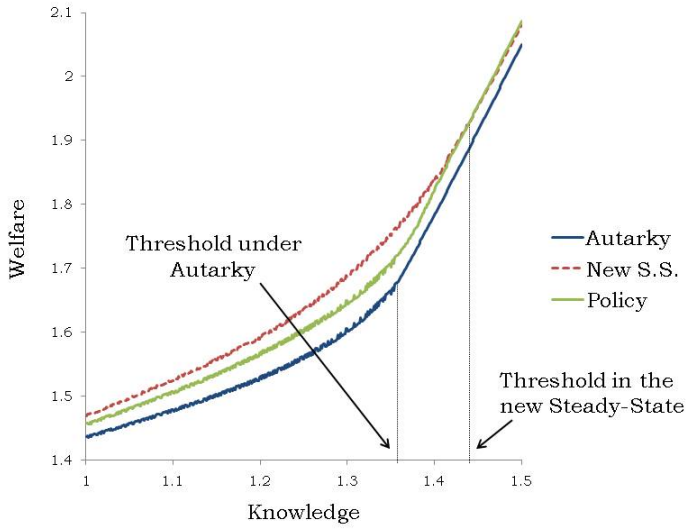
The above plots suggest that the incumbent firms experience a decline in output, irrespective of the rate of learning. According to Aitken and Harrison (1999), FDI was accompanied by a decline in the productivity of domestically owned firms in Venezuela. This decline, the authors report, is due to a contraction in output of domestic firms due to the "market stealing effects" of foreign firms.[37] In this model, the output reduction is a natural consequence of complementarity in production and learning. Under integration, the most knowledgeable workers are hired by the MNEs leaving less knowledgeable workers to work for incumbent domestic firms. I call this the "worker stealing effect". Despite this effect, the Home managers are actually better off if learning is fast enough.

Evidence regarding the impact of multinational production on wages has been mixed. Aitken et al. (1996) report that in Mexico and Venezuela, the wage spillover to domestic firms is negative and significant. On the other hand, Lipsey and Sjöholm (2004) find significant positive wage spillovers to domestic firms in Indonesia. In the previous section, we saw that if learning is fast enough, the wage schedule shifts down. This result, however, is not inconsistent with the finding of positive wage spillovers. In the above mentioned studies, the reported wage is the average of wages paid by all domestic firms. The average wage depends not only on the level of the wage schedule but also on the distribution of workers. With integration, as workers get matched with better managers and learn more, the mass of the distribution shifts to the right. This is confirmed in Figure 2. Therefore, a lowering of the wage schedule and a

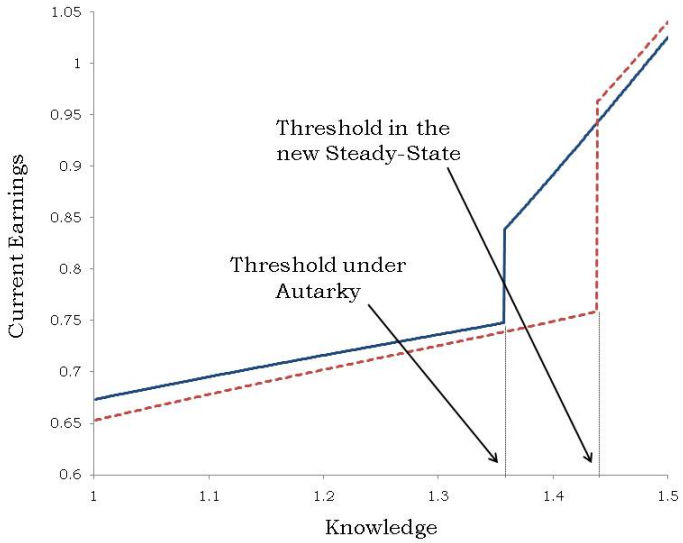
[35] The classic example is that of the Indian IT giant Infosys. In 1991, four employees working in the multinational Patni left to form Infosys, which then went on to become one of the largest IT firms in India, as well as, the world.

[36] It can be shown that the evolution of the value function during the transition is monotonic. This implies that for $\delta = 0.5$, agents are better-off compared to autarky at *each period* during the transition.

[37] With fixed costs of production, foreign firms with lower marginal costs can expand their output at the expense of domestic firms.



(a) *Welfare*



(b) *Inverse matching function*

Figure 5: *Effect of Integration when learning is fast*

higher average wage can go hand in hand.

The numerical results of this section confirm the analytical results obtained in the last section. When learning is slow, integration creates winners and losers. In the above example, the more knowledgeable

agents in the host country lose. But if agents learn fast enough, integration can make every agent better-off. In this case, there is a decline in the output of the incumbent firms, as well as, the wages of the workers. Therefore, a change in current wages or output could be misleading when it comes to assessing welfare gains from integration.

5.2 Earnings Dynamics

Although the economy as a whole does not grow in the steady-state, individual earning grows over the lifetime. Figure 6 plots the earnings path of the median worker for $\delta = 0.5$. The figure is drawn under the assumption that the actual knowledge he acquires every period is the expected knowledge that an agent with his level of knowledge would acquire. In the figure, the agent works for the first three periods and manages from the fourth period onwards.[38] Under integration, a lower wage in the first two periods is more than compensated by the increase in future profits. The lifetime earnings schedule under integration is also steeper than that under autarky.

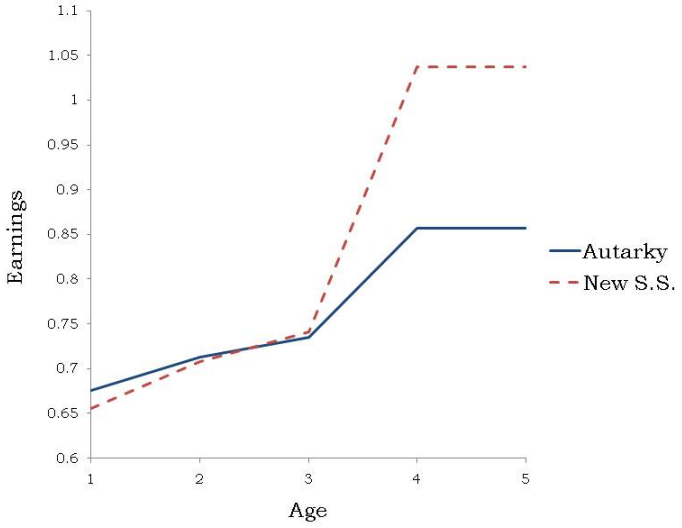


Figure 6: *Evolution of individual earnings*

Figure 7 provides an explanation for the greater jump in future profits. It shows the distribution of knowledge of the median worker after 5 periods.[39] Following integration, the distribution shifts to the right. On average, the agent becomes a more knowledgeable manager compared to autarky and hence, his expected profits are higher.

[38] Note that once the agent becomes a manager, his earnings do not change because he stops learning.

[39] Given the parameter values, the probability of the agent living for more than 5 periods is extremely small.

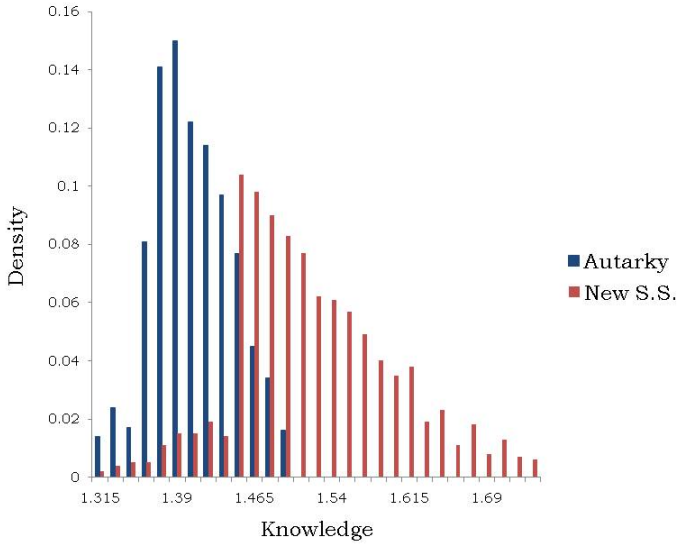


Figure 7: *Distribution of knowledge of an agent after 5 periods*

5.3 Inequality

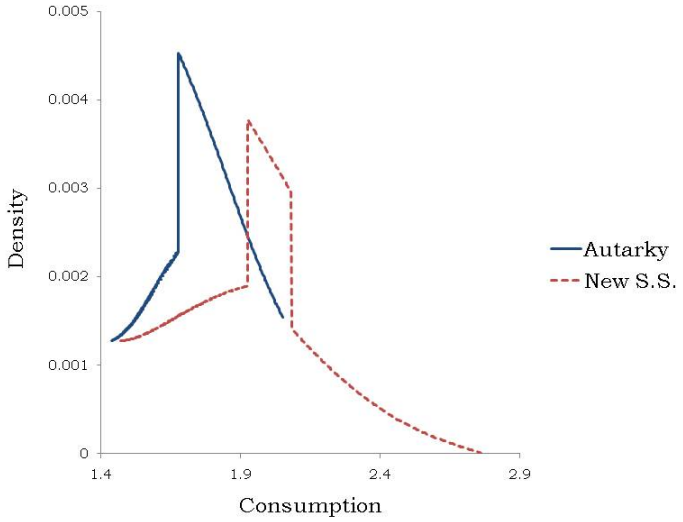


Figure 8: *Consumption distributions*

By generating a non-degenerate consumption distribution, the model also allows us to examine the

