Inefficient Labor Market Sorting

Carsten Eckel (LMU Munich)
Stephen R. Yeaple (Pennsylvania State University)

Discussion Paper No. 437

October 23, 2023

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Georg Weizsäcker, Ph.D., Humboldt University Berlin, 10117 Berlin, Germany
info@rationality-and-competition.de

Inefficient Labor Market Sorting*

Carsten Eckel† Stephen R. Yeaple‡

August 27, 2023

Abstract

A growing empirical literature attributes much of the productivity advantages of large, "superstar" firms to their adoption of best practice management techniques that allow them to better identify and use talented workers. The reasons for the incomplete adoption of these "structured management practices" and their welfare implications are not well understood. This paper provides a positive and normative analysis of these issues in a theoretical framework in which structured management practices induce sorting of talent across firms. Incomplete adoption arises because worker talent is in limited supply. In equilibrium there is excessive adoption of structured management practices and too much sorting of talented workers into large firms. In this second-best environment, policy changes that favor large firms, such as trade liberalization, have the potential to lower welfare.

Keywords: Labor Market Imperfections, Misallocation, Productivity, Wage Inequality, International Trade, Welfare.

JEL Classification: F12, F16, J31, J33, J42, M51.

*We are thankful to participants of seminars in UC Berkeley, U of Colorado, Dartmouth, UCD Dublin, Geneva, Hohenheim, Hong Kong, Leuven, the London School of Economics, Oxford, U of Pittsburgh, the Tinbergen Institute, the NBER ITI winter meeting, the LMU/CAS workshop on Public Economic Policy Responses to International Trade Consequences, the MaCCI Workshop on Multiproduct Firms in IO and Trade, the QUANTAGG workshop 2018, the 5th Hitotsubashi Summer Institute, ETSG Bern, and the German Economic Association 2019. The authors gratefully acknowledge support from the German Research Foundation (DFG) through SFB TR 190 and the Center for Economic Studies (CES), Munich. Yeaple also acknowledges support from NSF grant SES-1360209.

†University of Munich, CESifo and CEPR / carsten.eckel@econ.lmu.de.

‡Pennsylvania State University, CESifo and NBER

1 Introduction

Large and productive 'superstar' firms dominate the global economy (Bernard et al., 2007; Shane, 2012; Criscuolo et al., 2014; Caruso, 2015; Bernard and Okubo, 2016; Bureau of Labor Statistics, 2017b; Autor et al., 2020). To understand the origins of these firms' success, researchers have begun to investigate the role played by managerial quality. In a series of articles, summarized in Bloom et al. (2014), the authors report that highly productive and profitable firms perform well in terms of objective measures of their management quality. They find that superstar firms are better at managing human resources: not only are superstar firms better at attracting talented workers, but they are also better at monitoring production, at setting targets, and at providing incentives to their employees (e.g. Bender et al. 2018).[1] Following Bloom et al. (2019), we refer throughout this paper to this package of monitoring and incentive provision techniques as "structured management practices."

The emerging consensus from the empirical literature is that best practices exist but few firms fully implement them. This raises an uncomfortable question: Why, if there is a "best practice" and this practice is available to everyone, do only some firms follow this practice? An answer to this question has important policy implications because on the basis of the empirical correlations alone one would be tempted to propose subsidies to adopting superstar firms.[2] This motivates a related question. What are the welfare implications of incomplete adoption of the package of structured management practices? The theory necessary to answer these questions is absent from the literature.[3]

This paper develops a theory that addresses both questions. This theory contains the crucial elements identified by the empirical literature: workers that are heterogeneous in their ability (or skills, used as synonyms throughout), ability that is costly to observe, and management intensive techniques that raise employee productivity by providing incentives to increase their effort. In its simplest version, ex ante identical firms compete in a monopolistically competitive output market and in a competitive labor market and choose either to

[1] Bender et al. (2018) directly document the connection between firms surveyed management practices and their stock of talent labor, writing "(...) better managed firms are able to build up a superior stock of employees through the selective hiring and attrition. In particular, examining job inflows and outflows at the plants in our sample, we find that those with higher management scores are more likely to recruit higher ability workers (measured by the permanent component in their earnings) and are less likely to layoff or fire the highest skilled workers."(p. S374)

[2] For instance, Bloom et al (2019) suggest that incomplete adoption occurs because many firms do not know what the best practices are. They propose, but do not formally model, a situation in which spillovers from superstar firms might promote learning and so might justify subsidies to adopters.

[3] Indeed, the literature on personnel economics focuses on outcomes in partial equilibrium and largely ignores the welfare effects of organizational practices. See Lazear and Oyer (2013, p 480) who write, "Personnel economics typically focus on welfare within a given employment relationship rather than on the overall social welfare function."

produce using rudimentary management practices or to adopt so-called structured management practices. The rudimentary management practices features low worker productivity because it cannot provide individual incentives: Only the aggregate output of their workers can be observed. Best-practice management allow adopters to monitor individual worker output and so allows them to raise worker productivity by incentivizing effort but comes at the cost of greater managerial overhead expenses. We capture the emphasis of the importance of identifying worker ability by assuming that the output return to better incentives is increasing in ability.

In equilibrium, the model predicts that only some firms adopt best-practice management techniques, and these firms exhibit superstar characteristics. These superstar firms attract the most able workers, pay the highest wages, exhibit the highest labor productivity, and are more likely to engage in international trade. Despite the better observed performance of these adopting firms, in equilibrium firms that do not adopt what appear to be the package of best practice techniques coexist with the superstars. Intuitively, adopting firms have lower marginal cost of production relative to non-adopters, but pay high fixed costs to procure better management inputs. The marginal cost advantage is a function of the price of skilled labor, and this price will rise until firms are indifferent between adopting and not adopting the apparently "best-practice" techniques. Thus, the answer to our first question above is that not all firms adopt best-practice management techniques because non-adopters can still break even despite their apparent inefficiency because they do not require high overhead costs.

Turning to the second question, we show that the market equilibrium features the excessive allocation of resources to superstar firms relative to the social optimum. Here, imperfectly observable ability is crucial. Workers that opt to produce for firms that do not adopt best-practice management techniques are paid a wage that reflects the average productivity of their colleagues whereas workers that are employed by firms using the best-practice management techniques (those with monitoring practices) are paid a wage that reflects their individual ability (and effort). This has two important consequences: First, on the margin, the ability of the least able worker at the intensively managed firm is above the average ability for workers employed with by passive managers. And second, in a competitive labor market a sorting equilibrium arises where the least able worker employed by adopting firms is paid her outside option: the wage paid by firms that do not adopt. This means that the marginal worker is paid the same at both adopting and non-adopting firms but has a higher marginal revenue product at the non-adopting firm.

The way that this inefficiency manifests itself depends on the assumptions regarding firm entry. In our baseline case, entry as an adopting or non-adopting firm is free. Because entry is free, the free entry condition pins down the effective cost of labor in equilibrium so that firms' sizes are optimal. This has the flavor of the analysis of Dixit and Stiglitz (1977). The inefficiency instead shows up as excessive entry of adopting firms (with high fixed cost) and too little variety available in the market. In our second case, we consider the other extreme in which the measures of adopting and non-adopting firms are fixed so that firms are endowed with their management styles. In this case, the market equilibrium features large, high-wage firms that are too large and too profitable. In either case, a social planner charged with maximizing the utility of a representative agent could reach the first-best by subsidizing employment at small (non-adopting) firms.

To explore the welfare implications of policy changes, we consider the case of trade liberalization: As international trade favors large scale operations, it will shift resources away from small to superstar firms (Bernard et al., 2007; Freund and Pierola, 2015; Bernard et al. 2018). The information asymmetries induced by the incomplete adoption of management techniques cause globalization to create problems of the second best. On the one hand, there is an increase in the direct (and standard) gains from trade that obtain from imported varieties at lower cost as trade costs fall. On the other hand, lower trade costs shift resources from small firms where their shadow value is high to large firms where their shadow value is low. This is a negative welfare effect. Which effect dominates depends on the initial size of the gap in the shadow value of the reallocated workers and this in turn depends on how rare superstar firms are initially.

To further understand the welfare implications of the model we extend the model to a two-period setting in which the wages paid by adopting firms in the first period provide public information about worker types in the second. In this setting, there are no information frictions regarding older workers with experience at high wage firms and some of them will switch jobs mid-career eliminating the misallocation of older workers. However, because young workers anticipate the benefit of starting their career at an adopting firm, the misallocation of labor becomes more severe in the first period.

Our paper contributes to several related literatures. First, and foremost, our paper contributes to the literature on the adoption of management practices that has been revolutionized in recent years by Bloom and Van Reenan (2007).[4] This largely empirical literature has convincingly demonstrated that a serious problem facing management is the difficulty of identifying, monitoring, and thereby rewarding productive workers. Said differently, the literature demonstrates that the firms using "best-practice" management techniques have

[4] Other related papers that analyze the problem of identifying talented workers is the firm organization and personnel economics literature. See Lazear and Shaw (2007), Oyer and Schaefer (2011) and Bandiera et al (2015) for recent overviews and empirical evidence. This literature has generally neglected the general equilibrium implications of firm management practices.

information advantages that other firms lack and this is associated with these firms having "superstar" characteristics. Our contribution is to show that the asymmetries associated with incomplete adoption across firms is a natural outcome in the labor market and that it has important welfare implications. To our knowledge, ours is the first paper to provide a positive and normative analysis of this phenomenon.

Our paper also contributes to the large and growing literature on the role of market imperfections in the misallocation of resources across firms and their implications for the gains from trade. This literature has focused on issues of market power in output and input markets. For instance, Arkolakis et al. (2019) and Feenstra (2018) study the welfare effects in the presence of inefficiencies in the product market due to variable mark-ups. An important assumption in these studies is that differences in measured productivity are driven by exogenous differences in marginal factor requirements, reflecting true differences in social efficiencies across firms. Consequently, the key resource problem in these studies is that the most efficient firms are too small from a social point of view because they charge the highest mark-ups (Edmond et al., 2015; Nocco et al., 2019). On the input side, recent empirical analyses of monopsony power, such as Kroft et al. (2023) also are built on the premise that misallocation is induced by large firms restricting their output to lower input prices.

Our paper abstracts from these traditional sources of misallocation and instead focuses on the role of incomplete adoption of management practices on the allocation of talent across firms. We show that for the misallocation that arises from this source of heterogeneity across firms, the welfare effects are quite different. In fact, since the equilibrium is characterized by excessive consumption of resources by large firms (due to their artificially low cost of labor), a reallocation of resources from small firms to large superstar firms constitutes a first order *negative* welfare effect.

The remainder of this paper is organized into six sections. Section 2 introduces the model assumptions and characterizes the equilibrium. Section 3 provides an analysis of the welfare implications of labor market imperfections. The resource allocation and welfare implications of international trade liberalization are explored in section 4. Section 5 provides two important extensions to our model that are designed to understand how changes in model assumptions affect how imperfections manifest themselves in equilibrium outcomes. Section 6 provides guidance on how the welfare implications of our model could be assessed quantitively when data appropriate for "insider econometrics" is available for estimating model parameters to use in calibration. The final section summarizes and discusses the results. Additional extensions can be found in the appendix.

2 Model

In this section, we present the closed economy version of our model. We begin with the model assumptions and then characterize the equilibrium.

2.1 Key Assumptions

2.1.1 Demand

There are L households in the economy, and each houshold is endowed with one unit of labor and M/L units of management. Housholds receive utility from consuming goods and disutility from providing effort for labor:

$$U = U_c(\mathbf{q}) - \gamma\varepsilon, \tag{1}$$

where $U_c(\mathbf{q})$ is utility from consuming a set $\mathbf{q}$ of differentiated goods $q(i)$, ε is effort provided for labor and γ is the marginal disutility of effort. The disutility from effort is important for how workers respond to incentives, but we assume that it is negligable relative to the utility from consumption when it comes to welfare assessments. Technically, we assume that γ is very small relative to U_c, so that $\gamma/U_c \to 0$ and $U \to U_c$.

The utility of consumption, U_c, is CES:

$$U_c(\mathbf{q}) = \left(\int_{i\in\tilde{\Omega}} q(i)^{\frac{\sigma-1}{\sigma}} di\right)^{\frac{\sigma}{\sigma-1}}, \tag{2}$$

where $q(i)$ is the quantity consumed, σ is the elasticity of substitution between any two varieties, and $\tilde{\Omega}$ is the set of potentially consumable varieties.

Direct demand for variety $i \in \Omega$ (the set of actually produced varieties) is then given by

$$x(i) = EP^{\sigma-1}p(i)^{-\sigma}, \tag{3}$$

where $x(i)$ is economy-wide output of variety i and E is aggregate income in the economy. P stands for the price index, defined by

$$P \equiv \left(\int_{i\in\Omega} p(i)^{1-\sigma} di\right)^{\frac{1}{1-\sigma}}. \tag{4}$$

2.1.2 Production

There are two types of factors of production: Management M and labor L. Management is a homogeneous factor that is used as our numéraire. As in Yeaple (2005), labor consists of a continuum of heterogeneous workers with skills (or ability) z. The distribution of skills in the economy is described by the continuous probability density function $g(z)$ with positive support over $[\underline{z}, \infty)$ $(\underline{z} > 0)$ and its cumulative distribution function $G(\tilde{z}) = \int_{\underline{z}}^{\tilde{z}} g(z)\, dz$.

Production of a variety $x(i)$ requires fixed costs f in units of management plus marginal costs in units of (effective) labor $\tilde{c}$. These marginal costs are constant with respect to output and consist of a unit labor requirement α (in units of effective labor) and a factor cost component c: $\tilde{c} = \alpha c$. In our baseline scenario, we abstract from differences in α across firms and/or varieties and normalize α to one: $\alpha = 1$. In our appendix we discuss firm specific unit labor requirements as well as variety specific unit labor requirements in the context of multiproduct firms.[5] The fixed costs f and the factor cost component c are firm specific and will be indexed by a firm subscript j.

The productivity of an individual worker depends on their skill $z \in [\underline{z}, \infty)$ and on their effort $\varepsilon \in [1, \bar{\varepsilon}]$. These two variables determine a worker's effective supply of labor $e(\varepsilon, z)$. Both skill and effort increase a worker's productivity: $\partial e / \partial \varepsilon > 0$ and $\partial e / \partial z > 0$. In addition, we assume that $e(\varepsilon, z)$ is supermodular in ε and z: $\partial^2 e / (\partial \varepsilon \partial z) > 0$. This implies that the returns to effort are increasing in a worker's skill. For notational simplicity, we set $e(1, z) = z$ so that the productivity of a worker with minimal effort is given just by their skill z. Given these assumptions, a firm can have a more productive workforce for two reasons: Because its workforce puts in more effort, or because its workforce has better skills. Both of these channels will be important.

Two key features of structured management practices in the literature are monitoring and incentives (Bloom and Van Reenen, 2007). Firms that adopt structured management practices track and review performances of individual workers, put appropriate sanctions and rewards in place, recruit and promote talented workers and reward their workers according to their performances.[6] We follow this line of literature and translate these features into two key assumptions: First, structured management practices enable firms to observe the true productivity of individual workers (monitoring). Second, firms that adopt structured management practices reward their workers based on their true productivity (incentives). In return, firms that do *not* adopt structured management practice do not observe a worker's individual productivity (no monitoring) and, as a consequence, cannot pay a performance-based compensation (no incenitves). It is this information asymmetry between adopters and non-adopters that gives rise to the market imperfection.

The differences in monitoring and incentives between adopting and non-adopting firms

[5] See sections 8.4 and 8.5 in the appendix.