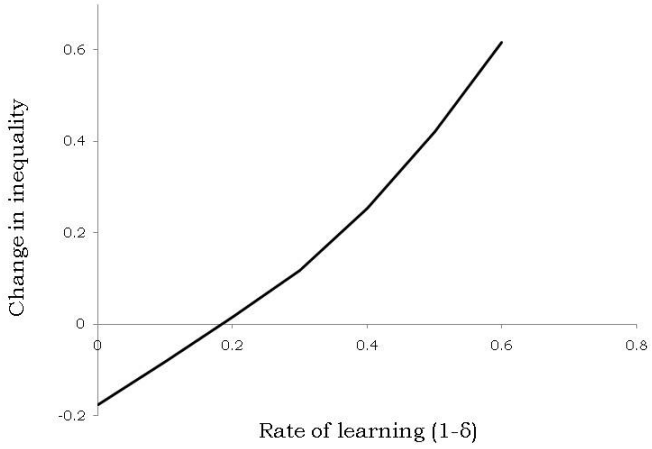


Figure 9: *Change in inequality*

effect of integration on inequality. Figure 8 plots the Home consumption distributions under autarky and integration. δ is set to 0.5 as before. With integration, the distribution stretches out, as mass shifts to the upper tail (The maximum consumption under autarky is 2.05, while that under integration is 2.77). Figure 9 plots the percentage change in the gini coefficient due to integration, as a function of δ. For $\delta = 0.5$, inequality rises by about 40%. But this rise in inequality is not a general phenomenon. For higher values of δ, integration actually leads to a reduction in inequality. Moreover, there is a monotonic relation between inequality and the rate of learning.

When agents learn fast enough, integration gives an advantage to the those who are born as the most knowledgeable workers. They work for the most knowledgeable Foreign managers, learn a lot, and in turn, become knowledgeable managers in the future. Agents who are born with very little knowledge continue to be matched with the less knowledgeable incumbent Home managers and accordingly, learn less - learning amplifies the initial inequality in the economy.

The above predictions could partly explain the experience of India since the early 90s. The annual FDI inflows to India increased from US $ 654 million to US $ 3083 million between the periods 1993-94 and 1999-00 (Ministry of Commerce and Industry, Government of India). During this time, there has been an increase in inequality, as measured by per-capita expenditure (Deaton and Dreze, 2002). This increase in inequality has occurred not only across regions, but within urban areas as well. If better managers hire more knowledgeable workers, and there is wide discrepancy in the knowledge levels of individuals,[40] then part of this rising inequality could be due to differences in opportunities faced by the

[40]In spite of churning out almost 400,000 engineers every year, only one in four of India's engineers are employable in the

Indian workers.

5.4 Pattern of MNE activity

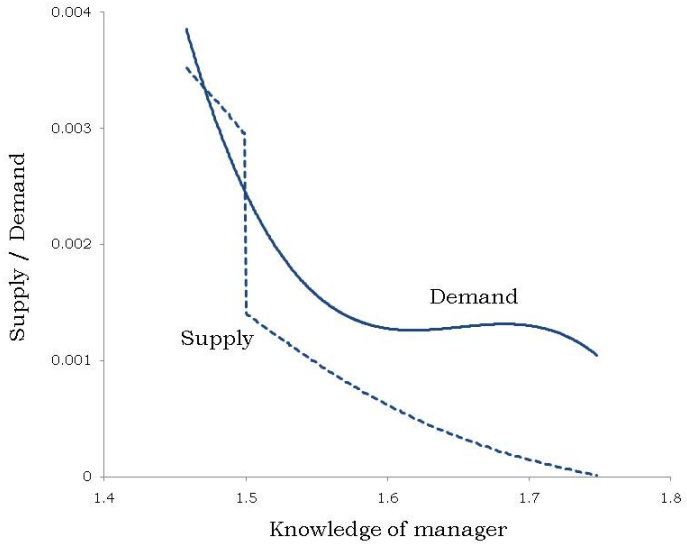


Figure 10: *Supply and demand for managers*

As mentioned in Section 4, integration leads to the creation of a new class of Home managers, who are as productive as their counterparts in the Foreign country. This can be seen in Figure 10 which plots the supply and demand for managers at Home in the integrated steady-state equilibrium. The supply of managers is simply the part of the knowledge distribution that lies above the threshold. Recall that the upper bound of the Home newborn distribution is 1.5. Hence, there is a discrete drop in the density of newborn agents to the right of 1.5, which explains the discontinuity at 1.5. The demand for managers is obtained by looking at the number of workers of each type and the demand for manager per worker.[41]

Figure 10 suggests that, in the integrated steady-state equilibrium, there are Home managers who are as knowledgeable as their Foreign counterparts. At the same time, the supply of Home managers is not sufficient to meet the demand. In the new equilibrium, some of the Foreign managers hire Home workers and hence, Home firms and Foreign MNEs operate together.[42] Figure 10 also throws light on the pattern of multinational activity. The supply of Home managers falls short of demand, and there are

software industry (*NYTimes, 17th October, 2006*).

[41]The demand for managers per worker is simply the reciprocal of the span of control. This is a special feature of the span of control depending only on the knowledge of the worker.

[42]See Markusen and Venables (1999) for the case where FDI leads to the development of local industry, but is driven out as the industry develops enough.

almost no Home MNEs in this equilibrium.[43] Moreover, most of the MNEs operating at Home are the best Foreign firms. Thus, the MNEs, on average, are bigger and more productive than the Home firms. PAM implies that a worker in a MNE, on average, is more knowledgeable than a worker in a Home firm. Therefore, the former employees of MNEs are also, on average, more productive managers.

In a survey of firms in Ghana, Görg and Strobl (2005) investigate whether knowledge spillovers occur through worker mobility. They combine information on whether or not the owner of a domestic firm had previous experience in a multinational with information on firm-level productivity. They show that firms which are run by owners who worked for foreign multinationals in the same industry immediately prior to opening their firm, are more productive than other domestic firms. Using data on Danish firms, Malchow-Mø ller et al. (2007) show that previous experience in foreign-owned firms increases a worker's current wage. Both pieces of evidence are consistent with my model.

6 Conclusion

Despite both formal and anecdotal evidence on knowledge spillover from foreign multinationals, we do not have a good understanding of, for example, how the distribution of knowledge changes and occupational choice evolves as a country gradually integrates with the rest of the world. I believe that this paper is a step in that direction. I show that allowing domestic workers to learn from foreign managers not only generates novel predictions about lifetime earnings profiles and consumption inequality among other things, ignoring the dynamic nature of knowledge diffusion could lead us to draw incorrect conclusions about welfare. Much remains to be done, however.

As pointed out in the introduction, in order to focus on matching and occupational choice, I assume that the transfer of knowledge is a costless process. In reality, firms spend a substantial part of their resources on training their workforce. If there is some degree of mobility of workers, firms' incentive to train their workers might be affected, which in turn would affect the diffusion of knowledge. How would the entry of multinationals affect domestic workers and firms in this setting?

This paper highlights a new channel through which domestic managers can gain indirectly from foreign multinationals, even in the absence of any flow of knowledge between them. But mobility of workers from more productive multinationals to less productive domestic firms could be another channel through which domestic managers gain. Workers with experience in multinationals could move to domestic firms for various reasons. A domestic firm could lure a multinational trained worker with higher wages. Or, in a frictional labor market, a multinational trained worker could accept a job in a small domestic firm, rather than wait to be matched with a more productive firm. Whatever be the case, domestic firms benefit from the superior knowledge that these workers bring along with them. I leave the examination of such interesting scenarios for future work.

[43] Here we follow AGR in assuming that, a manager will hire workers in the other country only if he strictly prefers doing so.

References

Aitken, B., Harrison, A. and Lipsey, R. E. (1996), 'Wages and foreign ownership: A comparative study of Mexico, Venezuela, and the United States', *Journal of International Economics* **40**(3-4), 345 – 371.

Aitken, B. J. and Harrison, A. E. (1999), 'Do Domestic Firms Benefit from Direct Foreign Investment? Evidence from Venezuela', *American Economic Review* **89**(3), 605–618.

Antràs, P. (2003), 'Firms, Contracts and Trade Structure', *Quarterly Journal of Economics* **118**(4), 1375–1418.

Antràs, P., Garicano, L. and Rossi-Hansberg, E. (2006), 'Offshoring in a Knowledge Economy', *Quarterly Journal of Economics* **121**(1).

Antràs, P. and Helpman, E. (2004), 'Global Sourcing', *Journal of Political Economy* **112**(3), 552–580.

Barba Navaretti, G. and Venables, A. (2004), *Multinational Firms in the World Economy*, Princeton University Press.

Beaudry, P. and Francois, P. (2010), 'Managerial Skills Acquisition and the Theory of Economic Development', *Review of Economic Studies* **77**(1), 90–126.

Becker, G. S. (1962), 'Investment in Human Capital: A Theoretical Analysis', *Journal of Political Economy* **70**(s5), 9.

Becker, G. S. (1973), 'A theory of marriage: Part i', *The Journal of Political Economy* **81**(4), pp. 813–846.

Bloom, N. and Van Reenen, J. (2010), 'Why do management practices differ across firms and countries?', *Journal of Economic Perspectives* **24**(1), 203–24.

Burstein, A. T. and Monge-Naranjo, A. (2009), 'Foreign Know-How, Firm Control, and the Income of Developing Countries', *Quarterly Journal of Economics* **124**(1), 149–195.

Carmichael, L. (1983), 'Firm-specific human capital and promotion ladders', *The Bell Journal of Economics* **14**(1), pp. 251–258.

Chari, V. V. and Hopenhayn, H. (1991), 'Vintage Human Capital, Growth, and the Diffusion of New Technology', *Journal of Political Economy* **99**(6), 1142–65.

Deaton, A. and Dreze, J. (2002), 'Poverty and inequality in india: A re-examination', *Economic and Political Weekly* **37**(36), pp. 3729–3748.

Easterly, W. (2001), *The Elusive Quest for Growth: Economists' Adventures and Misadventures in the Tropics*, The MIT Press.

Filson, D. and Franco, A. (2006), 'Spin-outs: Knowledge diffusion through employee mobility', *Rand Journal of Economics* **37**(4), 841–60.

Garetto, S. (2008), Input Sourcing and Multinational Production, mimeo, University of Chicago.

Garicano, L. (2000), 'Hierarchies and the Organization of Knowledge in Production', *Journal of Political Economy* **108**(5), 874–904.

Giarratana, M., Pagano, A. and Torrisi, S. (2004), The Role of Multinational Firms in the Evolution of the Software Industry in India, Ireland and Israel, *in* A. Arora and A. Gambardella, eds, 'From Underdogs to Tigers: Bridging the Gap', Oxford University Press, New York, pp. 207–235.

Goldberg, P. K. and Pavcnik, N. (March 2007), 'Distributional Effects of Globalization in Developing Countries', *Journal of Economic Literature* **45**, 39–82(44).

Görg, H. and Strobl, E. (2005), 'Spillovers from foreign firms through worker mobility: An empirical investigation', *Scandinavian Journal of Economics* **107**(4).

Grossman, G. M. and Helpman, E. (2003), 'Outsourcing versus FDI in Industry Equilibrium', *Journal of the European Economic Association* **1**(2/3), 317–327.

Hashimoto, M. (1981), 'Firm-specific human capital as a shared investment', *The American Economic Review* **71**(3), pp. 475–482.

Helpman, E. (1984), 'A Simple Theory of International Trade with Multinational Corporations', *Journal of Political Economy* **92**(3), 451–471.

Holmes, T. J. and Schmitz, James A., J. (1990), 'A Theory of Entrepreneurship and its Application to the Study of Business Transfers', *Journal of Political Economy* **98**(2), 265–294.

IMF (2007), *World Economic Outlook, October 2007: Globalization and Inequality*, International Monetary Fund.

Jovanovic, B. and Rob, R. (1989), 'The Growth and Diffusion of Knowledge', *Review of Economic Studies* **56**(4), 569–582.

Klepper, S. (2002), 'The capabilities of new firms and the evolution of the US automobile industry', *Industrial and Corporate Change* **11**(4), 645–666.

Klepper, S. and Sleeper, S. (2005), 'Entry by spinoffs', *Management Science* **51**(8), 1291–1306.

Kremer, M. (1993), 'The o-ring theory of economic development', *The Quarterly Journal of Economics* **108**(3), 551–575.

Kremer, M. and Maskin, E. (2006), Globalization and Inequality, mimeo, Harvard University.

Lipsey, R. E. and Sjöholm, F. (2004), 'Fdi and Wage Spillovers in Indonesian Manufacturing', *Review of World Economics* **140**(2), 287–310.

Malchow-Mø ller, N., Markusen, J. R. and Schjerning, B. (2007), Foreign Firms, Domestic Wages, Working paper, NBER.

Markusen, J. R. and Venables, A. J. (1999), 'Foreign direct investment as a catalyst for industrial development', *European Economic Review* **43**(2), 335 – 356.

Monge-Naranjo, A. (2007), Foreign Firms, Domestic Entrepreneurial Skills and Development, mimeo, Northwestern University.

Ohnsorge, F. and Trefler, D. (2007), 'Sorting It Out: International Trade with Heterogeneous Workers', *Journal of Political Economy* **115**(5), 868–892.

Oman, C. (2000), *Policy competition for foreign direct investment: A study of competition among governments to attract FDI*, OECD.

Poole, J. (2006), Multinational Spillovers through Worker Turnover, mimeo, University of San Diego.

Ramondo, N. (2008), Size, Geography and Multinational Production, mimeo, University of Texas at Austin.

Rodríguez-Clare, A. (2007), Trade, Diffusion and the Gains from Openness, Working paper, NBER.

Rosen, S. (1972), 'Learning and Experience in the Labor Market', *Journal of Human Resources* **7**(3), 326–342.

Rosen, S. (1982), 'Authority, Control, and the Distribution of Earnings', *Bell Journal of Economics* **13**(2), 311–323.

Sattinger, M. (1993), 'Assignment Models of the Distribution of Earnings', *Journal of Economic Literature* **31**(2), 831–880.

Appendix

Micro-foundation for the learning technology : Suppose workers have to exert effort in order to learn. If a worker, working for a manager with knowledge y, exerts e, his knowledge next period is $l(e,y)$. For simplicity, I assume that $l(e,y)$ has the following functional form - $\xi(y)e$, with $\xi' > 0$. So all workers, working for a particular manager, can potentially learn the same. Effort, however, is costly. Let us denote this cost by $q_x(e)$ with $q'_x > 0$. Note that the cost of effort depends on the knowledge of the worker. I assume that $q'_x(e)$ is decreasing in x. Let the continuation value of a worker with knowledge $l(e,y)$ in the next period be $C(l(e,y))$. If the Separation Theorem holds, the worker will choose e to maximize $V(l(e,y)) - q_x(e)$. The first-order condition for utility maximization yields $\frac{dC}{dl}\frac{dl}{de} = \frac{dC}{dl}\xi(y) = \frac{dq_x}{de}$. The left-hand side of this equation is constant. Given the assumption about q_x, it is then straightforward to show that more knowledgeable workers will choose higher e and accordingly have higher knowledge next period.

Proof of Lemma 1: Since y is indifferent along $\widetilde{w}_t(x,y)$, for any x_1 and x_2 we must have $f(y)n(x_1) - \widetilde{w}_t(x_1,y)n(x_1) = f(y)n(x_2) - \widetilde{w}_t(x_2,y)n(x_2)$. Letting $x_2 = x_1 + h$ and re-arranging, we have $\widetilde{w}_t(x_1 + h, y)n(x_1 + h) - \widetilde{w}_t(x_1,)n(x_1) = f(y)n(x_1 + h) - f(y)n(x_1)$. Using Taylor series approximation of $n(x_1 + h)$ around h (small), we have $n(x_1 + h) = n(x_1) + n'(x_1)h + o_2$. Replacing this in the above equation, we have $[\widetilde{w}_t(x_1 + h, y) - \widetilde{w}_t(x_1, y)]n(x_1) + \widetilde{w}_t(x_1 + h, y)n'(x_1)h = f(y)n(x_1 + h) - f(y)n(x_1)$. Dividing by h and taking the limit as $h \to 0$, we get $\frac{\partial \widetilde{w}_t(x_1,y)}{\partial x}n(x_1) + \widetilde{w}_t(x_1,y)n'(x_1) = f(y)n'(x_1)$. Re-arranging, we have $\frac{\partial \widetilde{w}_t(x_1,y)}{\partial x} = \frac{n'(x_1)(f(y)-\widetilde{w}_t(x_1,y))}{n(x_1)}$. Since $f(y) - \widetilde{w}_t(x_1,y) > 0$ (in equilibrium, profits must be positive) and $n'(x_1) > 0$, it follows that $\frac{\partial \widetilde{w}_t(x_1,y)}{\partial x} > 0$. Since x_1 was chosen randomly, the result follows. Furthermore, in equilibrium, $\frac{\partial \widetilde{w}_t(x,y)}{\partial x} = \frac{n'(x)(f(y)-\widetilde{w}_t(x))}{n(x)}$. Differentiating with respect to y, $\frac{\partial^2 \widetilde{w}_t(x,y)}{\partial x \partial y} = \frac{n'(x)f'(y)}{n(x)} > 0$.

Proof of Lemma 2: The worker's optimization problem is given by

$$V_W(x) = \max_y [\widetilde{w}_t(x,y) - \widetilde{r}_t(x,y) + \int V_W(x')dL(x'|x,y)].$$

The first-order condition is given by $\frac{\partial \widetilde{r}_t(x,y)}{\partial y} = \frac{\partial \widetilde{w}_t(x,y)}{\partial y} + \beta \frac{\partial}{\partial y}\int V_W(x')dL(x'|x,y)$. Differentiating with respect to x, $\frac{\partial^2 \widetilde{r}_t(x,y)}{\partial x \partial y} = \frac{\partial^2 \widetilde{w}_t(x,y)}{\partial x \partial y} + \beta \frac{\partial^2}{\partial x \partial y}\int V_W(x')dL(x'|x,y)$. The first term on the left-hand side is positive (Lemma 1). The second term is also positive because of Assumption 2c and $V_w(x')$ being increasing in x'.