[6] See appendix I.A in Bloom and van Reenen (2007) for a full list of management practices. 10 out of the 18 practices listed relate to *Monitoring* and *Incentives*.

have consequences for the wage contracts firms can offer, for the level of effort the two types of firms can elicit from their employees, and consequently for the productivity of their workforce. Without monitoring and incentives, workers receive a salary that is independent of their effort, and they put in only the minimum effort required by their contract, normalized to one: $\varepsilon = 1$ and $e(1, z) = z$. With monitoring and incentives, workers in adopting firms receive a wage that is based on their true effective supply of labor so that both skill and effort are rewarded: A worker with skill z and effort ε receives a wage $w(\varepsilon, z) = c_j e(\varepsilon, z)$, where c_j represents the piece rate paid by firm j, or the effective wage (per units of effective labor supplied, w/e). Given this wage contract, and given our utility in (1), workers put in extra effort in adopting firms if $\partial e(1, z) / \partial \varepsilon > \gamma / c_j$. Since we assume that γ is very small, and $\gamma / c_j \to 0$, this holds for all $\varepsilon \in [1, \bar{\varepsilon}]$, and the optimal effort is $\bar{\varepsilon} > 1$.[7] Consequently, workers incentivized by a piece rate exhibit a higher productivity than if the same workers were not incentivized, $e(\bar{\varepsilon}, z) > z$, as emphasized by a large literature in personnel economics on performance pay (see Lazear, 1986, 2000 and Lazear and Oyer, 2013, for an overview).[8]

Because of these productivity enhancing effects, firms would always choose to adopt structured management practices if these practices were available costlessly. However, adopting and implementing structured management practices features high overhead costs associated with the need to hire more (or better) managers, or buy more expensive software, etc. A good example of these costs is presented by Lazear (2000) who studies the effect of piecework on productivity for the Safelite Glass Corporation: "Piecework requires measurement of output. In Safelite's case, the measurement comes about through a very sophisticated information system. But the system involves people and machines that are costly. Indeed, in equilibrium, firms that pay hourly wages or monthly salaries are probably those for whom measurement costs exceed the benefits from switching to output-based pay."[9] Consequently, we assume that firms that adopt structured management practices incur a higher fixed cost f_a (in units of management) where

$$f_a > f_n. \tag{5}$$

[7] In appendix 8.1 we elaborate more on these microfoundations, based on Lazear (1986).

[8] Differences in effort are not necessarily the only reason why firms adopting structured management practice experience an increase in the productivity of their workforce. Investments in technology (lean manufacturing, TQM, IT) can be complementary to the adoption of best management practices (Bloom and Van Reenen, 2007; Black and Lynch, 2001, 2004). These complementarities reinforce the mechanisms described here.

[9] Cf. Lazear (2000, p. 1357-8). Other examples of such overhead costs are the factory information system described by Levitt, List, and Syverson (2013) that links the speed of output and the prevalence of quality defects to workers in real time, the data-driven decision technologies described by Brynjolfsson and McElheran (2016) and the artificial intelligence software used by top firms to identify and retain the best workers (e.g. Cowgill 2020). See also Freeman and Kleiner (2005) for other types of costs associated with piecework.

where f_n is the overhead cost of non-adopters. Throughout the paper, an index a refers to firms adopting structured management practices (*adopting firms*) and an index n to firms without structured management practices (*non-adopting firms*).

For robustness, we study two variances of our main framework below. In section 5.2, we consider a version of the model in which the number of adopting and non-adopting firms is exogenous. And in appendix 8.2 we discuss the case where structured management practices increase variable costs instead of fixed (overhead) costs. The main mechanisms remain in place, so that the assumption of free entry and the type of costs associated with structured management practices is not essential for our main message.

2.1.3 Market Structure and Timing

The market for the homogeneous factor management M is perfectly competitive, and the wage of a unit of management is normalized to one. Workers L are fully informed about their own productivity z but firms that do not monitor output know only the distribution of productivity in the population, $G(z)$, which is common knowledge.[10] This is a one shot game that occurs in four stages. All agents have perfect foresight.

In stage 1, firms enter and decide whether they want to pay f_a and adopt structured management practices or not to adopt and pay $f_n < f_a$. There is a continuum of firms of both types and their masses will be denoted by n_j ($j \in \{a, n\}$).

Once firms have made their entry and management decisions, the labor market opens. While there is a single competitive labor market, for expositional purposes it is useful to think of there being two labor markets (or pools) into which workers sort. Let the set of workers that ultimately choose to be in labor market j be denoted as Z_j.

One labor market corresponds to the set of adopting firms, $j = a$, where firms monitor the performance of individual workers and offer a wage that is contingent on their *individual* output once hired (incentivized). In this labor market, all information regarding workers in that pool is effectively known by adopting firms. Perfect competition implies that the wage of worker 1 relative to worker 2 with skills z_1, $z_2 \in Z_a$ and productivities $e(\bar{\varepsilon}, z_1)$ and $e(\bar{\varepsilon}, z_2)$ satisfy the no arbitrage condition $w_1/w_2 = e(\bar{\varepsilon}, z_1)/e(\bar{\varepsilon}, z_2)$.

The other labor market corresponds to the firms that do not adopt management practices, $j = n$. These firms do not monitor and incentivize their workforce and make a wage offer that reflects the anticipated *average* output of their employees. Individual worker productivities are known only to the workers. The inability of firms $j = n$ to verify workers' individual productivities requires that there must be a single wage $w = w_n$ for all $z \in Z_n$.

[10] This assumption can be relaxed so that firms are uninformed about the exact distribution of worker skills. Qualitatively similar results obtain.

In stage 2, workers choose whether to seek work at adopting firms or at non-adopting firms. They make this choice with perfect foresight regarding the wage they would receive from each type of employer.

In stage 3, the labor market clears. And finally in stage 4, production occurs and product markets are cleared. Firms compete via monopolistic competition. Individual products are atomistic and there is no strategic interaction.

2.2 Closed Economy Equilibrium

This section characterizes the equilibrium to our closed economy model. Each stage is analyzed in sequence starting from stage 4 and progressing backward to stage 1.

2.2.1 Product Market Clearing

Given demand (3) and a market structure of monopolistic competition, the profit-maximizing price of firm $j \in \{a, n\}$ is a constant mark-up over its marginal costs:

$$p(c_j) = \frac{\sigma}{\sigma - 1} c_j. \tag{6}$$

Since all firms have access to the same technologies, and demands are symmetric across all products, all firms within one type will be symmetric. Since firms of different types are drawing their workers from different pools of laborers, their factor costs c_j may be different, hence the subscript j.

In order to simplify notation we define

$$A \equiv (\sigma - 1)^{\sigma-1} \sigma^{-\sigma} E P^{(\sigma-1)}. \tag{7}$$

This parameter A depends only on aggregate income E, the price index P, and the elasticity of substitution σ. Since firms are atomistic, A is exogenous to the firm.

Given (3), (6) and (7), output of firm j can be written as

$$x(c_j) = (\sigma - 1) A c_j^{-\sigma}, \tag{8}$$

and revenues are

$$r(c_j) \equiv p(c_j) x(c_j) = \sigma A c_j^{1-\sigma} \tag{9}$$

Finally, profits are variable profits $p(c_j) x(c_j) / \sigma$ minus fixed costs f_j:

$$\pi(c_j) = A c_j^{1-\sigma} - f_j. \tag{10}$$

2.2.2 Factor Market Clearing

Worker sorting in stage two leads to segmentation of labor markets by firm type. The labor market equilibrium for type $j \in \{a, n\}$ is

$$n_j x_j = \tilde{L}_j, \tag{11}$$

where $\tilde{L}_j$ is the effective supply of labor available to firms of type j. Since workers sort in stage two, this variable is given at this stage, and the labor market equilibrium determines the effective labor cost, c_j, facing firms of type j. In both labor markets $j \in \{a, n\}$ firms are atomistic and take wages as given.

Market clearing of the numéraire factor (management M) is implied in general equilibrium:

$$n_a f_a + n_n f_n = M. \tag{12}$$

2.2.3 Worker Sorting

Workers observe whether a firm has adopted structured management practices or not. Thus, they decide whether they want to apply for a job in a firm with monitoring and incentives or in a firm without these management practices by choosing the respective labor pool. Differences in non-pecuniary job returns (including the disutility from effort) are quantitatively negligable ($\gamma \rightarrow 0$), so this decision is entirely based on differences in wages.

The labor market of adopting firms (a) is perfectly competitive. Hence, in equilibrium, firms offer a wage schedule where effective wages (conditional on worker's output) are the same across adopting firms. Anticipating correctly the effective wage c_a determined in stage 3, and given $\varepsilon = \bar{\varepsilon}$ from section 2.1.2, firms of type-a pay

$$w_a(z) = c_a e(\bar{\varepsilon}, z). \tag{13}$$

The labor market of non-adopting firms (n) is also competitive, but firms cannot condition the wage to be paid on individual worker ability because they are not able to observe the output of individual workers. They do know the distribution of productivities of the workers that will opt to accept a job in their labor pool, however. Consequently, the wage rate cannot be conditioned on the true productivity of any particular worker, but rather depends on the expected productivity of a representative bundle of workers in this labor market segment:

$$w_n = c_n \mathbb{E}_n(Z_n), \tag{14}$$

where c_n is the effective wage rate in this labor market segment. Because firms do not reward effort in this labor pool, workers put in only minimum effort $\varepsilon = 1$, so that their effective supply of labor is given by $e(1, z) = z$, and $\mathbb{E}_n(e(Z_n)) = \mathbb{E}_n(Z_n)$.

Given that wages differ between these two types of firms, each worker can decide whether he or she wants to apply for a job at adopting or non-adopting firms. The wage of a worker with skill z is thus

$$w = \max\{c_n \mathbb{E}_n(Z_n); c_a e(\bar{\varepsilon}, z)\}. \tag{15}$$

The following proposition describes the sorting outcome:

Proposition 1 (Sorting) *In an economy that features both adopting and non-adopting firms, there exists at least one stable equilibrium that is characterized by a $\tilde{z}$ so that workers with $z > \tilde{z}$ will choose to work for adopting firms, and workers with $z < \tilde{z}$ will choose to work for non-adopting firms. The critical $\tilde{z}$ is determined by*

$$c_n \bar{z}_n(\tilde{z}) = c_a e(\bar{\varepsilon}, \tilde{z}), \tag{16}$$

where $\bar{z}_n(\tilde{z}) \equiv \int_{\underline{z}}^{\tilde{z}} z dG(z)/G(\tilde{z})$. This equilibrium is stable if $\bar{z}_n(\tilde{z})/e(\bar{\varepsilon}, \tilde{z})$ is decreasing in $\tilde{z}$.

Proof. Assume a $\tilde{z}$ exists, so that $\mathbb{E}_n(Z_n) = \int_{\underline{z}}^{\tilde{z}} z dG(z)/G(\tilde{z}) = \bar{z}_n(\tilde{z})$. Then rewrite condition (16) as $\bar{z}_n(\tilde{z})/e(\bar{\varepsilon}, \tilde{z}) = c_a/c_n$. Using L'Hôpital's rule, we can determine the limits of $\bar{z}_n(\tilde{z})/e(\bar{\varepsilon}, \tilde{z})$ as $\tilde{z}$ approaches the boundaries of the support: $\lim_{\tilde{z}\to\underline{z}}[\bar{z}_n(\tilde{z})/e(\bar{\varepsilon}, \tilde{z})] = 1$ and $\lim_{\tilde{z}\to\infty}[\bar{z}_n(\tilde{z})/e(\bar{\varepsilon}, \tilde{z})] = 0$.[11] Since $\bar{z}_n(\tilde{z})/e(\bar{\varepsilon}, \tilde{z})$ is differentiable, this proves existence of (at least) one equilibrium with $\underline{z} < \tilde{z} < \infty$ for $c_a < c_n$. Furthermore, this equilibrium implies sorting where the most skilled workers work for adopting firms and the least skilled work for non-adopting firms: $c_a e(\bar{\varepsilon}, z) > c_n \bar{z}_n(\tilde{z})$ for $z > \tilde{z}$ and $c_a e(\bar{\varepsilon}, z) < c_n \bar{z}_n(\tilde{z})$ for $z < \tilde{z}$. This equilibrium is stable if for $\zeta < \tilde{z}$, $c_n \bar{z}_n(\zeta) > c_a e(\bar{\varepsilon}, \zeta)$, and for $\zeta > \tilde{z}$, $c_n \bar{z}_n(\zeta) < c_a e(\bar{\varepsilon}, \zeta)$. Thus, stability implies that $\bar{z}_n(\zeta)/e(\bar{\varepsilon}, \zeta)$ is decreasing in ζ at $\zeta = \tilde{z}$ and requires that

$$\Gamma \equiv \frac{e'(\bar{\varepsilon}, \tilde{z})\tilde{z}}{e(\bar{\varepsilon}, \tilde{z})} - \frac{\tilde{z}g(\tilde{z})}{G(\tilde{z})}\frac{[\tilde{z} - \bar{z}_n(\tilde{z})]}{\bar{z}_n(\tilde{z})} > 0. \tag{17}$$

Since $\bar{z}_n(\tilde{z})/e(\bar{\varepsilon}, \tilde{z})$ is decreasing globally (from 1 to 0), at least one stable equilibrium must exist. This equilibrium is unique if $\bar{z}_n(\tilde{z})/e(\bar{\varepsilon}, \tilde{z})$ is monotonically decreasing. ■

[FIGURE 1 here]

[11] For $z = \underline{z}$ we assume that $e(\varepsilon, \underline{z}) = e(1, \underline{z})$.

In Figure 1 we illustrate the equilibrium and its stability graphically. For illustrative purposes, the function $\bar{z}_n(\zeta)/e(\bar{\varepsilon},\zeta)$ is not monotonic. Clearly, if $c_n\bar{z}_n(\zeta) > c_a e(\bar{\varepsilon},\zeta)$, a worker with skill ζ earns higher wages in non-adopting firms than in adopting firms. Thus, if ζ was a sorting cutoff, this would not be an equilibrium because the marginal worker would want to work for non-adopting firms, leading to an increase in this cutoff. Therefore, a stable equilibrium requires that the $\bar{z}_n(\zeta)/e(\bar{\varepsilon},\zeta)$-function intersects c_a/c_n from above. In our Figure 1, equilibria $E1$ and $E3$ are stable, $E2$ is unstable. In what follows we only consider stable equilibria, so we assume that (17) holds locally.

One important implication of the sorting equilibrium is that

$$c_a = \frac{\bar{z}_n(\tilde{z})}{e(\bar{\varepsilon},\tilde{z})} c_n < c_n. \tag{18}$$

Thus, firms adopting structured management practices pay a lower effective wage rate (in efficiency units) than non-adopting firms. This has to hold in equilibrium because the productivity of the marginal worker in adopting firms is discretely higher than the average productivity of all workers in non-adopting firms: $e(\bar{\varepsilon},\tilde{z}) > \bar{z}_n(\tilde{z})$. Therefore, non-adopting firms have to pay a premium on the effective wage rate of adopting firms in order to compensate their above-average workers for pooling them with below-average workers.

Note that the differences in efforts elicited from workers in adopting firms enlarges the wage differences in the two labor market segments, but is not a necessary condition for the labor market sorting.

Corollary 1 *The difference in efforts provided by workers in adopting and non-adopting firms is not a necessary condition for the sorting equilibrium.*

Proof. If $\bar{\varepsilon} = 1$, so that $e(\bar{\varepsilon},z) = z$, equation (18) reduces to $c_a = [\bar{z}_n(\tilde{z})/\tilde{z}]\, c_n$, where the term $\bar{z}_n(\tilde{z})/\tilde{z}$ is larger than the term $\bar{z}_n(\tilde{z})/e(\bar{\varepsilon},\tilde{z})$ in proposition 1 but is still between zero and one and behaves identically at the limits. ■

This corollary states that even if there are no productivity effects of management practices (no incentive effects), there is still incomplete adoption and sorting. This is a fundamental difference to Yeaple (2005). Yeaple (2005) has shown that differences in technologies combined with comparative advantages of skilled workers in certain types of technologies can lead to positive assortative matching of workers to firms. Here, we show that this sorting is reinforced by information asymmetries in the labor market. In fact, we even show that these information asymmetries *alone* can lead to a sorting equilibrium where skilled workers choose a different working environment than unskilled workers. As we will see in section 3, it is this additional cost advantage induced by information asymmetries that is the source of the inefficiency caused by incomplete adoption.[12]

In a sorting equilibrium, we can now also determine the effective supplies of labor $\tilde{L}_j$ for the two types of firms from (11):

$$\tilde{L}_n = LG(\tilde{z})\,\bar{z}_n(\tilde{z}) \qquad \text{and} \qquad \tilde{L}_a = L\left[1 - G(\tilde{z})\right]\bar{e}_a(\bar{\varepsilon}, \tilde{z}), \tag{19}$$

where $\bar{e}_a(\bar{\varepsilon}, \tilde{z}) \equiv \int_{\tilde{z}}^{\infty} e(\bar{\varepsilon}, z)\, dG(z) / \left[1 - G(\tilde{z})\right]$.