Proof of Lemma 3: I shall prove this lemma by contradiction. I drop the subscript t for simplicity. If we do not have PAM in equilibrium, the there must be $x_1 < x_2$ and $y_1 < y_2$ such that $m(x_1) = y_2$ and $m(x_2) = y_1$. Then,

$$\begin{aligned}\pi(y_2) &= [f(y_2) - \widetilde{w}(x_1,y_2) + \widetilde{r}(x_1,y_2)]n(x_1) \\ &\geq [f(y_2) - \widetilde{w}(x_2,y_2) + \widetilde{r}(x_2,y_2)]n(x_2)\end{aligned}$$

Similarly, we must have

$$\begin{aligned}\pi(y_1) &= [f(y_1) - \widetilde{w}(x_2,y_1) + \widetilde{r}(x_2,y_1)]n(x_2) \\ &\geq [f(y_1) - \widetilde{w}(x_1,y_1) + \widetilde{r}(x_1,y_1)]n(x_1)\end{aligned}$$

Combining the above two inequalities and using the fact that $n(x_1) < n(x_2)$, we can write

$$\begin{aligned}&[\widetilde{w}(x_2,y_2) - \widetilde{w}(x_1,y_2)] - [\widetilde{w}(x_2,y_1) - \widetilde{w}(x_1,y_1)] \\ &> [\widetilde{r}(x_2,y_2) - \widetilde{r}(x_1,y_1)] - [\widetilde{w}(x_1,y_2) - \widetilde{w}(x_1,y_1)]\end{aligned}$$

Defining $x_2 = x_1 + \Delta$ and $y_2 = y_1 + \Delta$ and taking the limit as $\Delta \to 0$, we have $\frac{\partial^2 \widetilde{w}(w,y)}{\partial x \partial y} > \frac{\partial^2 \widetilde{r}(w,y)}{\partial x \partial y}$. But from Lemma 2, we know that $\frac{\partial^2 \widetilde{w}(w,y)}{\partial x \partial y} < \frac{\partial^2 \widetilde{r}(w,y)}{\partial x \partial y}$. Hence, we get a contradiction.

Proof of Lemma 4: Equilibrium in the labour market implies that

$$\int_{\underline{k}}^{k_t^*} \psi(s)ds = \int_{k_t^*}^{\overline{k}} n(m_t^{-1}(s))\psi(s)ds$$

where the LHS is the supply of workers while the RHS is the demand for workers. Define

$$\mathcal{L}(k_t^*) = \int_{\underline{k}}^{k_t^*} \psi(s)ds - \int_{k_t^*}^{\overline{k}} n(m_t^{-1}(s))\psi(s)ds$$

Now, $\mathcal{L}(\underline{k}) = -\int_{\underline{k}}^{\overline{k}} n(m_t^{-1}(s))\psi(s)ds < 0$, while $\mathcal{L}(\overline{k}) = \int_{\underline{k}}^{\overline{k}} \psi(s)ds > 0$. Moreover, $\frac{\partial L(k_t^*)}{\partial k_t^*} = [1 + n(m_t^{-1}(k_t^*)]\psi(k_t^*) > 0$. Hence, by the Intermediate Value Theorem, $\exists$ a unique k_t^* such that $\mathcal{L}(k_t^*) = 0$.

Proof of Lemma 5: The value function for the manager is given by

$$V_M(k, w_t) = \max_x\{f(k)n(x) - w_t(x)n(x)\} + (1-\delta)\max[V_W(k, w_{t+1}), V_M(k, w_{t+1})]$$

The value function for the worker is given by

$$V_W(k, w_t) = w_t(k) + (1-\delta)\int_k^{m_t(k)} \max[V_W(k', w_{t+1}), V_M(k', w_{t+1})]dL(k'|k, m_t(k))$$

Define the vector function $V = [V_M(k, w_t)\ V_W(k, w_t)]'$. Then $\max\{V_W, V_M\} = \max\{[1\ 0]V, [0\ 1]V\}$. Also, define $\alpha = [\max_x\{f(k)n(x) - w_t(x)n(x)\}\ w_t(k)]'$. Then we have the following equation:

$$\begin{aligned} V &= \alpha + (1-\delta)\begin{bmatrix} \max\{[1\ 0]V, [0\ 1]V\} \\ \int_k^{m_t(k)} \max\{[1\ 0]V, [0\ 1]V\}dL \end{bmatrix} \\ &= T(V) \end{aligned}$$

It can be established, using Blackwell's Sufficiency Conditions, that the operator T is a contraction in the space of continuous vector functions with norm $max[\sup_k |V_M(k)|, \sup_k |V_W(k)|]$. Therefore, a fixed point of V exists and is unique.

To prove that $V_W(k)$ is increasing in k, note that if $V_W'(k) < 0$, a worker will choose not to learn because learning reduces his continuation value. If workers do not learn, they do not pay rent. Consequently, the only payment that is made is wage and we are back to the static framework. But then more knowledgeable agents earn more and $V_W'(k) > 0$ trivially. Thus, we get a contradiction. It can be proved in a similar fashion that $V_M'(k) > 0$.

Proof of Lemma 6: First, let us derive the equilibrium conditions for a threshold equilibrium. Since k_t^* is indifferent between being a worker and a manager, we must have $V_W(k_t^*, w_t) = V_M(k_t^*, w_t)$. Furthermore, for k_t^* to be the threshold, it must be the case that $\overline{k}$ cannot hire $k_t^* + \epsilon$ and be strictly better-off. If $k_t^* + \epsilon$ is a manager, he earns $V_M(k_t^* + \epsilon)$. In order to hire $k_t^* + \epsilon$, the manager has to pay him a wage such that he is just indifferent between being a manager and a worker. Let this wage be ω. ω should satisfy

$$\omega + (1-\delta)\int V_M(k)dL(k|, k_t^* + \epsilon, \overline{k}) = V_M(k_t^* + \epsilon)$$

Therefore, period profit of $\overline{k}$ if he hires $k_t^* + \epsilon$ is given by

$$\begin{aligned}\pi_{k_t^*+\epsilon}(\overline{k}) &= (f(\overline{k}) - \omega)n(k_t^* + \epsilon) \\ &= f(\overline{k})n(k_t^* + \epsilon) - n(k_t^* + \epsilon)(V_M(k_t^* + \epsilon) - (1-\delta)\int V_M(k)dL(k|, k_t^* + \epsilon, \overline{k}))\end{aligned}$$

For k_t^* to be a threshold equilibrium, it must be the case that $\underset{\epsilon\to 0}{lim}\frac{\partial \pi_{k_t^*+\epsilon}(\overline{k})}{\partial \epsilon} \leq 0$ Now,

$$\begin{aligned}\lim_{\epsilon\to 0}\frac{\partial \pi_{k_t^*+\epsilon}(\overline{k})}{\partial \epsilon} &= f(\overline{k})n'(k_t^*) - n(k_t^*)(V_M'(k_t^*) - (1-\delta)\frac{\partial}{\partial k_t^*}\int V_M(k)dL(k|, k_t^*, \overline{k})) \\ &\quad - n'(k_t^*)V_M(k_t^*) - (1-\delta)\int V_M(k)dL(k|, k^*, \overline{k})\end{aligned}$$

From the manager's profit-maximizing problem, we have

$$f(\overline{k})n'(k_t^*) = w_t'(k_t^*)n(k_t^*) + w_t(k_t^*)n'(k_t^*)$$

Also, for a worker with knowledge k_t^*,

$$V_W(k_t^*) = w_t(k_t^*) + (1-\delta)\int V_M(k)dL(k|k_t^*, \overline{k})$$

Since k_t^* is the threshold, $\max[V_W(k), V_M(k)] = V_M(k)\ \forall k \geq k_t^*$. Differentiating w.r.t. k_t^*,

$$V_W'(k_t^*) = w_t'(k_t^*) + (1-\delta)\frac{\partial}{\partial k_t^*}\int V_M(k)dL(k|k_t^*, \overline{k})$$

Replacing in the expression for $\lim_{\epsilon\to 0}\frac{\partial \pi_{k_t^*+\epsilon}(\overline{k})}{\partial \epsilon}$ and using the fact that $V_W(k_t^*) = V_M(k_t^*)$, we have

$$\begin{aligned}\lim_{\epsilon\to 0}\frac{\partial \pi_{k_t^*+\epsilon}(\overline{k})}{\partial \epsilon} &= [w_t'(k_t^*) - V_M'(k_t^*) - V_W'(k_t^*) + w_t'(k_t^*)]n(k_t^*) \\ &= V_W'(k_t^*) - V_M'(k_t^*)\end{aligned}$$

where we use the fact that $V_W(k_t*, w_t) = V_M(k_t^*, w_t)$. Hence, $\underset{\epsilon\to 0}{lim}\frac{\partial \pi_{k_t^*+\epsilon}(\overline{k})}{\partial \epsilon} < 0$ implies that

$$V_W'(k_t^*) < V_M'(k_t^*)$$

The above condition needs to be satisfied for k_t^* to be the equilibrium threshold. We shall prove this proposition in a slightly different way. First, we shall prove the existence of the threshold equilibrium, assuming that the equilibrium is unique. Then we shall show that the sufficient condition for existence is also sufficient for uniqueness.