2.2.4 Firm Entry

All types of firms can enter and exit freely. Within types, firms are symmetric. This implies that their respective profits are driven down to zero. Given (10), this implies that

$$Ac_j^{1-\sigma} = f_j. \tag{20}$$

Taking ratios for $j = a, n$ we obtain

$$\frac{c_a}{c_n} = \left(\frac{f_n}{f_a}\right)^{\frac{1}{\sigma-1}}. \tag{21}$$

For convenience, we define

$$\Phi \equiv \left(\frac{f_n}{f_a}\right)^{\frac{1}{\sigma-1}} < 1. \tag{22}$$

Now combining (16) and (21), and using the definition in (22), we obtain

$$\frac{\bar{z}_n(\tilde{z})}{e(\bar{\varepsilon}, \tilde{z})} = \Phi. \tag{23}$$

We can now prove the following proposition:

Proposition 2 (Co-existence) *In a free entry equilibrium, adopting and non-adopting firms will coexist.*

Proof. First note from (5) that adopting firms have higher fixed costs. Therefore, a necessary condition for coexistence with free entry is that $c_a < c_n$, which is met [see (18)]. Second, we can show that an equilibrium with only one type of firm is inconsistent with free entry:

[12] Our assumption that better workers are more important to incentivize can be interpreted as a microfoundation for Yeaple (2005) in the absence of information asymmetries. This would lead to the familiar fixed cost marginal cost trade off. Here, it is the information asymmetry and the sorting that drives our results, and the incentive effects are only second order.

If $\tilde{z} \to \infty$ (no adoption), $\lim_{\tilde{z}\to\infty} c_a = c_n \lim_{\tilde{z}\to\infty} [\bar{z}_n(\tilde{z})/e(\bar{\varepsilon},\tilde{z})] = 0$ and $\lim_{\tilde{z}\to\infty} \pi_a = +\infty$. Hence, some firms must adopt structured management practices. If $\tilde{z} \to \underline{z}$ (all firms adopt), $\lim_{\tilde{z}\to\underline{z}} c_a = c_n \lim_{\tilde{z}\to\underline{z}} [\bar{z}_n(\tilde{z})/e(\bar{\varepsilon},\tilde{z})] = c_n$ and $\pi_n > \pi_a$ (because $f_n < f_a$). Hence, non-adopting firms must exist, too. Inequality (5) is a necessary condition for coexistence and hence incomplete adoption.[13] ■

Proposition 2 is at the core of our theory and it provides an explanation to why only a fraction of firms adopt the structured management practices associated with the best performing firms. It shows how firms with different cost structures can arise endogenously from ex ante identical firms due to the difficulties associated with observing worker skills and different strategies to deal with imperfect information. Adopting structured management practices requires high overhead costs that are sustainable only if they lower the marginal cost of production. With incomplete adoption, the piece rates in adopting firms are driven up to the point at which additional adoption by other firms would be unprofitable. In equilibrium, demand for skilled labor is kept in check by the high overhead costs and so the effective cost of labor remains lower in adopting firms.[14]

The fact that the effective cost of labor is lower in firms that pay the highest wages has important implications for measures of firm performance that have been associated with firms adopting structured management practices in the empirical literature. We present them here as corollaries of propositions 1 and 2:

Corollary 2 (Size) *Adopting firms have higher sales than non-adopting firms.*

Proof. It follows directly from $c_a < c_n$ that $r(c_a) > r(c_n)$. ■

Adopting firms have higher sales because they have lower marginal production costs.

Corollary 3 (Productivity) *Adopting firms are more productive than non-adopting firms as measured by revenue per worker.*

Proof. Using (6), (11), and (19), revenues per worker in adopting firms φ_a can be expressed as

$$\varphi_a \equiv \frac{p(c_a)\,x(c_a)}{L[1-G(\tilde{z})]/n_a} = \frac{\sigma}{\sigma-1} c_a \bar{e}_a(\bar{\varepsilon},\tilde{z}) \tag{24}$$

Similarly, revenues per worker in non-adopting firms φ_n can be expressed as

$$\varphi_n \equiv \frac{p(c_n)\,x(c_n)}{LG(\tilde{z})/n_n} = \frac{\sigma}{\sigma-1} c_n \bar{z}_n(\tilde{z}) \tag{25}$$

[13] This prove also requires that $\bar{z}_n(\tilde{z})/e(\bar{\varepsilon},\tilde{z})$ approaches zero and one at the upper and lower limit. If productivity levels are bounded above, or if the productivity effect of effort does not go to zero for the lowest skill level (fn 11), dominance of one type of firm is possible.

[14] If the adoption of structured management practices lead to higher variable costs that are passed on to workers, results are qualitatively identical. See also appendix 8.2.

Then, using the sorting condition (16), the ratio of the two productivity measures can be expressed as

$$\frac{\varphi_a}{\varphi_n} = \frac{\bar{e}_a\left(\bar{\varepsilon},\tilde{z}\right)}{e\left(\bar{\varepsilon},\tilde{z}\right)} > 1 \tag{26}$$

■

There are two channels contributing to the higher productivity of adopting firms to non-adopting firms. First, productivity is higher because workers put in more effort in adopting firms due to the piece rate ($\bar{\varepsilon} > 1$). The larger this difference in effort, the larger the difference in productivity: $\partial\left(\varphi_a/\varphi_n\right)/\partial\bar{\varepsilon} > 0$.[15] Second, productivity is higher because adopting firms attract workers with better skills. Even if $\bar{\varepsilon} = 1$ (no difference in effort), $\bar{e}_a\left(1,\tilde{z}\right) = \int_{\tilde{z}}^{\infty} z dG\left(z\right)/\left[1 - G\left(\tilde{z}\right)\right] \equiv \bar{z}_a\left(\tilde{z}\right)$ and $\varphi_a/\varphi_n = \bar{z}_a\left(\tilde{z}\right)/\tilde{z} > 1$. Therefore, adopting firms generate higher revenues per worker because they elicit more effort and employ higher skilled workers.

In appendix 8.5 we show one additional implication in the context of multiproduct firms: Since adopting firms pay a lower effective wage rate than non-adopting firms ($c_a < c_n$) they can expand into less efficient activities and produce more varieties than non-adopting firms.

Corollary 4 (Wages) *Adopting firms pay higher average wages.*

Proof. Non-adopting firms pay a flat wage of $w_n = c_n\bar{z}_n$. The average wage in adopting firms is $\bar{w}_a \equiv \int_{\tilde{z}}^{\infty} w\left(z\right) dG\left(z\right)/\left[1 - G\left(\tilde{z}\right)\right] = c_a\bar{e}_a\left(\bar{\varepsilon},\tilde{z}\right)$. Again using (16), the relative average wage in adopting firms is

$$\frac{\bar{w}_a}{w_n} = \frac{\bar{e}_a\left(\bar{\varepsilon},\tilde{z}\right)}{e\left(\bar{\varepsilon},\tilde{z}\right)}\left(=\frac{\varphi_a}{\varphi_n}\right) > 1. \tag{27}$$

■

Adopting firms appear more productive despite paying higher wages because they have a more productive labor pool and pass on the gains from the higher labor productivity only incompletely.

The following figure shows the profile of wages as a function of worker productivity.

[FIGURE 2 here]

In Figure 2, the thick green line depicts the hockey stick profile of wages as a function of workers' productivities. Workers in the range $z \in [\underline{z},\tilde{z})$ self-select into non-adopting firms and provide effort $\varepsilon = 1$. They receive a flat wage given by $w_n = c_n\bar{z}_n$. Above $\tilde{z}$, workers decide to work for firms that adopt structured management practices, provide

[15] Because of the supermodularity of e with respect to ε and z the productivity effects of skill are larger with higher effort.

effort $\bar{\varepsilon} > 1$ and receive a wage $w_a(z) = c_a e(\bar{\varepsilon}, z)$. This figure also illustrates nicely why a sorting equilibrium implies that the effective wage c_n paid by non-adopting firms has to be larger than the effective wage c_a in adopting firms. If non-adopting firms paid the same effective wage as adopting firms, $w_n = c_a \bar{z}_n$, then the wage for workers with above-average productivity $z > \bar{z}_n$ would be discretely lower in non-adopting firms than in adopting firms [$c_a \bar{z}_n < c_a e(\bar{\varepsilon}, z)$ for all $z \in (\bar{z}_n, \tilde{z})$]. Consequently, this could not be a sorting equilibrium. Instead, non-adopting firms have to pay a premium on the effective wage rate, $c_n > c_a$, in order to compensate their above-average workers for pooling them with below-average worker, so that $c_n \bar{z}_n = c_a e(\bar{\varepsilon}, \tilde{z})$. Put differently, adopting firms are able to obtain a rent from their workers in the form of a lower effective wage rate. This rent comes from allowing more productive workers to avoid being pooled with less productive workers. In this sense, workers share rents on their skills with firms rather than the other way around as is common in much of the literature on "fair wages."

Lazear (2000) presents an influential case study that describes the effects of the introduction of monitoring and incentive practices on productivity measures for the Safelite Glass Corporation in the 1990s. All of our corollaries above are mirrored in his findings: These management practices increased the average levels of output per worker [our $\bar{e}(\varepsilon, z)$] by about 40 percent, where approximately half of this increase in due to the incentive effect (higher ε in adopting firms), and half is due to the selection effect (higher z in adopting firms). In addition, pay-per-worker went up (our $\bar{w}$), but less than productivity $\bar{e}$, so that workers are sharing rents with firms. Finally, Lazear also finds an increase in the variance in output that is also present in our model due to the supermodularity of the e-function. Similar findings are also echoed by Bloom and Van Reenen (2011) and Lazear and Oyer (2013).

2.2.5 General Equilibrium

For completeness we derive aggregate statistics that will be important in the welfare calculations below. With profits driven down to zero, aggregate income consists of labor income and compensation for managers. Since management is used as our numéraire, their compensation is normalized to one:

$$E \equiv L \int_{\underline{z}}^{\infty} w(z)\, dG(z) + M = L \left[c_n \int_{\underline{z}}^{\tilde{z}} z dG(z) + c_a \int_{\tilde{z}}^{\infty} e(\bar{\varepsilon}, z)\, dG(z) \right] + M. \qquad (28)$$

With CES demand, a constant fraction of revenues goes to fixed costs, and variable factors receive the remaining (constant) fraction. In our framework, this implies that $E = \sigma M$, and thus

$$E = \frac{\sigma}{\sigma - 1} L \left\{ c_n G(\tilde{z}) \bar{z}_n(\tilde{z}) + c_a \left[1 - G(\tilde{z})\right] \bar{e}_a(\bar{\varepsilon}, \tilde{z}) \right\} = \sigma M. \tag{29}$$

With E determined, and A pinned down by (20), the price index P can be derived easily from (7).

3 Welfare Implications of Labor Market Imperfections

This section analyzes the welfare implications of the information advantage of adopting firms with structured management practices. We consider the problem of a social planner that seeks to maximize the output of bundles of the composite differentiated good subject to the technological and resource constraints facing agents in the model, plus the behavioral constraint regarding effort provision. The social planner chooses the number of varieties to produce, whether or not to reward effort, and the level of output of each variety. This maximization problem indirectly requires the planner to choose which workers to assign to each type of firm.

We make three preliminary arguments to simplify the exposition of the planner's problem. First, with respect to the level of output chosen of each variety, there are only two levels of output that the planner would assign to the produced varieties, x_n^* and x_a^*, for non-adopting and adopting firms, respectively. This follows immediately from the symmetry of technology within types and the existence of diminishing marginal utility to the consumption of any given variety.

Second, the planner can choose what type of worker to allocate to which type of firm, and can elicit effort from workers, despite the information constraint. For this, the planner first chooses the share of managers allocated to adopting and non-adopting firms, M_a and M_n. Given f_a and f_n, this essentially fixes the mass of the two types of firms, n_a^* and n_n^*. Workers can only work for type-a firms if they produce at least $e(\bar{\varepsilon}, \tilde{z}^*)$, otherwise they earn nothing. The remaining workers are allocated to type-n firms. This induces the desired sorting $\tilde{z}^*$. Effort can be elicited in type-a firms by allowing firms to compete for workers.

Third, it is straightforward to see that the optimal allocation of workers to firms involves $z > \tilde{z}^*$ being allocated to adopting firms and $z < \tilde{z}^*$ being allocated to non-adopting firms. This can be established by contradiction. Consider two workers 1 and 2 with $z_1 > z_2$. Suppose that the planner optimally assigned worker 1 to a non-adopting firm, eliciting effort $\varepsilon = 1$ and worker 2 to an adopting firm, eliciting effort $\bar{\varepsilon}$. Given the supermodularity of the function e in effort and skill, output could be increased by having workers swap jobs.

Given these observations, the planner's problem is to choose the measures of varieties, n_n^*

and n_a^*; the quantities of output per variety, x_n^* and x_a^*; and an allocation of labor to firms, $\tilde{z}^*$, to maximize

$$\max_{x_a^*, x_n^*, n_n^*, n_a^*, \tilde{z}^*} \left(n_a^* \left(x_a^*\right)^{\frac{\sigma-1}{\sigma}} + n_n^* \left(x_n^*\right)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}$$

subject to the managerial resource constraint, the effective labor resource constraint for a varieties, and the effective labor resource constraint for n varieties. These constraints are given by

$$\begin{aligned} n_a^* f_a + n_n^* f_n &\leq M, \\ n_a^* x_a^* &\leq L \int_{\tilde{z}^*}^{\infty} e\left(\bar{\varepsilon}, z\right) dG\left(z\right), \text{ and} \\ n_n^* x_n^* &\leq L \int_0^{\tilde{z}^*} z dG\left(z\right), \end{aligned}$$

and have Lagrangian multipliers given by λ_M, λ_a, and λ_n, respectively. The proof of the following proposition can be found in the appendix:

Proposition 3 (Efficiency) *(i) The market equilibrium leads to a socially inefficient allocation of labor across firm types. Compared to the social optimum, employment in adopting firms is too high in the sorting equilibrium, i.e.* $\tilde{z}^* > \tilde{z}$. *(ii) Firm sizes in the market equilibrium are socially optimal, i.e.* $x_a^*/x_n^* = x_a/x_n$. *(iii) The excess employment in adopting firms manifest itself as too many firms adopting structured management practices and too little aggregate variety, i.e.* $n_a^* < n_a$ *and* $n_a^* + n_n^* > n_a + n_n$.

Proof. See appendix 8.3. ■

We sketch the proof to the three parts of the proposition in turn. First, the misallocation of labor is due to the fact that workers in non-adopting firms receive a wage based on the average productivity of their entire labor pool, and not on the individual productivity of the marginal worker in that pool. This can be seen by comparing the relative shadow values of labor and the allocation of labor to firms in the solution to the planner's problem and in the market equilibrium. These are given by

$$\frac{\lambda_a}{\lambda_n} = \frac{\tilde{z}^*}{e(\bar{\varepsilon}, \tilde{z}^*)} = \Phi \tag{30}$$

in the planner's problem and by

$$\frac{c_a}{c_n} = \frac{\overline{z}_n(\tilde{z})}{e(\bar{\varepsilon}, \tilde{z})} = \Phi$$

in the market equilibrium. It is clear from these equations that the relative cost of labor is not distorted $[\lambda_a/\lambda_n = c_a/c_n]$ and so the sizes of firms in the market equilibrium is optimal (part ii of the proposition). Instead, the problem lies with the optimal cutoff. Because the value of the marginal worker in the market equilibrium exceeds the average value of a worker, i.e. $\tilde{z} > \overline{z}_n(\tilde{z})$, there is an excess allocation of labor to adopting firms (part i of the proposition). From the resource constraints, it immediately follows that there is excessive entry of adopting firms (excess adoption) and because they have higher fixed costs than non-adopting firms there are too few varieties in equilibrium (part iii of the proposition).