By assuming uniqueness, we are basically assuming that the set of workers and managers has to be connected in equilibrium (See AGR). Given that there exists a unique market-clearing threshold k_t^*, we check whether the threshold satisfies the equilibrium condition $V_W'(k_t^*) < V_M'(k_t^*)$. Dropping the time subscript, we have

$$V_M(k^*) = \frac{1}{\delta}(f(k^*) - w(\underline{k}))n(\underline{k})$$

Using the Envelope Theorem,

$$V_M'(k^*) = \frac{1}{\delta}f'(k^*)n(\underline{k})$$

Also,

$$\begin{aligned} V'_W(k^*) &= w'(k^*) + (1-\delta)\frac{\partial}{\partial k^*}(\int V_M(k)dL(k|k^*,\overline{k})) \\ &= \frac{(f(\overline{k}) - w(k^*))n'(k^*)}{n(k^*)} + (1-\delta)\frac{\partial}{\partial k^*}(\int V_M(k)dL(k|k^*,\overline{k})) \end{aligned}$$

where the second line follows from the manager's profit-maximization condition. Therefore, for k^* to be an equilibrium, it must be the case that

$$\frac{(f(\overline{k}) - w(k^*))n'(k^*)}{n(k^*)} + (1-\delta)\frac{\partial}{\partial k^*}(\int V_M(k)dL(k|k^*,\overline{k})) \leq \frac{1}{\delta}f'(k^*)n(\underline{k})$$

If $\delta = 1$, this condition reduces to

$$\frac{(f(\overline{k}) - w(k^*))n'(k^*)}{n(k^*)} \leq f'(k^*)n(\underline{k})$$

Since $w(k^*) > 0$, for the above inequality to hold, we need to find the conditions under which $\frac{f(\overline{k})n'(k^*)}{n(k^*)} \leq f'(k^*)n(\underline{k})$, or $f(\overline{k})n'(k^*) \leq f'(k^*)n(\underline{k})$, since $n(k^*) \geq 1$.

But $f(\overline{k})n'(k^*) \leq f(\overline{k})n'(\underline{k})$ ($\because n''(.) \leq 0$) and $f'(k^*)n(\underline{k}) \geq f'(\underline{k})n(\underline{k})$ (because $n''(.) \geq 0$). Hence, it follows that

$$f(\overline{k})n'(k^*) \leq f(\overline{k})n'(\underline{k}) \leq f'(\underline{k})n(\underline{k}) \leq f'(k^*)n(\underline{k})$$

where the inequality in the middle follows from Assumption 3. Thus for $\delta = 1$, the condition on technology is sufficient for an equilibrium. But when $\delta \neq 1$, we need to determine the magnitude of $\frac{\partial}{\partial k^*}(\int V_M(k)dL(k|k^*,\overline{k}))$, since this term is positive by assumption on the learning technology. This term is endogenous and it depends on the invariant distribution, which in turn is determined by the learning distribution. This term is bounded above, since the domain is compact. Hence by the Least Upper Bound Property, the supremum exists. Let

$$\zeta = \sup\{\frac{\partial}{\partial k^*}(\int V_M(k)dL(k|k^*,\overline{k}))\}$$

Define δ^* as the value of δ that satisfies

$$f(\overline{k})n'(\underline{k}) + (1-\delta^*)\zeta = f'(\underline{k})n(\underline{k})$$

This can be re-written as

$$\frac{n'(\underline{k})}{n(\underline{k})} + (1-\delta^*)\frac{\zeta}{f(\overline{k})n(\underline{k})} = \frac{f'(\underline{k})}{f(\overline{k})}$$

The fact that $\frac{n'(\underline{k})}{n(\underline{k})} < \frac{f'(\underline{k})}{f(\overline{k})}$ implies that $\delta^* < 1$. Hence $\forall \delta \in [\delta^*, 1]$, we have

$$f(\overline{k})n'(\underline{k}) + (1-\delta)\zeta \leq f'(\underline{k})n(\underline{k})$$

Thus,

$$\begin{aligned} f(\overline{k})n'(k^*) + (1-\delta)\frac{\partial}{\partial k^*}(\int V_M(k)dL(k|k^*,\overline{k})) &\leq f(\overline{k})n'(\underline{k}) + (1-\delta)\zeta \\ &\leq f'(\underline{k})n(\underline{k}) \\ &\leq f'(k^*)n(\underline{k}) \end{aligned}$$

$\delta \leq 1$ implies that $f'(k^*)n(\underline{k}) \leq \frac{1}{\delta} f'(k^*)n(\underline{k})$. Therefore,

$$f(\overline{k})n'(k^*) + (1-\delta)\frac{\partial}{\partial k^*}(\int V_M(k)dL(k|k^*,\overline{k})) \leq \frac{1}{\delta} f'(k^*)n(\underline{k})$$

This completes our proof about the existence of equilibrium. As mentioned before, showing uniqueness entails showing that the set of workers and managers is connected. Suppose not. WLOG let us assume that the knowledge distribution has the following partition - $([\underline{k}, k_1], [k_1, k_2], [k_2, k_3], [k_3, k_4])$. Workers in $[\underline{k}, k_1]$ work for managers in $[k_1, k_2]$ while workers in $[k_2, k_3]$ work for managers in $[k_3, k_4]$. For this to be an equilibrium, it must be the case that k_2 must be indifferent between being a worker and a manager. In other words, a deviation involving k_3 hiring $k_2 - \epsilon$ should not make both k_3 and $k_2 - \epsilon$ better off. Using a similar logic as developed above, one can show that that the condition for equilibrium is $V'_W(k_2) > V'_M(k_2)$. One can then show that if $\frac{n'(\underline{k})}{n(\underline{k})} < \frac{f'(\underline{k})}{f(\overline{k})}$, then for δ high enough, this condition will always be violated. Therefore, an allocation with disconnected sets of workers and managers can never be sustained as an equilibrium implying that the only equilibrium is the threshold equilibrium.

Proof of Lemma 7: There is an alternative way of looking at the evolution of knowledge. Let A be any Borel set of $[\underline{k}, \overline{k}]$. Then the transition function for the knowledge distribution satisfies, for every $k \in [\underline{k}, \overline{k}]$,

$$P_t(k, A) = \begin{cases} (1-\delta)\int\limits_A dL(s|k, m_t(k)) + \delta \int\limits_A d\Phi(s) & \text{if } k \in \overline{W} \\ \delta \int\limits_A d\Phi(s) & \text{if } k \in \overline{M} \end{cases}$$

Suppose P is monotone, has the Feller property and satisfies a mixing condition. Then P has a unique, invariant probability measure Ψ^* (Stokey, Lucas with Prescott, 1989). Define the operator T as

$$(Tf)(k) = \int f(k')P(k, dk'), \quad \text{all } k \in [\underline{k}, \overline{k}]$$

where $f : [\underline{k}, \overline{k}] \to \mathbb{R}$ is a bounded function. If f is non-decreasing, then the first-order stochastic dominance property of the learning distribution implies that Tf is also non-decreasing. (Monotone Property) It is straightforward to verify that if f is bounded and continuous, then the same holds for Tf, i.e., $T : C(k) \to C(k)$ (Feller Property). The mixing condition requires that $\exists c \in [\underline{k}, \overline{k}]$, $\epsilon > 0$ and $N \geq 1$ such that $P^N(\underline{k}, [c, \overline{k}]) \geq \epsilon$ and $P^N([\underline{k}, c], \overline{k}]) \geq \epsilon$. Choose $k' \in [\underline{k}, \overline{k}]$. Define $\epsilon_1 = \int_{[k', \overline{k}]} d\Psi_N(s)$ and $\epsilon_2 = \int_{[\underline{k}, k']} d\Psi_N(s)$. By the assumption on $\Psi_N(.)$, we know that both these objects are greater than 0. Choose $\epsilon = \delta \min\{\epsilon_1, \epsilon_2\}$ and $N = 1$. Then $P(\underline{k}, [k', \overline{k}]) \geq \epsilon$ and $P([\underline{k}, k'], \overline{k}]) \geq \epsilon$. Therefore all the conditions for the existence and uniqueness of the invariant distribution are satisfied.