For the purpose of intuition, it is useful to see the cause of the misallocation of labor (part i) from the perspective of the marginal worker. Because the productivity of the average worker in the non-adopting labor pool is lower than the productivity of the marginal worker $[\bar{z}_n(\zeta) < \zeta \; \forall \; \zeta \in (0, \infty)]$, relative wages in non-adopting firms are too low compared to the relative productivity of the marginal worker in the two types of firms. As a consequence, fewer workers self-select into non-adopting firms than what is socially desirable and the mispricing of labor manifests itself in too little employment in non-adopting firms. Put differently, while adopting firms have desirable characteristics (larger economies of scale, higher labor productivity, higher wages, see corollaries 2 to 4), from a social perspective they collectively consume too many resources.

The second part of the proposition states that the misallocation of labor does not show up as excessively large or small production of individual varieties. This has the flavor of the original Dixit and Stiglitz (1977) result. In the market equilibrium, profits are driven to zero for both types a and n firms so that the relative marginal cost, c_a/c_n, is pinned down by their relative fixed cost, f_a/f_n. As free entry prevents the cost of the effective cost of labor from being affected by the excessive supply of effective labor to adopting firms, the misallocation manifests itself as excessive entry of adopting firms. This is the third part of the proposition. Essentially, demand for workers must expand through entry to absorb the supply of workers who seek to avoid pooling with less able workers. Because adopting firms with structured management practices use more managerial resources than non-adopting firms, the equilibrium features too few varieties relative to the social optimum.

In equilibrium, the firms face the socially optimal effective cost of labor but workers are willing to supply excessively large amounts of labor at that price in order to have the privilege to have their output observed and so to avoid the artificially low wage at non-adopting firms. It is this excessive supply of effective units of labor available to adopting firms that induces their inefficient entry and starves non-adopting firms of managerial inputs.

[FIGURE 3 here]

Figure 3 illustrates the misallocation of labor graphically. It is based on our Figure 2 and shows the market valuation and the social valuation of workers with different skills in a specific allocation $\tilde{z}$. The market valuation of a worker with skill z in adopting firms is given by $w(z) = c_a e(\bar{\varepsilon}, z)$ (the thick black line). Workers in adopting firms are correctly valued so the social valuation is equal to the market valuation. The market valuation of a worker with skill z in non-adopting firms is independent of z and given by $w_n = c_n \bar{z}_n(\tilde{z})$ (the thick blue line). The intersection of the two gives the market allocation $\tilde{z}$. However, the social valuation of workers in non-adopting firms is not aligned with the market valuation. From a social point of view, workers with skill z in non-adopting firms should be valued at $c_n z$ (the thick red line). The socially optimal allocation $\tilde{z}^*$ is then given by the intersection of the two social valuations $c_a e(\bar{\varepsilon}, z)$ and $c_n z$. All workers with skill levels $z < \tilde{z}$ are optimally allocated. They work for non-adopting firms, and their social valuation in non-adopting firms is higher than the social valuation in adopting firms: $c_n z > c_a e(\bar{\varepsilon}, z)$. All workers with skill levels $z > \tilde{z}^*$ are also optimally allocated. They work for adopting firms ($z > \tilde{z}$), and their social valuation in adopting firms is higher: $c_a e(\bar{\varepsilon}, z) > c_n z$. It is workers with skill levels $z \in (\tilde{z}, \tilde{z}^*)$ that are misallocated. These workers work for adopting firms ($z > \tilde{z}$), but their social valuation is higher in non-adopting firms $[c_n z > c_a e(\bar{\varepsilon}, z)]$. Thus, it is only a subset of workers that are misallocated, and we can now turn to a straightforward option to correct this misallocation.

The misallocation of labor creates an incentive to subsidize employment in non-adopting firms. Since $\tilde{z} < \tilde{z}^*$, it follows that a reallocation of labor from adopting to non-adopting firms (an increase in $\tilde{z}$) increases aggregate output. To see how such a subsidy can increase welfare assume that the government can subsidize employment in non-adopting firms and finance this subsidy with a non-distorting per capita tax on income. This changes equations (20) (for $j = n$) and (28):

$$A \left[c_n (1-s)\right]^{1-\sigma} = f_n, \tag{31}$$

$$E = \left[c_n (1-s) \int_{\underline{z}}^{\tilde{z}} z dG(z) + c_a \int_{\tilde{z}}^{\infty} e(\bar{\varepsilon}, z) dG(z)\right] L + M = \sigma M, \tag{32}$$

where s is the subsidy rate, $c_n (1-s)$ are after subsidy effective labor costs in non-adopting firms, and $s c_n \bar{z}_n(\tilde{z}) G(\tilde{z}) L$ is the total subsidy paid.

The allocation of labor in (23) changes to

$$\frac{\bar{z}_n(\tilde{z})}{e(\bar{\varepsilon}, \tilde{z})} = \Phi (1-s). \tag{33}$$

Because $\bar{z}_n(\tilde{z}) / e(\bar{\varepsilon}, \tilde{z})$ is decreasing in $\tilde{z}$ by (17), an increase in s increases $\tilde{z}$ and the subsidy is effective in raising employment in non-adopting firms.

Because the effective real wage in non-adopting firms, $c_n (1-s) P^{-1}$, continues to be pinned down by the free entry condition, relative marginal costs are unaffected by the subsidy so that the per variety output remains at their social optimum. Thus, the subsidy affects welfare only through the allocation of labor $\tilde{z}$, and the optimal subsidy s^* can be calculated as the percentage difference between the marginal and the average productivity in the non-adopting labor pool, evaluated at the optimal allocation:

$$s^* = \frac{\tilde{z}^* - \bar{z}_n(\tilde{z}^*)}{\tilde{z}^*}. \tag{34}$$

Proposition 4 (Subsidy) *There exists an optimal subsidy rate on employment in non-adopting firms $s^* \in (0, 1)$ that corrects the misallocation of labor and reaches the social optimum, so that $\tilde{z}(s^*) = \tilde{z}^*$.*

In this setup, the market imperfections in the labor market create an incentive to subsidize small firms without structured management practices that exhibit low productivity. The inability of these firms to identify and to reward talent results in their having a high effective cost of labor. They can survive only because their overhead costs are low enough to stay in business so they coexist with adopters, but in order for them to expand their aggregate use of labor, they would have to raise their wages for all of their workers including the inframarginal workers with relatively low productivity. As a consequence, the employment share of small, non-adopting, low productivity firms is too small compared to the social optimum, and a subsidy on employment in these firms can be welfare improving.[16]

The corrective subsidy raises welfare because it induces a reallocation of workers from adopting firms to non-adopting firms through exit and entry. The workers that switch pools in response to the subsidy are more skilled than the workers previously working for non-adopting firms. As a consequence, average productivity rises in subsidized firms. This implication is in line with recent evidence that shows that subsidies targeted to small firms do indeed raise value added per worker in treated firms as implied by our model (Lombardi et al., 2018).

In the Extensions, we show that under the alternative assumption of a fixed number of firms the misallocation of labor manifests itself as adopting firms that are too large and inefficient relative to the social optimum. In that case the corrective subsidy works by inducing non-adopting firms to expand while inducing adopting firms to downsize.

4 Open Economy

The existence of a market failure associated with incomplete adoption of management practices means that the analysis of policy must occur in the world of the second best. One such

[16] The intuition is similar to that pointed out in Greenwald and Stiglitz (1986) in a different context.

environment in which issues of the second best arise is that of international trade. To that end, we now consider international trade in an open economy setting. We will first discuss the implications of international trade in our theoretical framework and then address how these results can be used to quantify the empirical importance of the mechanisms discussed here.

There are two identical countries and international trade is costly in two dimensions: Entering a foreign market creates fixed costs of exporting f^x, and shipping goods to foreign locations is subject to variable (iceberg) trade costs $\tau > 1$. The two types of costs are identical for all firms. Export participation depends on the size of trade costs relative to type-specific fixed costs. We focus on the case where only adopting firms engage in exporting and assume that $f_n < \tau^{\sigma-1} f^x < f_a$ (Melitz condition).[17]

Free entry ensures that profits of domestic non-adopting firms, given by (10) for $j = n$, and exporting adopting firms, $\pi_a^t = A\left(1 + \tau^{1-\sigma}\right) c_a^{1-\sigma} - f_a - f^x$, are driven down to zero. By combining the new zero profit conditions with the sorting condition in (16) we can determine the market allocation of labor in the open economy:

$$\frac{\bar{z}_n\left(\tilde{z}_t\right)}{e\left(\bar{\varepsilon}, \tilde{z}_t\right)} = \Phi_t, \tag{35}$$

where $\Phi_t^{\sigma-1} \equiv f_n\left(1 + \tau^{1-\sigma}\right) / \left(f_a + f^x\right)$ and $d \ln \Phi_t / d \ln \tau = -\left(1 + \tau^{\sigma-1}\right)^{-1} < 0$. Since $\Phi_t \in (0, 1)$ and, hence, $\underline{z} < \tilde{z}_t < \infty$, both types of firms coexist and proposition 2 continues to hold.[18]

In this setup, we can derive three major propositions regarding the effects of trade liberalization defined as a reduction in variable trade costs τ:

Proposition 5 (Employment Shares) *Trade raises the employment share in adopting firms* $1 - G\left(\tilde{z}\right)$.

Proposition 6 (Real Wages) *Trade raises real wages in adopting firms/for high-skill workers and lowers real wages in non-adopting firms/for low-skill workers.*

Proposition 7 (Welfare Effects) *The welfare effects of trade liberalization depend on two effects: A direct effect of the reduction in trade costs that tends to raise welfare, and an indirect effect of the reallocation of labor that tends to lower welfare. The welfare reducing effect dominates when trade costs are high and non-adopting firms have a high share in employment.*

[17] We refer to this case as the Melitz (2003) condition, because trade costs are higher than manufacturing fixed costs in non-adopting firms and the equilibrium exhibits sorting into exporting where only adopting firms export.

[18] If $\tau^{\sigma-1} f^x < f_l < f_h$, all firms export, and $\Phi_t^{\sigma-1} \equiv \left(f_l + f^x\right) / \left(f_h + f^x\right)$ with very similar implications.

Proposition 5 is straightforward. Only exporting firms benefit from a reduction in variable trade costs, and since only adopting firms export, they expand and their employment share rises. From (35) and (17) we have $d\ln\tilde{z}_t/d\ln\tau > 0$.[19]

For proposition 6 we express real wages $w_j^r \equiv w_j/P$ as

$$\ln w_n^r = \ln\xi + \ln\bar{z}_n(\tilde{z}_t) \qquad \ln w_a^r(z) = \ln\xi + \ln e(\bar{\varepsilon}, z) + \ln\Phi_t \tag{36}$$

where $\xi \equiv (1 - 1/\sigma)(M/f_n)^{\frac{1}{\sigma-1}} = c_n/P$ is a constant and equal to the effective real wage in non-adopting firms in a free entry equilibrium.

A closer evaluation of these two equations reveals two relevant facts: First, $w_a^R(z)$ depends directly on Φ_t and, thus, negatively on trade costs τ. Therefore, the wage of workers in adopting firms rises because these workers work in exporting firms that benefit directly from a reduction in trade costs. In fact, since $\ln(c_a/P) = \ln w_a^r(z) - \ln e(\bar{\varepsilon}, z) = \ln\xi + \ln\Phi_t$, their *effective* real wage (or real piece rate) rises in Φ_t. Second, w_n^R does not depend on Φ_t. Consequently, workers in non-adopting firms are not directly affected by changes in trade costs, and their *effective* real wages $\ln(c_n/P) = \ln w_n^r - \ln\bar{z}_n(\tilde{z}_t) = \ln\xi$ remain unaffected. However, their (non-effective) real wages w_n^r fall because their most productive workers leave to the adopting firms' labor pool, leading to a decline in the average worker productivity $\bar{z}_n$. Hence, high-skilled workers in adopting firms see their real wages rise, while low-skilled workers in non-adopting firms experience a decline in their real wages.

It is worth noting that the average worker skill in both labor pools falls when the most productive workers in the non-adopting firms' pool move to the adopting firms' pool (where they are the least skilled workers). However, this change in the composition of the workforce has very different effects in the two pools: In adopting exporting firms, workers are directly affected by any cost shocks transmitted through international markets, but are unaffected by changes in the composition of the workforce.[20] In non-adopting non-exporting firms, real wages are not directly affected by any trade cost shocks, but instead respond to changes in the composition of the workforce. This latter prediction has a very distinct implication: Workers continuously employed by small non-adopting non-exporting firms with low productivity will experience a reduction in their real wages in response to a reduction in trade costs.

The composition effect on wages can be identified empirically as a team effect. If a worker or a set of workers with above average skills leave non-adopting firms they trigger a team effect on the wages of the remaining workers. This is not the case in firms that have adopted structured management practices: If a worker (or a set of workers) with below average skills

[19] Mathematically, $d\ln\tilde{z}_t/d\ln\tau = \left(1 + \tau^{\sigma-1}\right)^{-1}\Gamma^{-1} > 0$, where Γ is defined by (17).

[20] The term $e(\bar{\varepsilon}, z)$ is a z-specific productivity term that has no compositional component.

leaves the pool of workers for adopting firms, there is no team effect on remaining workers. Given the insights from the economics of management literature in Bloom et al. (2014), this is an important and very intuitive distinction: Compositional effects are more important for teams if workers are not incentivized, as is the case in our non-adopting firms. In firms with best practice management techniques and proper incentives in place (our large, exporting, high productivity adopting firms) these compositional effects do not matter.

Finally, let us turn to welfare. Given (2), (7), (20), (22), and (29), and defining $\Xi \equiv \xi L\sigma/(\sigma-1)$, aggregate welfare $W = E/P$ can be expressed as a function of $\tilde{z}_t$ and Φ_t,

$$W(\tilde{z}_t, \Phi_t) = \Xi\left(\int_{\underline{z}}^{\tilde{z}_t} z dG(z) + \Phi_t \int_{\tilde{z}_t}^{\infty} e(\bar{\varepsilon}, z)\, dG(z)\right), \tag{37}$$

and the change in welfare in response to a reduction in variable trade costs can be decomposed into two effects:

$$\frac{d\ln W}{d\ln\tau} = -\frac{1-\lambda(\tilde{z}_t)}{1+\tau^{\sigma-1}} + \frac{\lambda(\tilde{z}_t)\,\theta(\tilde{z}_t)}{(1+\tau^{\sigma-1})\,\Gamma}, \tag{38}$$

where $\lambda(\tilde{z}_t) \equiv c_n \int_{\underline{z}}^{\tilde{z}_t} zdG(z) / \int_{\underline{z}}^{\infty} w(z)\, dG(z) \in [0,1]$ is defined as the share of labor in non-adopting firms, and $\Gamma > 0$ is defined in (17).

The first effect is the direct effect of τ on welfare, $(\partial \ln W/\partial \ln \Phi_t)(d\ln\Phi_t/d\ln\tau)$. It is negative, indicating that a reduction in variable trade costs tends to raise welfare directly by lowering costs. The size of this effect depends on the share of labor in adopting firms, $1-\lambda(\tilde{z}_t)$, and on the size of the cost reduction, $1/(1+\tau^{\sigma-1})$.

The second effect, $(\partial \ln W/\partial \ln \tilde{z}_t)(d\ln\tilde{z}_t/d\ln\tau)$, is the indirect effect working through the reallocation of labor $\tilde{z}_t$. Here, $\theta(\tilde{z}_t)$ is a measure of the extent of the deviation of the market equilibrium from the social optimum:[21]

$$\theta(\tilde{z}_t) \equiv \frac{\tilde{z}_t - \bar{z}_n(\tilde{z}_t)}{\bar{z}_n(\tilde{z}_t)} \frac{g(\tilde{z}_t)\,\tilde{z}_t}{G(\tilde{z}_t)} > 0. \tag{39}$$

This indirect effect is positive, indicating that a reduction in variable trade costs tends to lower welfare. This second effect exists because the market equilibrium is not socially optimal $[\theta(\tilde{z}_t) > 0]$. There is already too much labor in adopting firms so that $\partial \ln W/\partial \ln \tilde{z}_t > 0$ (see part (i) of proposition 3) and an additional boost of employment in adopting firms through trade liberalization exacerbates the misallocation and has a negative first order effect on welfare. This effect depends on the share of labor in non-adopting firms, $\lambda(\tilde{z}_t)$, on the size of the misallocation, $\theta(\tilde{z}_t)$, as well as on the extend of the reallocation of labor $d\ln\tilde{z}_t/d\ln\Phi_t$.

[21] By combining (23) and (30) we obtain $\bar{z}_n(\tilde{z})/\tilde{z} = [\tilde{z}^*/e(\bar{\varepsilon}, \tilde{z}^*)] / [\tilde{z}/e(\bar{\varepsilon}, \tilde{z})]$. Hence, the relative difference between $\tilde{z}$ and $\bar{z}_n(\tilde{z})$ reflects the relative difference between $\tilde{z}$ and $\tilde{z}^*$.

It is important to note that this reallocation of labor from non-adopting to adopting firms constitutes a *negative* welfare effect despite the fact that workers in adopting firms exert more effort. This is surprising since the effort effect increases average worker productivity in the economy. But because of the pooling in non-adopting firms, the differences in productivity based on effort $[e(\bar{\varepsilon}, \tilde{z}_t) > \tilde{z}_t]$ are more than offset by the differences in productivity based on sorting: $\Phi_t e(\bar{\varepsilon}, \tilde{z}_t) < \tilde{z}_t$. Thus, the effort effects are only second order in the welfare calculations.

There are two explanations for the ambiguity of the aggregate welfare effects: (i) An income-based explanation and (ii) an efficiency-based explanation. For the income-based explanation remember that welfare is a function of the average real wage in the economy. However, the real wages in the two types of firms are moving in different directions: They increase in adopting firms and decrease in non-adopting firms. The aggregate effect depends on the share of the two types of firms in employment. That is why real income rises when adopting firms have a large employment share in the economy.

The efficiency-based explanation highlights how trade liberalization affects labor productivity. On the one hand, the reduction in variable trade costs lowers labor costs of exporting and this tends to increase labor productivity (the direct effect). On the other hand, marginal productivities of labor are not equalized across firm types because of the labor market inefficiency. By reallocating the marginal worker to adopting firms, aggregate labor productivity falls (the indirect effect). The magnitude of the indirect effect depends on the extent of the misallocation, i.e. on the difference between $\bar{z}_n(\tilde{z}_t)$ and $\tilde{z}_t$. And since this difference is increasing in $\tilde{z}_t$, the indirect effect with its negative welfare implication is larger for larger employment shares of non-adopting firms.

There is one important case where the welfare effect of trade liberalization is clearly positive: When the optimal subsidy s^* is in place. With the optimal subsidy in place, marginal productivities of labor are equalized across firm types, and the negative welfare effect disappears. This is a straightforward application of the envelop theorem: The optimal subsidy is chosen so that $\partial \ln W / \partial \ln \tilde{z}_t(s^*) = 0$. In this case, only the direct effect remains and the welfare effects of trade liberalization are clearly positive:

Corollary 5 (Welfare with Subsidy) *With the optimal subsidy s^* in place, the welfare effects of trade are unambiguously positive.*

5 Extensions

In this section we want to address some extensions of our basic framework that will serve as robustness checks of our main result with respect to specific assumptions of our framework.

5.1 Learning and Secondary Markets for Workers

In our main framework we assume that firms that do not adopt structured management practices do not monitor workers and, therefore, do not (ever) know the true productivity of workers. This is true in a static environment where workers cannot switch firms. But if non-monitoring non-adopting firms can observe the wages that are being paid to workers in the adopting firms, they can deduce their skill levels from this observation. Thus, in a dynamic framework they can learn about worker productivity because they can observe a worker's work history, even if they cannot or do not assess, screen or monitor workers individually. In this extension we study how the possibility of learning about workers' skills and the existence of secondary markets (after learning) affect the equilibrium and the distortion.

In order to capture the dynamic nature of the learning process we study the market equilibrium in two periods. Period 1 is essentially as in section 2.2.3: Firms that invest in structured management practices assess their workers' true skills, monitor their individual output, offer a piece rate and elicit extra effort from their workers. Firms that do not invest in structured management practices only know the distribution of skills in their labor pool, do not monitor their workers' individual output and pay them a salary that is based on the average skill in their labor pool. Consequently, their workers only put in the minimum effort required. Sorting leads to a cutoff $\tilde{z}_1$, where the index 1 refers to period 1, so that workers with skills in the range $(\underline{z}, \tilde{z}_1)$ work for non-adopting firms and receive a wage of $w_{n1} = c_{n1}\bar{z}_n(\tilde{z}_1)$, while workers with skills in the range $(\tilde{z}_1, \infty)$ work for firms that have adopted structured management practices and receive $w_{a1} = c_{a1}e(\bar{\varepsilon}, z)$.

Period 2 is when the learning takes place and secondary markets open: In period 2 firms can observe the work history of workers and the distribution of wages in period 1. However, this does not mean that all firms know the true productivities of all workers. Firms can only observe what is actually observable: For workers with skills in the range $(\underline{z}, \tilde{z}_1)$ this is still very little: Since these workers have not been assessed by their employers in period 1, no new information is available. This is different for workers with skills in the range $(\tilde{z}_1, \infty)$: These workers have worked in period 1 for firms with structured management practices, have been screened and monitored, and have received a wage that was based on their true productivity. This wage and the wage distribution are now observable, and it allows other firms to deduce the true skills z of these workers without having to screen them again. As a consequence, firms that have not adopted structured management practices (and have not paid the higher fixed costs f_a) can now offer wages that are based on the true skills of individual workers for workers with skills $z > \tilde{z}_1$. They cannot, however, reward effort ε because they are still unable to monitor the individual output of a worker. Hence, workers continue to put in only minimal effort $\varepsilon = 1$ in non-adopting firms.

The wage that non-adopting firms can offer to workers with skills $z > \tilde{z}_1$ is now $w_n(z) = c_n e(1, z) = c_n z$. Thus, these (previously screened) workers are no longer pooled with lesser skilled workers, and they can now benefit from the higher effective wages $c_n > c_a$ in non-adopting firms. This makes it attractive for workers in the lower part of the skill range $(\tilde{z}_1, \infty)$ to move from adopting firms to non-adopting firms. In fact, the marginal worker $\tilde{z}_{nl}$ in (16) who is indifferent between the two labor pools in the scenario without learning strictly prefers to work for non-adopting firms when learning is possible: $w_a(\tilde{z}_{nl}) = c_a e(\bar{\varepsilon}, \tilde{z}_{nl}) = c_n \bar{z}_n(\tilde{z}_{nl}) < c_n \tilde{z}_{nl}$, where nl stands for *no learning*. Workers in the upper part of the skill distribution still prefer to work for adopting firms because of the productivity effect of higher effort. Since a worker's productivity is supermodular in effort and skill, the effort effect dominates for highly skilled workers, while the effective wage effect dominates for lower skilled workers. These relationships are also visualized in our Figure 3 where the black line and the red line intercept at some $z > \tilde{z}$.

The new sorting condition in period 2 is now $c_{a2} e(\bar{\varepsilon}, \tilde{z}_2) = c_{n2} \tilde{z}_2$, or, using (21) and (22),

$$\frac{\tilde{z}_2}{e(\bar{\varepsilon}, \tilde{z}_2)} = \Phi. \tag{40}$$

When comparing this sorting outcome to our original sorting in (23) without learning it is straightforward that $\tilde{z}_2 > \tilde{z}_{nl}$. Thus, with learning, employment in period 2 is larger in non-adopting firms and smaller in adopting firms compared to an equilibrium without learning.

In fact, this sorting condition in period 2 is identical to the condition for a socially optimal sorting in (30). The social optimum is reached because there is no more information asymmetry at the margin: The productivity of the marginal worker is known to all firms. There is still pooling for workers who have never worked for adopting firms $z < \tilde{z}_1$ and whose productivity is still unknown. But the marginal worker $\tilde{z}_2$ has been screened by adopting firms in period 1 and this information is becoming public knowledge in period 2, so there is no more information asymmetry at the margin.

In period 2 we have three groups of workers: Workers with skills $z < \tilde{z}_1$ have never been screened, receive a salary based on the average pool productivity and provide no effort. Workers with skills $z \in (\tilde{z}_1, \tilde{z}_2)$ have been screened in period 1 in adopting firms. They move to non-adopting firms in period 2 and receive a salary based on their true productivity. They provide only minimum effort in period 2 since non-adopting firms do not reward effort. Finally, workers with skills $z > \tilde{z}_2$ work for adopting firms in both periods. They receive a piece rate in both periods and provide maximum effort in both periods.

Workers are forward looking and internalize the benefit from being screened in period 1 in their sorting decisions. Thus, they compare in period 1 the salary in non-adopting firms in both periods with their income in adopting firms in period 1 and in non-adopting firms in period 2. The new sorting condition is then: $c_{n1}\bar{z}_n(\tilde{z}_1) + c_{n2}\bar{z}_n(\tilde{z}_1) = c_{a1}e(\bar{\varepsilon}, \tilde{z}_1) + c_{n2}\tilde{z}_1$. Using (21) and (22), the sorting outcome in period 1 can be written as

$$[1 - \vartheta(\tilde{z}_1)] \frac{\bar{z}_n(\tilde{z}_1)}{e(\bar{\varepsilon}, \tilde{z}_1)} = \Phi \tag{41}$$

where $\vartheta(\tilde{z}_1) \equiv \frac{c_{n2}}{c_{n1}} \frac{\tilde{z}_1 - \bar{z}_n(\tilde{z}_1)}{\bar{z}_n(\tilde{z}_1)} = 1 - \frac{c_{a1}e(\bar{\varepsilon}, \tilde{z}_1)}{c_{n1}\bar{z}_n(\tilde{z}_1)} \in (0, 1)$.

When comparing (41) with learning to (23) without learning we can determine that employment in non-adopting firms is lower in period 1 with learning than in the scenario without learning: $\tilde{z}_1 < \tilde{z}_{nl}$.[22] In fact, since $\vartheta(\tilde{z}_1) \propto [\tilde{z}_1 - \bar{z}_n(\tilde{z}_1)] / \bar{z}_n(\tilde{z}_1)$, this difference is increasing in the inefficiency. We can summarize these findings in the following proposition:

Proposition 8 (Sorting with Learning) *In a dynamic scenario where non-adopting firms can learn about the productivity of workers from their work history in adopting firms, the inefficiency is even larger in period* 1 *(before learning takes place) than in a scenario without learning, but disappears in period* 2 *(after learning has taken place).*

The inefficiency disappears in period 2 because non-adopting firms can learn the true productivity of workers in adopting firms in period 1. Hence, there is no more information asymmetry at the margin. Instead, our inefficiency now manifests itself in excessive labor market sorting, and too much resources spent on management practices, before learning has taken place.

5.2 Exogenously Better Management

In section 3, we showed in our benchmark model with endogenous adoption of structured management practices that the distortion created by imperfect information manifested itself as excess worker sorting into the labor pool available to adopters. The resulting labor supply misallocation meant that there were too many firms adopting structured management technologies from the perspective of a social planner but free entry had the implication that both adopting and non-adopting firms were individually of optimal size.

In this section, we consider the case in which the measure of each type of firm, n_i where $i \in \{a, n\}$, is exogenously fixed.[23] We will see that the distortion created by imperfect information then manifests itself as adopting firms that are too large relative to the social

[22] Note that $\bar{z}_n(\tilde{z}_{nl}) / e(\bar{\varepsilon}, \tilde{z}_{nl}) = \Phi < \bar{z}_n(\tilde{z}_1) / e(\bar{\varepsilon}, \tilde{z}_1)$.

[23] This could be due to a fixed measure of managers that had the ability to screen workers or due to a single free entry condition that yielded firms that can screen with probability $n_a/(n_a + n_n)$.

optimum and non-adopting firms that are too small. In this case, from a social point of view it would be better if bigger (adopting) firms were smaller.

Without free entry, there is no zero profit condition and the size of firms is determined by the labor market clearing condition:

$$x_i = \tilde{L}_j / n_j. \tag{42}$$

Consequently, relative sizes are given by

$$\frac{x_n}{x_a} = \frac{n_a}{n_n} \frac{G(\tilde{z})\,\bar{z}_n(\tilde{z})}{[1 - G(\tilde{z})]\,\bar{e}_a(\bar{\varepsilon}, \tilde{z})} \tag{43}$$

Given (8) and using ζ to represent the ability of the cutoff worker, we show that the relative labor market equilibrium can be written

$$\frac{c_a}{c_n} = \left(\frac{n_a}{n_n} \frac{\int_{\underline{z}}^{\zeta} \zeta dG(\zeta)}{\int_{\zeta}^{\infty} e(\bar{\varepsilon}, \zeta) dG(\zeta)} \right)^{\frac{1}{\sigma}}. \tag{44}$$

Equation (44) is depicted in Figure 5 as an increasing locus in $(c_a/c_n, \zeta)$ space ("Labor Market Equilibrium"). It is increasing because a higher cutoff level ζ implies a higher relative supply in the type-n labor pool and requires in equilibrium a lower relative effective wage in that pool to boost relative labor demand.

[Figure 5 here]

As in the benchmark model, the "Sorting Condition" is given by $c_a/c_n = \bar{z}_n(\zeta)/e(\bar{\varepsilon}, \zeta)$ and is decreasing by condition (17). In the market equilibrium, the cutoff skill is given by $\tilde{z}$ and the relative effective wage facing producers is $c_a/c_n = \bar{z}_n(\tilde{z})/e(\bar{\varepsilon}, \tilde{z})$.

Also as in the benchmark model, the "Social Planner" would allocate labor so that the relative shadow price of labor in each of the labor pools equates the marginal social return of the cutoff worker in the two pools, i.e. $c_a^*/c_n^* = \zeta/e(\bar{\varepsilon}, \zeta)$. This condition is shown as a decreasing curve in figure 5. Crucially, this locus lies everywhere above the "Sorting Condition" curve [since $\zeta > \bar{z}_n(\zeta)$]. It follows immediately from the figure that too much labor is allocated to the adopting labor pool, i.e. $\tilde{z} < \tilde{z}^*$, and the effective cost of labor at adopting firms is too low relative to non-adopting firms, i.e. $c_a/c_n(\tilde{z}) < c_a^*/c_n^*(\tilde{z}^*)$. The latter result contrasts with the benchmark model where c_a/c_n is flat at its optimal level due to the technologically imposed free entry condition. With exogenous firm heterogeneity, a low equilibrium $\tilde{z}$ reduces c_a/c_n in order to make adopting firms grow to absorb the excess labor and to induce non-adopting firms to shrink to conserve labor. The following proposition follows immediately from figure 5.

Proposition 9 (Optimal Firm Size) *Without free entry, adopting firms are relatively too large compared to the social optimum.*

In their effort to avoid pooling with less able workers, workers at or just above the cutoff ability choose to share rents on their ability with their employers and this has the implication that their employers grow excessively large relative to the social optimum.[24]

6 Quantifying the Welfare Effects of Sorting Distortions

The distortion described here is driven by the difference between the average and the marginal productivity of workers in non-adopting firms. Given our sorting result, this difference is increasing in the size of the labor pool in non-adopting firms: $d\left(\tilde{z}-\bar{z}_n\left(\tilde{z}\right)\right)/d\tilde{z}>0$. Hence, if only few firms have adopted management practices and $\tilde{z}$ is large, the distortion is also large, and if adoption is widely spread and $\tilde{z}$ is small, the distortion is also small. This is in line with our welfare result in proposition 7 where the negative welfare effect dominates when the employment share of non-adoption firms is high. This provides a first idea about where this distortion is expected to matter: In countries and industries with low adoption rates of management practices. Bloom and van Reenen (2007) report that adoption rates differ significantly across countries and industries, and these data provide a first step toward identifying sectors and countries where the distortion matters.