Proof of Proposition 1: Let the number of people being born every period be normalized to 1. Cohort t at time t are all newborns. All agents in $[\underline{k}, k_A^*]$ are workers. The measure of these agents is $\int_{\underline{k}}^{k_A^*} \phi_H(k)dk$. A worker with knowledge k demands $\frac{1}{n(k)}$ managers. Therefore, the total demand for managers by cohort t workers is $\int_{\underline{k}}^{k_A^*} \frac{\phi_H(k)}{n(k;\beta)}dk$. The supply of cohort t managers is simply the measure of agents in $[k_A^*, \overline{k}_H]$. This is given by $\int_{k_A^*}^{\overline{k}_H} \phi_H(k)dk$. Let us consider the distribution of cohort $t-1$ agents at time t. A fraction $1-\delta$ of every type of agent in $[\underline{k}, k_A^*]$ survive in period t. Out of the ones that survive, a fraction π of every type of agent do not learn and remain where they are. Hence, the total demand for managers by cohort $t-1$ workers is $\int_{\underline{k}}^{k_A^*} \frac{\pi(1-\delta)\phi_H(k)}{n(k;\beta)}dk$. Similarly, a fraction $1-\delta$ of the cohort $t-1$ managers in $[k_A^*, \overline{k}]$ survive in period t. These agents do not learn. Moreover, $(1-\theta)(1-\delta)\int_{\underline{k}}^{k_A^*} \phi_H(k)dk$ agents move into this interval from $[\underline{k}, k_A^*]$. They are the cohort $t-1$ agents who were workers in period $t-1$ but become managers in period t. Therefore, the supply of cohort $t-1$ managers is $(1-\delta)[\int_{k_A^*}^{\overline{k}_H} \phi_H(k)dk + (1-\pi)\int_{\underline{k}}^{k_A^*} \phi_H(k)dk]$. The supply and demand for managers in other cohorts can be obtained in a similar fashion. Adding up the demand for managers and the supply of managers in each cohort, we

get

$$\text{Demand for managers } = \frac{1}{1-\theta(1-\delta)} \int_{\underline{k}}^{k_A^*} \frac{\phi_H(k)}{n(k;\beta)} dk$$

$$\text{Supply of managers} = \frac{1}{\delta(1-\theta(1-\delta))}[[1-\theta(1-\delta)] \int_{k_A^*}^{\overline{k}_H} \phi_H(k)dk + (1-\theta)(1-\delta) \int_{\underline{k}}^{k_A^*} \phi_H(k)dk]$$

In equilibrium, supply must equal demand. Equating the above two expressions and after a bit of algebra, we obtain the following :

$$\int_{\underline{k}}^{k_A^*} \frac{\phi_H(k)}{n(k;\beta)} dk = \int_{k_A^*}^{\overline{k}_H} \phi_H(k)dk + (1-\theta)(\frac{1-\delta}{\delta})$$

In order to derive the properties of k_A^*, we use the Implicit Function Theorem. Differentiating the above equation w.r.t. π,

$$\frac{\partial k_A^*}{\partial \pi} \frac{\phi_H(k_A^*)}{n(k_A^*;\beta)} = -\frac{\partial k_A^*}{\partial \pi} \phi_H(k_A^*) - (\frac{1-\delta}{\delta})$$

Therefore,

$$\frac{\partial k_A^*}{\partial \pi}[\frac{\phi_H(k_A^*)}{n(k_A^*;\beta)} + \phi_H(k_A^*)] = -(\frac{1-\delta}{\delta})$$

Since the LHS is positive while the RHS is negative, $\frac{\partial k_A^*}{\partial \pi} < 0$. In a similar fashion it can be shown that $\frac{\partial k_A^*}{\partial \delta} < 0$.

Differentiating the labour market clearing condition w.r.t. β,

$$\frac{\partial k_A^*}{\partial \beta} \frac{\phi_H(k_A^*)}{n(k_A^*;\beta)} + \int_{\underline{k}}^{k_A^*} \frac{-\phi_H(k)}{n(k;\beta)^2} \frac{\partial n(k;\beta)}{\partial \beta} dk = -\frac{\partial k_A^*}{\partial \beta} \phi_H(k_A^*)$$

Re-arranging terms, we have

$$\frac{\partial k_A^*}{\partial \beta}[\frac{\phi_H(k_A^*)}{n(k_A^*;\beta)} + \phi_H(k_A^*)] = \int_{\underline{k}}^{k_A^*} \frac{\phi_H(k)}{n(k;\beta)^2} \frac{\partial n(k;\beta)}{\partial \beta} dk$$

Since $\frac{\partial n(k;\beta)}{\partial \beta} > 0$, both the LHS and the RHS are positive. Therefore, $\frac{\partial k_A^*}{\partial \beta} > 0$

Proof of Lemma 8: Let G f.o.s.d. H. Let g and h be the corresponding densities. Also, let $\xi(k)$ be the demand for manager per worker, where the worker has knowledge k. Since the span of control is only a function of the worker's knowledge, a worker with knowledge k works in a firm of size $n(k)$. Hence $\xi(k)$ is simply the reciprocal of $n(k)$. Therefore $\xi'(k) < 0$ (this follows from $n'(k) > 0$). Also, let k^* be the threshold under H.

We shall prove the lemma by contradiction. Let k^* also be the threshold for G. We can have two possibilities - (i) $g(k) < h(k)$ for all $k < k^*$. In this case, the demand for managers under $G = \int_{\underline{k}}^{k^*} \xi(k)g(k)dk < \int_{\underline{k}}^{k^*} \xi(k)h(k)dk =$ demand for managers under H. But the supply of managers under $G = 1 - G(k^*) > 1 - H(k^*) =$supply of managers under H. Hence at k^*, there is an excess supply of managers under G. This means that the threshold for G must be greater than k^*. (ii) There are n intervals $A_i \subset [\underline{k}, k^*]$, $i = 1, ...n$ such

that

$$\begin{aligned} g(k) &> h(k) \forall k \in A_i, \forall i \\ g(k) &< h(k) \text{ otherwise} \end{aligned}$$

Rank the $A_i s$ such that $A_i < A_j \Rightarrow maxA_i < minA_j$. We proceed as follows - We know that $\underline{k} < minA_1 = a_1(say)$ (otherwise H would f.o.s.d. G). Let $B = [\underline{k}, a_1]$. Then it must be the case that $g(k) < h(k)$ for all $k \in B$. G f.o.s.d. H implies that

$$\int_B g(k)dk + \int_{A_1} g(k)dk < \int_{A_1} h(k)dk + \int_B h(k)dk$$

Re-arranging the above equation,

$$\int_{A_1} [g(k) - h(k)]dk < \int_B [h(k) - g(k)]dk$$

Multiplying both sides by $\xi(a_1)$,

$$\int_{A_1} \xi(a_1)[g(k) - h(k)]dk < \int_B \xi(a_1)[h(k) - g(k)]dk$$

Now, $\xi'(k) < 0$ implies that $\xi(a_1) < \xi(k) \forall k \in B$ and $\xi(a_1) > \xi(k) \forall k \in A_1$. Replacing $\xi(a_1)$ in the above equation,

$$\int_{A_1} \xi(k)[g(k) - h(k)]dk < \int_B \xi(k)[h(k) - g(k)]dk$$

Here we are using the fact that $h(k) - g(k) > 0 \forall k \in B$ and $g(k) - h(k) > 0 \forall k \in A_1$. We re-arrange again to obtain

$$\int_B \xi(k)g(k)dk + \int_{A_1} \xi(k)g(k)dk < \int_{A_1} \xi(k)h(k)dk + \int_B \xi(k)h(k)dk$$

The LHS and the RHS are the demand for managers by workers in $B \cup A_1$ under G_1 and G_2 respectively. Define $maxA_1 = a_1'$ and $minA_2 = a_2$. Let $C = [a_1', a_2]$. G f.o.s.d. H implies that

$$\int_B g(k)dk + \int_{A_1} g(k)dk + \int_C g(k)dk + \int_{A_2} g(k)dk < \int_B h(k)dk + \int_{A_1} h(k)dk + \int_C h(k)dk + \int_{A_2} h(k)dk$$

Re-arranging, we have

$$\int_{A_2} [g(k) - h(k)]dk < (\int_B [h(k) - g(k)]dk - \int_{A_1} [g(k) - h(k)]dk) + \int_C [h(k) - g(k)]dk$$

Multiplying both sides by $\xi(a_2)$,

$$\int_{A_2} \xi(a_2)[g(k) - h(k)]dk < \xi(a_2)(\int_B [h(k) - g(k)]dk - \int_{A_1} [g(k) - h(k)]dk) + \int_C \xi(a_2)[h(k) - g(k)]dk$$

Since $\xi'(k) < 0$, we have

$$\begin{aligned} \int_{A_2} \xi(k)[g(k) - h(k)]dk &< \xi(a_2)(\int_B [h(k) - g(k)]dk - \int_{A_1} [g(k) - h(k)]dk) \\ &\quad + \int_C \xi(k)[h(k) - g(k)]dk \end{aligned}$$

Again, using $\xi'(k) < 0$ in the above inequality

$$\begin{aligned} LHS &< \xi(a_1)(\int_B [h(k) - g(k)]dk - \int_{A_1} [g(k) - h(k)]dk) \\ &+ \int_C \xi(k)[h(k) - g(k)]dk \\ &\int_B \xi(k)[h(k) - g(k)]dk - \int_{A_1} \xi(k)[g(k) - h(k)]dk \\ &+ \int_C \xi(k)[h(k) - g(k)]dk \end{aligned}$$

Re-arranging gives us that the demand for managers by workers in $B \cup A_1 \cup C \cup A_2$ under G is less than that under H. We can repeat this argument by expanding the set till we reach k^*. But then we have shown that the demand for managers under G is less than that under H. However the supply of managers under G is greater than that under H. Therefore, at k^*, there is an excess supply of managers under G. Hence the threshold under G has to be greater than k^*.