The mechanism developed in this paper is particularly relevant for industries where incentive pay and worker heterogeneity interact to shape the composition of the workforce. One necessary condition for our mechanism is that incentives actually matter, i.e. incentive pay increases the productivity of individual workers by raising their effort and pays higher wages to induce sorting. There is recent evidence from the management and personnel literature that this is indeed the case in many industries (Bloom and Van Reenen, 2011; Lazear and Oyer, 2013; Bureau of Labor Statistics, 2017a), and the Safelight Glass Corporation example from Lazear (2000) shows that even in a relatively simple industry such as car windshield replacement services, the productivity and sorting effects can be substantial. A second necessary condition is that worker heterogeneity matters for the productivity of workers, and

[24] Were we to introduce multiproduct firms, adopting firms would also opt to manage an excessive number of product lines while non-adopting firms would be inefficiently proscribed in their product offerings. See appendix 8.5 for details.

that this heterogeneity cannot easily be observed. There is also recent evidence that shows that unobserved worker heterogeneity is important in many industries and plays a key role in determining wage differentials across workers, both within and across industries and occupations (Helpman et al., 2016; Card et al., 2023). And finally, for co-existence we require the costs of implementing structured management practices to be non-negligible. Or as Lazear (2000, 1357-8) put it: "In equilibrium, firms that pay hourly wages or monthly salaries are probably those for whom measurement costs exceed the benefits from switching to output-based pay." Evidence is provided by Freeman and Kleiner (2005) who show that piece rates can improve productivity, but they also increase costs, including costs of monitoring, and for some firms (in their case for a U.S. show manufacturer) the costs effect can dominate.

In order to identify the distortion more rigorously it is useful to recall that the misallocation is driven by distortions in wages that do not reflect social opportunity costs. These distortions in wages lead to distortions in relative revenues across firm types, and these revenues are measurable.

Using (9) and (16) we obtain

$$\underbrace{\ln \bar{z}_n(\tilde{z}) - \ln \tilde{z}}_{\text{Distortion}} = \underbrace{\ln e(\bar{\varepsilon}, \tilde{z}) - \ln \tilde{z}}_{\text{Effort Effect}} - \frac{1}{\sigma - 1}(\ln r_a - \ln r_n) \tag{45}$$

The term on the left hand side is the log difference between average productivity and marginal productivity of the marginal worker in non-adopting firms. This is our measure of the misallocation that we want to identify and quantify. On the right hand side we have two terms: The second term consists of the log difference of revenues in adopting and non-adopting firms (r_a/r_n) and of the elasticity of substitution σ. Both are observable or estimable. The first term on the right hand side is the log difference of the productivity of the marginal worker in adopting firms and in non-adopting firms. Since this is the relative productivity of the same worker in the two types of firms, this difference is determined by the effort effect of the incentives in adopting firms. This effect can be quantified with insider econometrics like Lazear (2000) or Bandiera et al. (2005) (see Ichniowski and Shaw, 2013, for an overview). The difference between these two terms then provides an estimate for our inefficiency, where the hypothesis is that this difference is negative.[25]

This estimation procedure is quite robust to individual assumptions in our theory. First of all, it does not depend on free entry. In our figure 3, equation (45) measures the difference between *Social Planner* and *Sorting Condition* at $\tilde{z}$, independent of how $\tilde{z}$ is determined.

[25] Technically, because of the supermodularity between skills and effort, we also need to esimate $\tilde{z}$ which can be done by using employment shares of the two types of firms and putting some functional form assumption on $g(z)$.

Whether *Labor Market Equilibrium* is a flat horizontal line (free entry) or upward sloping (no entry) does not affect our estimation procedure. Second, the productivity between a worker in adopting firms may be higher than in non-adopting firms for other reasons than just the incentive pay effect, like complementary technological adoptions, peer effects etc. (see Ichniowski and Shaw, 2013, table 1). Again, this does not affect our estimation procedure as long as we can measure the aggregate productivity effect of management practices on workers. And finally, if the monitoring costs are variable costs and passed on to workers, effective wages are equalized and differences in revenues disappear. Instead, the inefficiency now shows up as a negative productivity effect among lesser skilled workers in adopting firms (see appendix 8.2).

Once we have identified and quantified our distortion, we can estimate our parameter θ in (39) and, using the functional form assumption $e(\varepsilon, z) = z^{\varepsilon}$ (see appendix 8.1), $\Gamma = \bar{\varepsilon} - \theta$. This allow us to calibrate our model and quantify the welfare effects using (38). In (38), λ is the observable share of labor in non-adopting firms, and trade costs τ can also be estimated.

7 Conclusion

A large and convincing empirical literature has established that a substantial portion of firm heterogeneity in productivity can be attributed to variation across firms in the adoption of structured management practices. Firms that adopt these practices are better able to monitor and to incentivize their workers and to attract more able workers. This literature has not analyzed the theoretical reasons for the incomplete adoption of these practices, nor has it analyzed the welfare implications of incomplete adoption.

In this paper, we specify a model that is based on the premises about the nature of structured management proposed by practitioners in the empirical literature and analyze the determinants of incomplete adoption and their normative consequences. We find that incomplete adoption can be explained as a natural outcome when the adoption of structured management practices is costly and talent in the workforce is scarce. From a normative point of view, we find that rather than there being too little adoption of structured management practices in equilibrium there is too much. This outcome arises because relatively talented workers want to be employed in an environment in which their individual output is monitored and rewarded to avoid having to be pooled with less capable workers where their wages would reflect the average ability of the pool.

Our model features a particular mechanism: structured management is interpreted as a situation in which output is monitored and firms pay piece rates. This mechanism is far more general than this particular micro-foundation. The key feature of our mechanism is that it allows skilled workers to avoid having to pool with less able workers by taking jobs at firms that monitor output. Many other micro-foundations would play a similar role. For instance, some firms offer sales commissions whereas others do not, allowing more able workers to sort to those that do. Other firms may screen entry level applicants more diligently than others. In each of these cases, some firms have better information about the abilities of their workers and naturally more able workers will tend to flock to these firms.

How this inefficiency shows up in the data depends on the extent to which structured management practices require overhead investments, the extent to which certain management teams are endowed with the ability to use these practices, and the extent of information spillovers over time made possible by observed early career wages. In any case, a trade liberalization event that induced reallocation of workers from small to large firms creates spillover effects that worsen the real wage outcomes of workers left behind. Insider econometrics that account for productivity changes as firms are induced to adopt structured management techniques can be used to identify the size of the misallocation induced by incomplete adoption.

The policy implications of the misallocation that arises in our model are starkly different than those that appear in other settings. In our model policies that discourage the adoption of structured management techniques can improve aggregate output while inducing a reallocation of resources from large to small firms. Papers that focus exclusively on the impact of market power in either the output or input market generally are built on assumptions that imply that large firms have stronger incentives to restrict output relative to small firms and so optimal policies require a reallocation of resources from small to large firms. Papers based on knowledge spillovers also tend to imply welfares gains from resource reallocations from small to large firms as larger firms are assumed to provide greater positive externalities. Our paper identifies a countervailing mechanism that demonstrates the need for caution in providing policy advice.

8 Appendix

8.1 Optimal Provision of Effort

In this appendix we provide some microfoundation for why firms that adopt structured management practices experience an increase in labor productivity among continuously employed workers. The theory is based on Lazear (1986, 2000).

A worker's objective function (worker surplus WS) is $WS = w(\varepsilon, z) - \gamma\varepsilon$, where $w(\varepsilon, z)$ denotes this workers (wage) income, depending on effort ε and skill z, and $\gamma > 0$ are the marginal (non-monetary) costs of providing effort. Throughout our analysis we assume that

γ is infinitesimally small so that it affects a worker's optimal provision of effort, but can be disregarded in the utility function (2) and in our welfare analysis.

The marginal benefits of providing effort are

$$\frac{dWS}{d\varepsilon} = \frac{\partial w(\varepsilon, z)}{\partial \varepsilon} - \gamma.$$

In the case of firms that do not monitor and incentivize their workers, workers receive a fixed salary w^0 and are required to work with a minimum effort $\varepsilon^0 = 1$. In this case, their surplus is given by

$$WS = \begin{cases} w^0 - \gamma\varepsilon & \text{if } \varepsilon \geq 1 \\ 0 & \text{if } \varepsilon < 1 \end{cases}.$$

Since $dWS/d\varepsilon = -\gamma < 0$ for all $\varepsilon \geq 1$, workers only provide the minimum effort required.

If firms pay a piece rate, the wage that workers receive in firm j is given by $w_j(\varepsilon, z) = c_j e(\varepsilon, z)$, where the piece rate c_j represents the effective wage per unit of effective labor supplied, w/e, by firm j. Their surplus is now given by

$$WS = c_j e(\varepsilon, z) - \gamma\varepsilon,$$

and the marginal surplus is

$$\frac{dWS}{d\varepsilon} = c_j \frac{\partial e(\varepsilon, z)}{\partial \varepsilon} - \gamma.$$

A necessary condition for eliciting effort through incentive pay is that $\partial e(1, z)/\partial\varepsilon > \gamma/c_j$, and we assume that γ is indeed small enough that this holds for all z. Second, we assume that there exists a maximum level of effort, $\bar{\varepsilon}$, and that $\partial e(\varepsilon, z)/\partial\varepsilon > \gamma/c_j$ for all $\varepsilon \in [1, \bar{\varepsilon}]$. A sufficient (but not necessary) condition for the second condition is $\partial^2 e(\varepsilon, z)/\partial\varepsilon^2 \geq 0$, i.e. if the marginal productivity effect of effort is not decreasing. The second assumption simplifies the analysis since it implies that optimal effort is fixed at $\varepsilon = \bar{\varepsilon}$ and changes in effective wages do not lead to changes in optimal effort.

With these assumptions, workers receiving a fixed salary work with effort $\varepsilon = 1$ and productivity $e(1, z) = z$ and workers receiving a piece rate work with effort $\varepsilon = \bar{\varepsilon} > 1$ and productivity $e(\bar{\varepsilon}, z) > z$.

A convenient functional form for $e(\varepsilon, z)$ that satisfies these conditions is $e(\varepsilon, z) = (z/\underline{z})^{\varepsilon}$ and $\underline{z} = 1$. In this case, $e(1, z) = z$, $w_j(\varepsilon, z) = c_j z^{\varepsilon}$ and $dWS/d\varepsilon = c_j z^{\varepsilon} \ln z - \gamma = w_j(\varepsilon, z) \ln z - \gamma$. Assuming again that γ is sufficiently small, workers with $z > \underline{z}$ have an incentive to work with effort $\bar{\varepsilon}$.

8.2 Marginal Monitoring Costs

Regarding the costs of monitoring, we assume in our main framework that these costs are borne by firms in the form of higher overhead costs. Lazear (1986) assumes that monitoring creates higher marginal costs that are passed on to workers. This is an alternative way of modelling these costs that can also be integrated in our framework and leads to qualitatively identical results.

Suppose workers in adopting firms can only spent a fraction $\chi \in (0,1)$ on their production tasks and $1-\chi$ is spent on reporting their progress (monitoring). This implies that adopting firms do not have to pay a higher fixed costs in order to monitor, but a variable iceberg cost per worker. Marginal unit labor requirements become $\chi^{-1} > 1$, and total variable labor requirements in adopting firms are $\chi^{-1}x_a$. Fixed costs are the same for both types of firms: $f_a = f_n = f$, so that $\Phi = 1$.

Following Lazear (1986), monitoring firms pass on the costs of monitoring to their workers, paying a discount on wages by a factor of χ. Hence, wages paid by adopting firms are now $w_a(z) = \chi c_a e(\bar{\varepsilon}, z)$. As a consequence, their profits are given by $\pi_a = A(\chi c_a/\chi)^{1-\sigma} - f = Ac_a^{1-\sigma} - f$. Non-monitoring firms continue to pay wages $w_n = c_n \bar{z}_n(\tilde{z})$ and earn profits $\pi_n = Ac_n^{1-\sigma} - f$. Free entry leads to zero profits of both types of firms and equalizes effective wages across firms: $Ac_n^{1-\sigma} = f = Ac_a^{1-\sigma}$, or $c_a = c_n$.

Sorting now requires that $w_n = c_n \bar{z}_n(\tilde{z}) = w_a(\tilde{z}) = \chi c_a e(\bar{\varepsilon}, \tilde{z})$, or, given $c_a = c_n$,

$$\frac{\bar{z}_n(\tilde{z})}{e(\bar{\varepsilon}, \tilde{z})} = \chi < 1. \tag{46}$$

Again, $\bar{z}_n(\tilde{z})/e(\bar{\varepsilon}, \tilde{z}) \in (0,1)$ implies incomplete adoption where $\tilde{z} \in (\underline{z}, \infty)$.

In the social planners problem, the second constraint now becomes $\chi^{-1} n_a^* x_a^* \le L \int_{\tilde{z}^*}^{\infty} e(\bar{\varepsilon}, z)\, dG(z)$ and the solution is $\tilde{z}^*/e(\bar{\varepsilon}, \tilde{z}^*) = \chi$. Hence, $\tilde{z} < \tilde{z}^*$, and the market solution is again inefficient.

In order to identify this distortion we can rewrite (46) as

$$\ln \bar{z}_n(\tilde{z}) - \ln \tilde{z} = [\ln e(\bar{\varepsilon}, \tilde{z}) - \ln \tilde{z}] + \ln \chi \tag{47}$$

In this case the productivity effect of monitoring depends on two counteracting effects: A positive effect of incentivizing workers to provide effort (the first term in square brackets), and a negative effect of the costs of monitoring (the second term, $\ln \chi < 0$). Because of the supermodularity of skills and effort, the positive effect dominates for highly skilled workers, whereas the negative effect dominates for workers in the lower distribution of Z_a. With free entry, revenues are equalized across firm types, and the two productivity effects cancel out in the aggregate. Using insider econometrics, our inefficiency would then show up as a negative productivity effect *at the margin*.

8.3 Social Planner's Problem

The social planner's problem as presented in section 4 implies the following Lagrangian:

$$\left(n_a^*\left(x_a^*\right)^{\frac{\sigma-1}{\sigma}}+n_n^*(q_n^*)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}+\lambda_M(M-n_a^*f_a-n_n^*f_n)+$$
$$\lambda_a\left(L\int_{\tilde{z}^*}^{\infty}e(\bar{\varepsilon},z)dG(z)-n_a^*x_a^*\right)+\lambda_n\left(L\int_0^{\tilde{z}^*}zdG(z)-n_n^*x_n^*\right).$$

Taking the derivative with respect to n_a^*, n_n^*, x_a^*, x_n^*, and $\tilde{z}^*$ we obtain after modest simplification

$$\frac{\sigma}{\sigma-1}\left(x_a^*\right)^{\frac{\sigma-1}{\sigma}}\left(n_a^*\left(x_a^*\right)^{\frac{\sigma-1}{\sigma}}+n_n^*(q_n^*)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}-1} = \lambda_M f_a+\lambda_a x_a^* \tag{48}$$
$$\frac{\sigma}{\sigma-1}\left(x_n^*\right)^{\frac{\sigma-1}{\sigma}}\left(n_a^*\left(x_a^*\right)^{\frac{\sigma-1}{\sigma}}+n_n^*(q_n^*)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}-1} = \lambda_M f_n+\lambda_n x_n^* \tag{49}$$
$$\left(x_a^*\right)^{-\frac{1}{\sigma}}\left(n_a^*\left(x_a^*\right)^{\frac{\sigma-1}{\sigma}}+n_n^*(q_n^*)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}-1} = \lambda_a \tag{50}$$
$$\left(x_n^*\right)^{-\frac{1}{\sigma}}\left(n_a^*\left(x_a^*\right)^{\frac{\sigma-1}{\sigma}}+n_n^*(q_n^*)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}-1} = \lambda_n \tag{51}$$
$$\frac{\lambda_a}{\lambda_n} = \frac{\tilde{z}^*}{e(\bar{\varepsilon},\tilde{z}^*)} \tag{52}$$

Using these first order conditions we can now prove the three parts of our proposition 3:

Part (ii): Substituting (50) and (51) into (48) and (49) and then dividing the resulting expression yields

$$\frac{x_n^*}{x_a^*}=\Phi^\sigma. \tag{53}$$

Equations (20), (21), (22) in the text collectively imply

$$\frac{x_n}{x_a}=\Phi^\sigma.$$

Hence, the social planner would choose relative output levels of n and a varieties that mirror those that the market would generate. Output levels of individual varieties are efficient.

Part (i): The ratio of (50) to (51) combined with (52) and (53) together imply

$$\frac{\tilde{z}^*}{e(\bar{\varepsilon},\tilde{z}^*)}=\Phi.$$

Comparing this expression to the market equilibrium cutoff shown in equation (22) we have

$$\frac{\tilde{z}^*}{e(\bar{\varepsilon},\tilde{z}^*)}=\frac{\overline{z}_n\left(\tilde{z}\right)}{e(\bar{\varepsilon},\tilde{z})}$$

Because $\overline{z}_n\left(\tilde{z}\right)<\tilde{z}$ for all $\tilde{z}$ if must be that $\tilde{z}^*>\tilde{z}$. Too many units of effective labor are allocated by the market to the a technology relative to the output maximizing social planner.

Part (iii): Because the social planner chooses to allocate more labor to the n technology and less to the a technology while choosing the same levels of output of each variety type as in the market equilibrium, it follows directly from the resource constraint that $n_a^*<n_a$ and $n_n^*>n_n$. Since $f_n<f_a$ this means that the market features too few varieties relative to the social optimum.

8.4 Firm-specific Unit Labor Requirements

In section 2.1.2 we abstract from differences in unit labor requirements and normalize α to one. In this section, we discuss firm-type specific unit labor requirements.

Let unit labor requirements of firms of type j be denoted by α_j. Then, equation (22) changes to

$$\Phi_\alpha\equiv\frac{\alpha_n}{\alpha_a}\left(\frac{f_n}{f_a}\right)^{\frac{1}{\sigma-1}}. \tag{54}$$

Coexistence of high- and low-tech firms requires that $\Phi_\alpha<1$, or

$$\frac{\alpha_n}{\alpha_a}<\left(\frac{f_a}{f_n}\right)^{\frac{1}{\sigma-1}}. \tag{55}$$

Since $f_a>f_n$, $\alpha_n=\alpha_a$ is a sufficient condition for coexistence.

But coexistence is also sustainable if high-tech firms have a lower unit labor requirement. In this case, it is important that the relative difference in unit labor requirements is not too large relative to the difference in fixed costs:

$$(\sigma-1)\left(\ln\alpha_n-\ln\alpha_a\right)\leq\left(\ln f_a-\ln f_n\right). \tag{56}$$

One interesting case is where the investment f_j can be interpreted as an innovation in management *and* production technology, where a higher f_j also reduces unit labor requirements α_j. Suppose

$$\alpha_j=f_j^{-\rho}. \tag{57}$$

In this case, coexistence requires that the responsiveness of unit labor requirements with respect to fixed costs is not too large, or

$$\rho < (\sigma - 1)^{-1}. \tag{58}$$

8.5 Multiproduct Firms

Our framework can be extended to make contact with the literature on multiproduct firms and international trade. In the baseline model we focus just on single-product firms, but in this extension we adopt the flexible manufacturing apparatus present in Eckel and Neary (2010) and demonstrate that multiproduct and single product firms can arise endogenously in our framework. As in Eckel and Neary (2010) and Bernard et al. (2011), trade liberalization induces firms to pare their high cost product lines that are sold only domestically but the reallocation of labor from small to large firms has the implication that the share of multiproduct firms in total output expands.

In order to extend this framework to multiproduct firms, we follow Eckel and Neary (2010) and assume that all firms possess a certain core competency for a specific variety where their unit labor costs is lowest for all products in their product range. All other products in their product range can then be identified by their (unidimensional) distance to the firm's core competency, denoted by $\omega > 0$. Production of multiple products is subject to flexible manufacturing, which implies that firms can add and drop products to and from their product range freely, but as they add products to their product range and move away from their core competency, unit labor requirements of these products increases. Thus, unit labor requirements α depend on the position ω of a product in a firm's product range, and are increasing in ω:

$$\alpha = \alpha(\omega) \qquad \text{and} \qquad \alpha'(\omega) \equiv \partial\alpha/\partial\omega > 0. \tag{59}$$

To simplify notation we normalize unit labor requirements at the core to one: $\alpha(0) = 1$. In addition, there is a fixed costs per product of f^p.

The profits of a multiproduct firm are

$$\pi_j^{mpf} = \int_0^{\omega_j} A c_j^{1-\sigma} \alpha(\omega)^{1-\sigma} \, d\omega - \omega_j f^p - f_j \tag{60}$$

and the optimal product range ω_j is given implicitly by $d\pi_j^{mpf}/d\omega_j = 0$:

$$A c_j^{1-\sigma} \alpha(\omega_j)^{1-\sigma} = f^p \tag{61}$$

Depending on the size of f^p there are three possible outcomes:

- $f_n < f_a \leq f^p$: All firms are single-product firms
- $f_n \leq f^p < f_a$: Adopting firms are multiproduct firms, non-adopting firms are single-product firms
- $f^p < f_n < f_a$: All firms are multiproduct firms

For this extension suppose that $f^p < f_n < f_a$ so that all firms are multiproduct firms. Combining zero profits with the optimal product range yields

$$\alpha\left(\omega_j\right)^{\sigma-1}\int_0^{\omega_j}\alpha\left(\omega\right)^{1-\sigma}d\omega-\omega_j=\frac{f_j}{f^p} \tag{62}$$

The left hand side is increasing in ω_j and the right hand side is increasing in f_j, so $f_a > f_n$ implies

$$\omega_a > \omega_n \tag{63}$$

Using the same procedure as in section 2.2.4 we obtain the following equation that determines the allocation of labor across firm types:

$$\frac{\bar{z}_n\left(\tilde{z}_{mpf}\right)}{e\left(\bar{\varepsilon},\tilde{z}_{mpf}\right)}=\Phi_{mpf}. \tag{64}$$

Here, Φ_{mpf} depends on the measure of product ranges and is defined as

$$\Phi_{mpf}^{\sigma-1}\equiv\left(\frac{\int_0^{\omega_a}\alpha\left(\omega\right)^{1-\sigma}d\omega}{\int_0^{\omega_n}\alpha\left(\omega\right)^{1-\sigma}d\omega}\right)\left(\frac{\omega_n f^p+f_n}{\omega_a f^p+f_a}\right)<1 \tag{65}$$

Since $\omega_a > \omega_n$ and $f_a > f_n$ implies that $\Phi_{mpf} < 1$, $\tilde{z}_{mpf} \in (\underline{z}, \infty)$ and both types of firms coexist. Furthermore, (16) implies that $c_a < c_n$. Thus, the misallocation and the welfare effects are qualitatively identical. We focused here on the closed economy but extending this extension to an open economy is rather straightforward.

Proposition 10 (Multiproduct Firms) *Adopting firms produce more products.*

Adopting firms pay a lower effective wage rate than non-adopting firms ($c_a < c_n$). This allows them to expand into less efficient activities and produce varieties further away from their core competency with higher unit labor requirements. They have an incentive to do so because management practices are applicable in all divisions within the firm, so that by adding products to their product range they can lower the fixed costs per product.

8.6 Multiple Export Destinations

Suppose the world is a circle with a continuum of countries, each located at a point on this circle. The distance between countries δ is the shortest arcdistance on the circumference of the circle. Variable trade costs between countries i and j are increasing in distance: $\tau_{ij} = \tau(\delta)$, $\tau(0) = 1$, $\tau'(\delta) > 0$, $\tau''(\delta) > 0$. Note that there is no index on $\tau(\delta)$ so this function is symmetric across all locations. In addition, there are fixed costs f^x per country exported to.

The profits of exporting for firm type j are now given by:

$$\pi_j^{mx} = 2\left(Ac_j^{1-\sigma}\int_0^{\delta_j}\tau(\delta)^{1-\sigma}\,d\delta - \delta_j f^x\right) \tag{66}$$

where δ_j measure the mass of countries firm j exports to. Note that countries are symmetrically exporting to the left and to the right of their location. Hence, profits are multiplied by factor two.

The optimal mass of export destinations is given by

$$\frac{d\pi_j^{mx}}{d\delta_j} = 2\left(Ac_j^{1-\sigma}\tau(\delta_j)^{1-\sigma} - f^x\right) = 0. \tag{67}$$

Since $\tau(0) = 1$ and variable trade costs of exporting to the closest destination are infinitesimally small, our conditions for exporting from section 4 have to be slightly altered:

- $f_n < f_a < f^x$: No exporting
- $f_n < f^x < f_a$: Exporting only by adopting firms
- $f^x < f_n < f_a$: All firms export

For this extension suppose that $f^x < f_n < f_a$ so that all firms export. Aggregate profits for firm of type j are then given by

$$\pi_j^{mx} = A\left(1 + 2\int_0^{\delta_j}\tau(\delta)^{1-\sigma}\,d\delta\right)c_j^{1-\sigma} - 2\delta_j f^x - f_j, \tag{68}$$

where mx stands for *m*ultiple e*x*port destinations. Combining free entry and (67) yields

$$\tau(\delta_j)^{\sigma-1}\left(1 + 2\int_0^{\delta_j}\tau(\delta)^{1-\sigma}\,d\delta\right) - 2\delta_j = \frac{f_j}{f^x} \tag{69}$$

Since the left hand side of (69) is increasing in δ_j and the right hand side is increasing in

f_j, $f_a > f_n$ implies

$$\delta_a > \delta_n. \tag{70}$$

Using the same procedure as in section 2.2.4 we obtain the following equation that determines the allocation of labor across firm types:

$$\frac{\bar{z}_n(\tilde{z}_{mx})}{e(\bar{\varepsilon}, \tilde{z}_{mx})} = \Phi_{mx}. \tag{71}$$

Here, Φ_{mx} depends on the measure of export destinations and is defined as

$$\Phi_{mx}^{\sigma-1} \equiv \left(\frac{1 + 2\int_0^{\delta_a} \tau(\delta)^{1-\sigma} \, d\delta}{1 + 2\int_0^{\delta_n} \tau(\delta)^{1-\sigma} \, d\delta}\right)\left(\frac{2\delta_n f^x + f_n}{2\delta_a f^x + f_a}\right) < 1 \tag{72}$$

Since $\delta_a > \delta_n$ implies that $\Phi_{mx} < 1$, $\tilde{z}_{mx} \in (\underline{z}, \infty)$ and both types of firms coexist. Furthermore, (16) implies that $c_a < c_n$. Thus, the misallocation and the welfare effects are qualitatively identical.

Proposition 11 (Multiple Export Destinations) *Adopting firms export to more destinations and ship longer distances.*

8.7 Multiple Degrees of Adoption

In this extension we illustrate that our assumption of just two degrees of adoption, adopt and non-adopt, and consequently just two types of firms, is not critical for the mechanism we describe. The simple case with just two types of firms is very useful for understanding the forces behind the mechanism, and for providing intuition, but it needs to be extended for empirical work. Here we show a straightforward way to extend this framework to include multiple degrees of adoption that lead to a higher degree of firm heterogeneity. We can think of these as different degrees of adopting what Bloom et al. (2014) refer to as "best practice" techniques.

Suppose there are N degrees of adopting management practices $j \in \{0, ..., N\}$, and each degree allows firms to observe workers up to a specific skill level. A degree j allows firms to observe up to skill level $\breve{z}_j$, so that all skills with $z < \breve{z}_j$ can be observed and all skills with $z > \breve{z}_j$ cannot. This also implies that only workers with skills $z < \breve{z}_j$ can be incentivized, and workers with skills $z > \breve{z}_j$ cannot. A higher degree of adoption (with a higher $\breve{z}_j$) is more expensive and requires strictly higher fixed costs, so that $f_{j+1} > f_j > 0$.[26]

[26] Our modelling with many screening technologies is similar to a discrete version of screening in Helpman, Itskhoki and Redding (2010).

Given that a firm with adoption level j (type-j firm) can only observe skills up to level $\breve{z}_j$, it can only pay a piece rate up to this level. Above this level, it has to pay a wage based on the average productivity.[27] Hence, the wage paid by a type-j firm to workers with skill z depends on the skill of these workers:

$$w_j(z) = c_j \times \begin{cases} e(\bar{\varepsilon}, z) & \text{for} \quad \tilde{z}_j < z < \breve{z}_j < \tilde{z}_{j+1} \\ \bar{e}(1, \breve{z}_j, \tilde{z}_{j+1}) & \text{for} \quad \tilde{z}_j < \breve{z}_j < z < \tilde{z}_{j+1} \end{cases}, \tag{73}$$

where $\bar{e}(\varepsilon, z_1, z_2) \equiv \int_{z_1}^{z_2} e(\varepsilon, z)\, dG(z) / [G(z_2) - G(z_1)]$. The set of workers that sort into the labor pool of type j is $Z_j = \{z : z \in (\tilde{z}_j, \tilde{z}_{j+1})\}$, so that $\tilde{z}_j$ and $\tilde{z}_{j+1}$ denote the lower and upper boundary of skills in firms of type j.

[Figure 4 here]

The profile of wages in type-j firms is illustrated in Figure 4. In the range between $\tilde{z}_j$ and $\breve{z}_j$, firms of type j can fully observe the true productivity of individual workers and pay a wage based on this productivity: $w_j(z) = c_j e(\bar{\varepsilon}, z)$. Above $\breve{z}_j$, these firms only know the average productivity of their workforce because their degree of adoption of management practices does not allow them to monitor workers with these skills. As a consequence, workers are also not incentivized, and the wage is based on $\bar{e}$: $w_j(z) = c_j \bar{e}(1, \breve{z}_j, \tilde{z}_{j+1})$. All workers in a type-$j$ labor pool receive the same effective wage rate c_j.

Note that there is a discrete jump in the wage profile within a labor pool. This jump is due to the fact that the average skills of workers in the range $(\breve{z}_j, \tilde{z}_{j+1})$ is discretely higher than the skills of worker $\breve{z}_j$. However, the jump is mitigated by the fact that workers in the range $(\breve{z}_j, \tilde{z}_{j+1})$ are not incentivized and, therefore, only provide the minimal effort $\varepsilon = 1$.

The boundaries of a particular labor pool are determined by the sorting conditions. In the case of $N + 1$ technologies there are N sorting conditions. For the boundary between firms of type j and $j - 1$, this sorting condition is

$$\frac{c_j}{c_{j-1}} = \frac{\bar{e}(1, \breve{z}_{j-1}, \tilde{z}_j)}{e(\bar{\varepsilon}, \tilde{z}_j)} \tag{74}$$

If this condition is fulfilled, workers with skill $\tilde{z}_j$ receive the same wage in the two labor pools j and $j-1$. Since $\tilde{z}_j > \breve{z}_{j-1}$, their skill is above the observable skill in type-$(j-1)$ firms, so their wage in the type-$(j-1)$ labor pool is $c_{j-1}\bar{e}(1, \breve{z}_{j-1}, \tilde{z}_j)$. In the type-$j$ labor pool their skill is below the threshold of observability, so firms with adoption degree j will pay them a wage based on their true productivity of $c_j e(\bar{\varepsilon}, \tilde{z}_j)$. We assume that condition (17) holds locally.

[27] We maintain the assumption that firms know the distribution of skills in their labor pool.

Given this sorting condition, the marginal skill $\tilde{z}_j$ is then determined by

$$\frac{\bar{e}\left(1, \breve{z}_{j-1}, \tilde{z}_j\right)}{e\left(\bar{\varepsilon}, \tilde{z}_j\right)} = \Phi_j, \tag{75}$$

where

$$\Phi_j^{\sigma-1} \equiv \frac{f_{j-1}}{f_j} < 1, \tag{76}$$

and our key corollaries with respect to relative revenues per worker (corollary 3) and average wages (corollary 4) also hold in this extension.

Proposition 12 (Discrete Adoption) *With multiple degrees of adoption, sorting leads to multiple labor pools where different degrees of adoption can coexist. Firms with a higher degree of adoption (a higher observable skill level $\breve{z}$) recruit more productive workers and exhibit higher revenues per worker and higher average wages.*

The market equilibrium in the case with multiple degrees of adoption is similarly distorted as the equilibrium with only two levels of adoption. Because wages of workers with above average skills in each labor pool receive wages based on their average productivity and not on their marginal productivity, the skill set of workers sorting into a particular labor pool ($j > 0$) is too low compared to the social optimum ($\tilde{z}_j < \tilde{z}_j^*$). As a consequence, for any degree of adoption $j > 0$, employment in firms with a lower degree of adoption is too low and in firms with a higher degree of adoption too high compared to the social optimum: $G\left(\tilde{z}_j\right) L < G\left(\tilde{z}_j^*\right) L$ and $\left[1 - G\left(\tilde{z}_j\right)\right] L > \left[1 - G\left(\tilde{z}_j^*\right)\right] L$.