Proof of Proposition 2: The derivation of the threshold is the same as in the proof of Proposition 3. Equating the supply of managers and the demand for managers, we have

$$\int_{\underline{k}}^{k_I^*} \frac{\phi_H(k) + \phi_F(k)}{2n(k;\beta)}dk = \int_{k_I^*}^{\overline{k}_F} \frac{\phi_H(k) + \phi_F(k)}{2}dk + (1-\theta)(\frac{1-\delta}{\delta})$$

Re-arranging, we have

$$\int_{\underline{k}}^{k_I^*} \frac{\phi_H(k)}{n(k;\beta)}dk - \int_{k_I^*}^{\overline{k}_H} \phi_H(k)dk - (1-\theta)(\frac{1-\delta}{\delta}) = \int_{k_I^*}^{\overline{k}_H} \phi_F(k)dk + (1-\theta)(\frac{1-\delta}{\delta}) - \int_{\underline{k}}^{k_I^*} \frac{\phi_F(k)}{n(k;\beta)}dk$$

Not that the LHS is the excess demand for managers in the Home country if the threshold is k_I^*, while the RHS is the excess supply of managers in the Foreign country if the threshold is k_I^*. If $k_I^* = k_A^*$, the LHS is equal to 0, i.e.

$$\int_{\underline{k}}^{k_A^*} \frac{\phi_H(k)}{n(k;\beta)}dk - \int_{k_A^*}^{\overline{k}_H} \phi_H(k)dk = (1-\theta)(\frac{1-\delta}{\delta})$$

Since $\phi_F(k)$ f.o.s.d. $\phi_H(k)$, from Lemma 6, we know that

$$\int_{\underline{k}}^{k_A^*} \frac{\phi_F(k)}{n(k;\beta)}dk - \int_{k_A^*}^{\overline{k}_H} \phi_F(k)dk < (1-\theta)(\frac{1-\delta}{\delta})$$

Therefore, for $k_I^* = k_A^*$, the RHS is positive. But this means that $k_I^* \neq k_A^*$. In particular, since the LHS is increasing in k_I^* and the RHS is decreasing, it must be the case that $k_I^* > k_A^*$.

Proof of Proposition 4: We know that an allocation A is a Pareto improvement over allocation B if $u(x_i^A) \geq u(x_i^B)$ for all i, and $u(x_j^A) > u(x_j^B)$ for some j. This suggests that in order to show that A *is not* a Pareto improvement over B, it is sufficient to show that $\exists$ individuals 1 and 2 s.t. $u(x_1^A) \geq u(x_1^B) \Rightarrow u(x_2^A) < u(x_2^B)$ and vice versa. From Lemma 6, we have

$$k_{A,NL}^* < k_{I,NL}^*$$

where NL refers to no-learning. If there are incumbent firms in the Home economy, this also means that

$$k^*_{I,NL} < \overline{k}_H$$

The above inequality suggests that under Integration, there are incumbent Home managers who continue to operate ($k \in [k^*_{I,NL}, \overline{k}_H]$). At the same time, under Autarky, $m^{-1}_{A,NL}(k^*_{A,NL}) = \underline{k} \Rightarrow m^{-1}_{A,NL}(k^*_{I,NL}) > \underline{k}$ (follows from PAM). While under Integration, $m^{-1}_{A,NL}(k^*_{I,NL}) = \underline{k}$, i.e., under Integration, the manager with knowledge $k^*_{I,NL}$ has a worse match. The present value of $k^*_{I,NL}$ is just the period profits $\pi_{I,NL}(k^*_{I,NL})$ divided by δ.

$$\begin{aligned}
\pi_{A,NL}(k^*_{I,NL}) &= (f(k^*_{I,NL}) - w_{A,NL}(m^{-1}_{A,NL}(k^*_{I,NL})))n(m^{-1}_{A,NL}(k^*_{I,NL})) \\
&\geq (f(k^*_{I,NL}) - w_{A,NL}(k))n(k) \quad \forall k \\
&\geq (f(k^*_{I,NL}) - w_{A,NL}(\underline{k}))n(\underline{k})
\end{aligned}$$

Note that $\pi_{I,NL}(k^*_{I,NL}) = (f(k^*_{I,NL}) - w_{I,NL}(\underline{k}))n(\underline{k})$. Therefore, the relation between $\pi_{A,NL}(k^*_{I,NL})$ and $\pi_{I,NL}(k^*_{I,NL})$ depends on the relation between $w_{A,NL}(\underline{k})$ and $w_{I,NL}(\underline{k})$. Let us consider the following cases -

(a) $w_{A,NL}(\underline{k}) < w_{I,NL}(\underline{k})$: In this case, $\pi_{A,NL}(k^*_{I,NL}) > \pi_{I,NL}(k^*_{I,NL}) \Rightarrow \underline{k}$ is strictly better-off under Integration but $k^*_{I,NL}$ is strictly worse-off.

(b) $w_{A,NL}(\underline{k}) \gg w_{I,NL}(\underline{k})$: Then it is possible to have, $\pi_{A,NL}(k^*_{I,NL}) < \pi_{I,NL}(k^*_{I,NL}) \Rightarrow k^*_{I,NL}$ is strictly better-off under Integration but $\underline{k}$ is strictly worse-off.

(c) $w_{A,NL}(\underline{k}) = w_{I,NL}(\underline{k})$: In this case, $\pi_{A,NL}(k^*_{I,NL}) \geq \pi_{I,NL}(k^*_{I,NL})$. This is not a negation of Pareto improvement. However let us choose the agent with knowledge $k^*_{I,NL} + \epsilon$ such that $m^{-1}_{I,NL}(k^*_{I,NL} + \epsilon) < m^{-1}_{A,NL}(k^*_{I,NL} + \epsilon)$. Since $m(.)$ is continuous, we can always find such an ϵ. Moreover, since $m(.)$ is a function, its inverse must be strictly monotonic. Hence $m^{-1}_{I,NL}(k^*_{I,NL} + \epsilon) > m^{-1}_{I,NL}(k^*_{I,NL}) = \underline{k}$. Now

$$w'_{A,NL}(\underline{k}) = \frac{(f(k'_{A,NL}) - w_{A,NL}(\underline{k}))n'(\underline{k}}{n(\underline{k})} < \frac{(f(k'_{I,NL}) - w_{I,NL}(\underline{k}))n'(\underline{k}}{n(\underline{k})} = w'_{I,NL}(\underline{k})$$

Combined with $w_{A,NL}(\underline{k}) = w_{I,NL}(\underline{k})$, this means that in the neighborhood of $k = \underline{k}$, $w_{A,NL}(k) < w_{I,NL}(k)$. Hence,

$$\begin{aligned}
\pi_{A,NL}(k^*_{I,NL} + \epsilon) &= (f(k^*_{I,NL} + \epsilon) - w_{A,NL}(m^{-1}_{A,NL}(k^*_{I,NL} + \epsilon)))n(m^{-1}_{A,NL}(k^*_{I,NL} + \epsilon)) \\
&\geq (f(k^*_{I,NL} + \epsilon) - w_{A,NL}(m^{-1}_{I,NL}(k^*_{I,NL} + \epsilon)))n(m^{-1}_{I,NL}(k^*_{I,NL} + \epsilon))
\end{aligned}$$

Using the fact that $w_{A,NL}(k^*_{I,NL} + \epsilon) < w_{I,NL}(k^*_{I,NL} + \epsilon)$, we have

$$\pi_{A,NL}(k^*_{I,NL} + \epsilon) > (f(k^*_{I,NL} + \epsilon) - w_{I,NL}(m^{-1}_{I,NL}(k^*_{I,NL} + \epsilon)))n(m^{-1}_{I,NL}(k^*_{I,NL} + \epsilon)) = \pi_{I,NL}(k^*_{I,NL} + \epsilon)$$

Therefore $k^*_{I,NL} + \epsilon$ is strictly worse-off. Hence, for all the 3 cases (a), (b) and (c), we have shown that at least one individual is worse-off. Since these cases are exhaustive, the result follows.