Finally, we want to point out that the assumption of a discrete number of degrees of adoption is important. If N goes to infinity, each labor pool becomes infinitesimal and the boundaries of the pools collapse onto their observable skills. Formally, $\tilde{z}_j$ approaches $\breve{z}_j$ from below, and $\tilde{z}_{j+1}$ approaches $\breve{z}_j$ from above. Then,

$$\lim_{\tilde{z}_{j+1} \to \breve{z}_j} \bar{e}_j\left(1, \breve{z}_j, \tilde{z}_{j+1}\right) = e_j\left(1, \breve{z}_j\right), \tag{77}$$

and there is no more pooling in the labor market and no more uncertainty about skill levels. Each firm gets exactly the skill level that they target with their technology, and effective wages are equalized across different degrees of adoption. As a consequence, the inefficiency associated with pooling also disappears.

Proposition 13 (Continuous Technologies) *With continuous adoption possibilities, there is no more pooling in the labor market and the inefficiency associated with pooling disappears.*

References

[1] Arkolakis, Costas, Arnaud Costinot, Dave Donaldson, and Andres Rodriguez-Clare. 2019. "The Elusive Pro-Competitive Effects of Trade." *Review of Economic Studies* 86(1): 46-80.

[2] Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen. 2017. "The Fall of the Labor Share and the Rise of Superstar Firms." *Quarterly Journal of Economics* 135(2): 645-709, 2020

[3] Bandiera, Oriana, Iwan Barankay, and Imran Rasul. 2005. "Social Preferences and the Response to Incentives: Evidence from Personnel Data." *Quarterly Journal of Economics* 120: 917–962.

[4] Bandiera, Oriana, Guiso, Luigi, Prat, Andrea, and Sadun, Raffaella. 2015. "Matching firms, managers, and incentives." *Journal of Labor Economics* 33(3): 623-681.

[5] Bender, Stefan, Nicholas Bloom, David Card, John Van Reneen, and Stephanie Wolter. 2018. "Management Practices, Workforce Selection and Productivity." *Journal of Labor Economics*, vol. 36, no. S1

[6] Bernard, Andrew B., J. Bradford Jensen, Stephen J. Redding, and Peter K. Schott. 2007. "Firms in International Trade." *Journal of Economic Perspectives* 21(3): 105-130.

[7] Bernard, Andrew B., J. Bradford Jensen, Stephen J. Redding, and Peter K. Schott. 2018. "Global Firms." *Journal of Economic Literature* 56(2): 565-619.

[8] Bernard, Andrew B. and Toshihiro Okubo. 2016. "Product Switching and the Business Cycle." NBER Working Paper No. 22649, NBER Working Paper Series: Cambridge.

[9] Bernard, Andrew B., Stephen J. Redding, and Peter K. Schott. 2011. "Multiproduct Firms and Trade Liberalization." *Quarterly Journal of Economics* 126(3): 1271-1318.

[10] Black, Sandra E., and Lisa M. Lynch. 2001. "How to Compete: The Impact of Workplace Practices and Information Technology on Productivity." *Review of Economics and Statistics*, vol. 83(3), 434-45.

[11] Black, Sandra E., and Lisa M. Lynch. 2004. "What's Driving the New Economy?: The Benefits of Workplace Innovation." *The Economic Journal*, 114, 97-116.

[12] Bloom, Nicholas, Erik Brynjolfsson, Lucia Foster, Ron Jarmin, Megha Patnaik, Itay Suporta-Ecksten, and John Van Reenan. 2019. "What Drives Differences in Management Practices." *American Economic Review* 109(5): 1648-1683.

[13] Bloom, Nicholas, Renata Lemos, Raffaella Sadun, Daniela Scur, and John Van Reenan. 2014. "The Empirical Economics of Management." *Journal of the European Economic Association* 12(4): 835-876.

[14] Bloom, Nicholas, and John Van Reenan. 2007. "Measuring and Explaining Management Practices across Firms and Countries." *Quarterly Journal of Economics* 122(4), 1351-1408

[15] Bloom, Nicholas and John Van Reenen. 2011. "Human Resource Management and Productivity." In: Card, David and Orley Ashenfelter (eds.). Handbook of Labor Economics (Chapter 19). Volume 4b. Elsevier: 1697-1767.

[16] Brynjolfsson, Erik and Kristina McElheran. 2016. "The Rapid Adoption of Data-Driven Decision-Making." *American Economic Review* 106(5): 133-139.

[17] Bureau of Labor Statistics. US Department of Labor. 2017a. "A look at incentive-based versus time-based pay in 2016." *The Economics Daily* at https://www.bls.gov/opub/ted/2017/a-look-at-incentive-based-versus-time-based-pay-in-2016.htm.

[18] Bureau of Labor Statistics. US Department of Labor. 2017b. "Distribution of private sector employment by firm size class: 1993/Q1 through 2017/Q1, not seasonally adjusted". at https://www.bls.gov/web/cewbd/table_f.txt.

[19] Card, David, Jesse Rothstein and Moises Yi. 2023. "Industry Wage Differentials: A Firm-Based Approach." NBER Working Paper 31588, DOI 10.3386/w31588

[20] Caruso, Anthony. 2015. "Statistics of U.S. Businesses Employment and Payroll Summary: 2012. Economy-Wide Statistics Briefs." Report Number: G12-SUSB. U.S. Census Bureau.

[21] Cowgill, Bo. 2020. "Bias and Productivity in Humans and Algorithms: Theory and Evidence from Resume Screening." Working Paper Columbia University.

[22] Criscuolo, Chiara, Peter N. Gal, and Carlo Menon. 2014. "The Dynamics of Employment Growth: New Evidence from 18 Countries." OECD Science, Technology and Industry Policy Papers No. 14. OECD Publishing: Paris.

[23] Dixit, Avinash K. and Joseph E. Stiglitz. 1977. "Monopolistic competition and optimum product diversity." *The American Economic Review* 67.3: 297-308.

[24] Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu. 2015. "Competition, Markups, and the Gains from International Trade." *American Economic Review* 105(10): 3183-3221.

[25] Eckel, Carsten and J. Peter Neary. 2010. "Multi-product Firms and Flexible Manufacturing in the Global Economy." *Review of Economic Studies* 77(1): 188–217.

[26] Feenstra, Robert C. 2018. "Restoring the Product Variety and Procompetitive Gains from Trade with Heterogeneous Firms and Bounded Productivity." *Journal of International Economics* 110: 16-27.

[27] Freeman, Richard B., and Morris M. Kleiner. 2005. "The last American shoe manufacturers: Decreasing productivity and increasing profits in the shift from piece rates to continuous flow production." *Industrial Relations: A Journal of Economy and Society* 44.2: 307-330.

[28] Freund, Caroline and Martha Denisse Pierola. 2015. "Export Superstars." *The Review of Economics and Statistics*, 97(5): 1023–1032.

[29] Greenwald, Bruce and Joseph Stiglitz. 1986. "Externalities in Economies with Imperfect Information and Incomplete Markets." *Quarterly Journal of Economics* 101(2): 229-264.

[30] Grossman, Gene M. 2013. "Heterogeneous workers and international trade." *Review of World Economics* 149.2: 211-245.

[31] Helpman, Elhanan, Itskhoki, Oleg, Muendler, Marc Andreas, & Redding, Stephen J. 2016. "Trade and inequality: From theory to estimation." *The Review of Economic Studies*, 84(1), 357-405.

[32] Helpman, Elhanan, Oleg Itskhoki, and Stephen J. Redding. 2010. "Inequality and Unemployment in a Global Economy." *Econometrica* 78(4), 1239-1283.

[33] Ichniowski, Casey and Kathryn Shaw. 2013. "Insider Econometrics: Empirical Studies of How Management Matters." in: The Handbook of Organizational Economics. R. Gibbons and J. Roberts., eds., (2012): 263-312

[34] Kroft, Kory, Yao Luo, Magne Mogstad, and Bradley Setzler (2023), "Imperfect Competition and Rents in Labor and Product Markets: The Case of the Construction Industry," NBER Working Paper 27325, DOI 10.3386/w27325

[35] Lazear, Edward P. 1986. "Salaries and Piece Rates." *The Journal of Business* 59(3): 405-431.

[36] Lazear, Edward P. 2000. "Performance Pay and Productivity." *American Economic Review* 90: 1346–1361.

[37] Lazear, Edward P and Paul Oyer. 2013. "Personnel Economics." In: Handbook of Organizational Economics, edited by Robert Gibbons and John Roberts: 479-519.

[38] Lazear, Edward P. and Kathryn L. Shaw. 2007. "Personnel Economics: The Economist's View of Human Resources." *Journal of Economic Perspectives* 21(4): 91-114.

[39] Levitt, Steve, John List, and Chad Syverson. 2013. "Toward an Understanding of Learning by Doing: Evidence from an Automobile Plant." *Journal of Political Economy* 121(4): 643-681.

[40] Lombardi, Stefano, Oskar Nordström Skans, and Johan Vikström. 2018. "Targeted wage subsidies and firm performance." *Labour Economics* 53: 33–45.

[41] Melitz, Marc J. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica* 71: 1695-1725

[42] Nocco, Antonella, Gianmarco Ottaviano, and Matteo Salto. 2019. "Geography, Competition, and Optimal Multilateral Trade Policy." *Journal of International Economics* 120: 145-161.

[43] Oyer, Paul, and Scott Schaefer. 2011. "Personnel economics: Hiring and incentives." In: Card, David and Orley Ashenfelter (eds.). Handbook of Labor Economics (Chapter 20). Volume 4b. Elsevier: 1769–1823.

[44] Shane, Scott (2012), "Small Business's Share of Employment is Shrinking" Forbes, 4/21/2012. https://www.forbes.com/sites/scottshane/2012/04/21/small-businesss-share-of-employment-is-shrinking/

[45] Yeaple, Stephen. 2005. "A Simple Model of Firm Heterogeneity, International Trade, and Wages." *Journal of International Economics* 65(1):1-20.

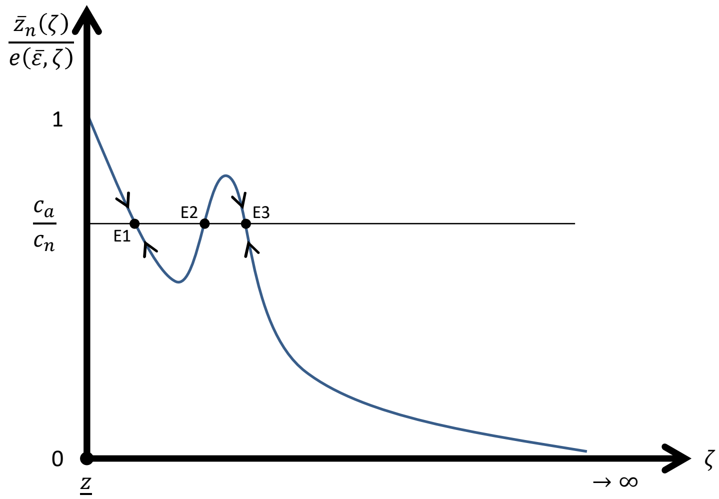

Figure 1: Stability of Sorting Equilibrium

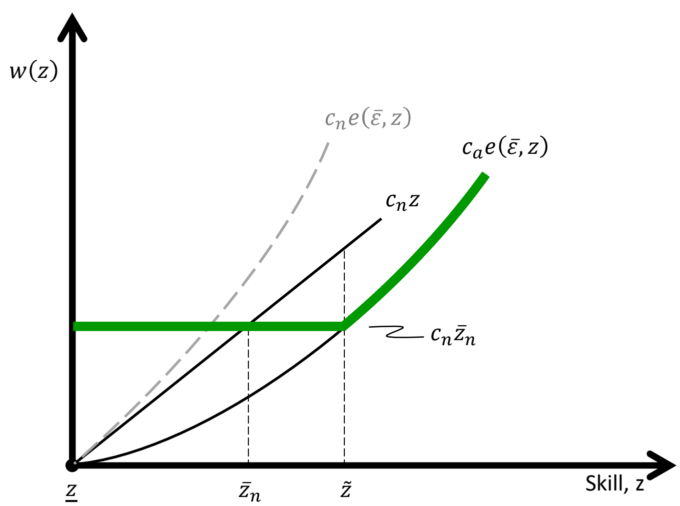

Figure 2: Hockey Stick Wage Profile

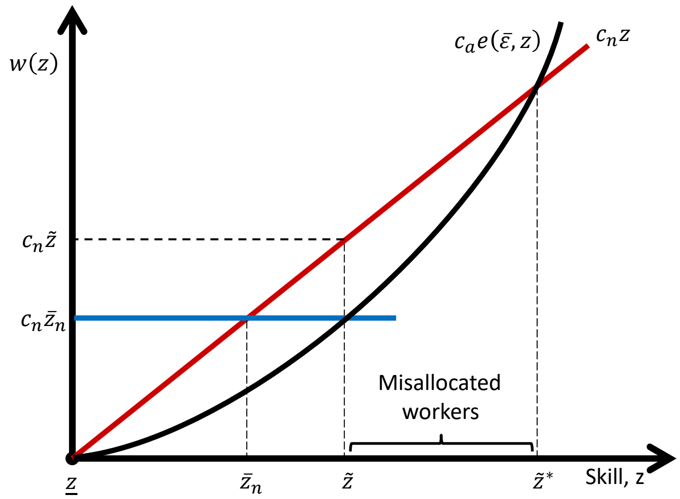

Figure 3: Allocation of Workers

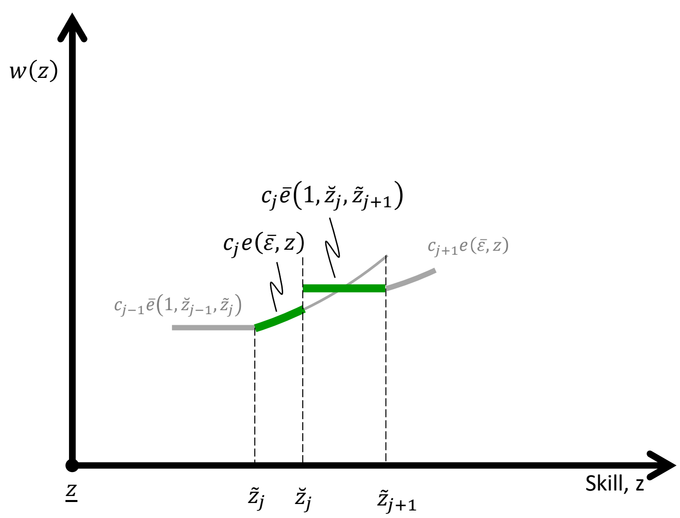

Figure 4: Multiple Screening Technologies

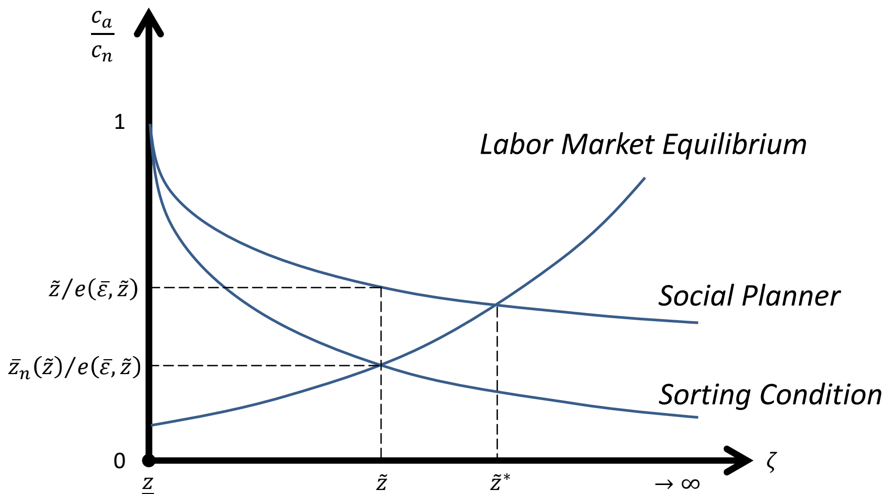

Figure 5: Exogenous Heterogeneity