Proof of Theorem 5: We shall proceed as follows - First, we shall find the condition under which $\overline{k}_H$ is better-off under Integration. Since $\overline{k}_H$ is a manager under both Autarky and Integration, we have to show that $V_{M,I}(\overline{k}_H) > V_{M,A}(\overline{k}_H)$. Since $\overline{k}_H$ is matched with k^*_A under Autarky and $\widetilde{k}$ under Integration, this implies that

$$[f(\overline{k}_H) - w_A(k^*_A)]n(k^*_A) < [f(\overline{k}_H) - w_I(\widetilde{k})]n(\widetilde{k})$$

Re-arranging, we have

$$f(\overline{k}_H)[n(k^*_A) - n(\widetilde{k})] < w_A(k^*_A)n(k^*_A) - w_I(\widetilde{k})n(\widetilde{k})$$

Now,

$$\begin{aligned} w_A(k_A^*)n(k_A^*) - w_I(\widetilde{k})n(\widetilde{k}) &= \frac{\delta+(1-\theta)(1-\delta)}{\delta}[f(k_A^*)n(\underline{k})\frac{n(k_A^*)}{1+n(k_A^*)} - f(k_I^*)n(\underline{k})\frac{n(k_I^*)}{1+n(k_I^*)}] \\ &+\frac{(1-\theta)(1-\delta)}{\delta}[f(\overline{k}_F)n(k_I^*)\frac{n(k_I^*)}{1+n(k_I^*)} - f(\overline{k}_H)n(k_A^*)\frac{n(k_A^*)}{1+n(k_A^*)}] \\ &+\frac{\delta+(1-\theta)(1-\delta)}{\delta}[\frac{n(k_A^*)}{1+n(k_A^*)}\int_{\underline{k}}^{k_A^*} f(m_A(k))n'(k)dk \end{aligned}$$

$$+\frac{n(k_I^*)}{1+n(k_I^*)}\int_{\widetilde{k}}^{k_I^*} f(m_I(k))n'(k)dk] - \int_{\underline{k}}^{\widetilde{k}} f(m_I(k))n'(k)dk$$

Let us consider each term on the RHS.

$$f(k_A^*)n(\underline{k})\frac{n(k_A^*)}{1+n(k_A^*)} - f(k_I^*)n(\underline{k})\frac{n(k_I^*)}{1+n(k_I^*)} > f(\underline{k})n(\underline{k})\frac{n(\underline{k})}{1+n(\underline{k})} - f(\overline{k}_H)n(\underline{k})\frac{n(\overline{k}_H)}{1+n(\overline{k}_H)}$$

$$f(\overline{k}_F)n(k_I^*)\frac{n(k_I^*)}{1+n(k_I^*)} - f(\overline{k}_H)n(k_A^*)\frac{n(k_A^*)}{1+n(k_A^*)} > f(\overline{k}_F)n(\underline{k})\frac{n(\underline{k})}{1+n(\underline{k})} - f(\overline{k}_H)n(\overline{k}_H)\frac{n(\overline{k}_H)}{1+n(\overline{k}_H)}$$

$$-\int_{\underline{k}}^{\widetilde{k}} f(m_I(k))n'(k)dk > -f(\overline{k}_H)[n(\overline{k}_H) - n(\underline{k})]$$

$$\frac{n(k_A^*)}{1+n(k_A^*)}\int_{\underline{k}}^{k_A^*} f(m_A(k))n'(k)dk > 0,\ \frac{n(k_I^*)}{1+n(k_I^*)}\int_{\widetilde{k}}^{k_I^*} f(m_I(k))n'(k)dk > 0$$

Replacing them in the above equation,

$$\begin{aligned} w_A(k_A^*)n(k_A^*) - w_I(\widetilde{k})n(\widetilde{k}) &> \frac{\delta+(1-\theta)(1-\delta)}{\delta}[f(\underline{k})n(\underline{k})\frac{n(\underline{k})}{1+n(\underline{k})} - f(\overline{k}_H)n(\underline{k})\frac{n(\overline{k}_H)}{1+n(\overline{k}_H)}] \\ &+\frac{(1-\theta)(1-\delta)}{\delta}[f(\overline{k}_F)n(\underline{k})\frac{n(\underline{k})}{1+n(\underline{k})} - f(\overline{k}_H)n(\overline{k}_H)\frac{n(\overline{k}_H)}{1+n(\overline{k}_H)}] \\ &-f(\overline{k}_H)[n(\overline{k}_H) - n(\underline{k})] \\ &= A(say) \end{aligned}$$

Furthermore,

$$\begin{aligned} f(\overline{k}_H)[n(k_A^*) - n(\widetilde{k})] &< f(\overline{k}_H)[n(\overline{k}_H) - n(\underline{k})] \\ &= B(say) \end{aligned}$$

Hence the sufficient condition for $\overline{k}_H$ to be strictly better-off under Integration is that $A > B$. After a bit of algebra, this condition reduces to

$$(1-\theta)(\frac{1-\delta}{\delta}) > \frac{2f(\overline{k}_H)[n(\overline{k}_H) - n(\underline{k})] + n(\underline{k})[\mu(\overline{k}_H)f(\overline{k}_H) - \mu(\underline{k})f(\underline{k})]}{\mu(\underline{k})[f(\overline{k}_F) + f(\underline{k})]n(\underline{k}) - \mu(\overline{k}_H)[n(\underline{k}) + n(\overline{k}_H)]f(\overline{k}_H)}$$

where $\mu(\underline{k}) = \frac{n(\underline{k})}{1+n(\underline{k})}$ and $\mu(\overline{k}_H) = \frac{n(\overline{k}_H)}{1+n(\overline{k}_H)}$. Of course, this only ensures that $\overline{k}_H$ is strictly better off. We need to show that every Home agent can be made better off.

Notice that for $k \in [\underline{k}, k_A^*]$, agents are workers under both regimes. For $k \in [k_A^*, k_I^*]$, agents are workers under Integration but managers under Autarky. Finally for $k \in [k_I^*, \overline{k}_H]$, agents are managers under both regimes. In the steady-state, $V_{M,i}(k) = \frac{1}{\delta}\pi_i(k)$, $i \in \{A, I\} \Rightarrow V'_{M,i}(k) = \frac{1}{\delta}\pi'_i(k) = \frac{1}{\delta}f'(k)n(m_i^{-1}(k))$. For $k \in [k_I^*, \overline{k}_H]$,

$$\begin{aligned} m_I^{-1}(k) &< m_A^{-1}(k) \Rightarrow \frac{1}{\delta}f'(k)n(m_I^{-1}(k)) < \frac{1}{\delta}f'(k)n(m_A^{-1}(k)) \\ &\Rightarrow V'_{M,I}(k) < V'_{M,A}(k) \end{aligned}$$

Suppose $V_{M,I}(\overline{k}_H) > V_{M,A}(\overline{k}_H)$. Since $V_{M,A}(.)$ is decreasing at a faster rate than $V_{M,I}(.)$ in the neighborhood $[k_I^*, \overline{k}_H]$, this implies that $V_{M,I}(k) > V_{M,A}(k)$ for $k \in [k_I^*, \overline{k}_H]$. In particular, $V_{M,I}(k_I^*) > V_{M,A}(k_I^*)$. For $k \in [k_A^*, k_I^*]$,

$$V_{W,I}(k) = \frac{1}{\delta + (1-\theta)(1-\delta)}w_I(k) + \frac{(1-\theta)(1-\delta)}{\delta(\delta + (1-\theta)(1-\delta))}f(m_I(k))n(k)$$

$$\Rightarrow V'_{W,I}(k) = \frac{1}{\delta + (1-\theta)(1-\delta)}f(m_I(k))n'(k) + \frac{(1-\theta)(1-\delta)}{\delta(\delta + (1-\theta)(1-\delta))}f'(m_I(k))n(k)m'_I(k)$$

Also, $V'_{M,A}(k) = \frac{1}{\delta}f'(k)n(m_A^{-1}(k))$. When $\delta = 1$, $V'_{W,I}(k) = f(m_I(k))n'(k)$ and $V'_{M,A}(k) = f'(k)n(m_A^{-1}(k))$. Now,

$$\frac{f'(k)}{f(m_I(k))} \geq \frac{f'(\underline{k})}{f(m_I(k))} > \frac{f'(\underline{k})}{f(\overline{k}_H)} > \frac{n'}{(}\underline{k})n(\underline{k}) \geq \frac{n'}{(}m_A^{-1}(k))n(m_A^{-1}(k)) > \frac{n'(k)}{n(m_A^{-1}(k))}$$

Hence, $V'_{M,A}(k) > V'_{W,I}(k)$. Therefore, $\exists\ \delta_1$ s.t. $\forall \delta > \delta_1$, $V'_{M,A}(k) > V'_{W,I}(k)$ and hence $V_{W,I}(k) > V_{M,A}(k)$ for $k \in [k_A^*, k_I^*]$. In particular, $V_{W,I}(k_A^*) > V_{M,A}(k_A^*)$. For $k \in [\underline{k}, k_A^*]$,

$$V'_{W,A}(k) = \frac{1}{\delta + (1-\theta)(1-\delta)}f(m_A(k))n'(k) + \frac{(1-\theta)(1-\delta)}{\delta(\delta + (1-\theta)(1-\delta))}f'(m_A(k))n(k)m'_A(k)$$

and

$$V'_{W,I}(k) = \frac{1}{\delta + (1-\theta)(1-\delta)}f(m_I(k))n'(k) + \frac{(1-\theta)(1-\delta)}{\delta(\delta + (1-\theta)(1-\delta))}f'(m_I(k))n(k)m'_I(k)$$

When $\delta = 1$, $V'_{W,A}(k) = f(m_A(k))n'(k) > f(m_I(k))n'(k) = V'_{W,I}(k)$. Therefore, $\exists\ \delta_2$ s.t. $\forall \delta > \delta_2$, $V'_{W,A}(k) > V'_{W,I}(k)$ and hence $V_{W,I}(k) > V_{w,A}(k)$ for $k \in [k_A^*, k_I^*]$. Hence, if we choose $\delta^* = \max\{\delta_1, \delta_2\}$, $\forall \delta > \delta^*$, $V_{W,I}(k) > V_{w,A}(k)$ for $k \in [\underline{k}, \overline{k}_H]$.