Too Much of a Good Thing? Labor Market Imperfections as a Source of Exceptional Exporter Performance*

Carsten Eckel† Stephen R. Yeaple‡

January 29, 2019

Abstract

Ex-post firm heterogeneity can result from different strategies to overcome labor market imperfections by ex-ante identical firms – with far-reaching consequences for the welfare effects of trade. With asymmetric information about workers' abilities and costly screening, in equilibrium some firms screen and pay wages based on the true productivity of their workers, and some firms do not screen and pay wages based on the average productivity of their workforce. Screening firms are larger, attract better workers and pay lower effective wages. This results in excessive consumption of resources by large firms relative to the social optimum. Trade liberalization then has an ambiguous effect on aggregate welfare: lower trade costs improve access to foreign goods but also exacerbate the labor market distortion as more resources are transferred to large firms. The model highlights the need to know why firms "excel" before drawing welfare conclusions regarding cross firm reallocations of resources.

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

JEL Classification: F12, F16, J31, J33.

*We are thankful to Mathieu Parenti and participants of seminars in Dartmouth, UCD Dublin, Geneva, Hohenheim, Hong Kong, Leuven, the London School of Economics, Oxford, 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, and the Tinbergen Institute. The authors gratefully acknowledge support from the German Research Foundation (DFG) through SFB TR 15 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, multiproduct, exporting firms account for a large share of output, employment and trade, and this share is growing (Bernard et al., 2007; Shane, 2012; Criscuolo and Menon, 2014; Caruso, 2015; Bernard and Okubo, 2016; U.S. Bureau of Labor Statistics, 2017). For some, this outcome is the result of market forces that efficiently allocate resources to the most efficient firms. Others are pointing out that these developments may be driven by market failures based on factor market imperfections and call for support for small businesses (for recent discussions of these issues see e.g. Hurst and Pugsley, 2011; Mirrlees et al., 2011; Restuccia and Rogerson, 2013; WTO, 2016 or Hsieh and Klenow, 2017). There are also concerns that these large firms may enjoy market power in labor markets, and that this may have contributed to the fall in the labor share of income in industrialized countries (Autor et al., 2017; Krueger, 2018).

This issue is particularly important in the context of international trade. Many empirical studies have shown that the majority of all export value is generated by large firms (Bernard et al., 2007; Freund and Pierola, 2015; Bernard et al. 2018), and it is these large exporting firms that are benefiting most from reductions in international trade costs, whether they are induced by trade policies or by technological advances. As a consequence international trade is an important driver of the reallocations of resources towards large firms, and these reallocations are an important component of the overall welfare effects of international trade.

In this paper we show that to assess the welfare implications of the increased concentration of economic activity in large firms requires a knowledge of exactly why these firms are so large. We present a new model where ex post differences in firm size and productivity are generated by different strategies to overcome labor market imperfections by ex ante identical firms. The resulting industry structure exhibits many facts that are consistent with recent empirical findings: Larger firms are more productive, more likely to export (to more destinations), produce more products, pay higher wages, employ a more productive workforce and screen their workers more intensively. However, the resulting equilibrium is inefficient because too much labor is employed in large firms. This has two important implications: First, there is an incentive to subsidize small firms, and second, a reallocation of labor towards large firms has a first order negative welfare effect that can even dominate traditional gains from trade.

The importance of reallocations for the welfare effects of trade is currently debated in a growing literature. Melitz and Redding (2015) and Dhingra and Morrow (2016) point out that if the open economy equilibrium is socially optimal, these reallocations must necessarily enhance the welfare gains from trade. This is an important result because it applies to

the most popular framework in current trade theory, the monopolistic competition model with CES demand (as in Melitz, 2003). Melitz and Redding (2015) also point out that if the allocation is socially efficient, the welfare effects of trade-induced reallocations are only second order effects, which follows directly from an application of the envelop theorem.

Motivated by the importance of the social efficiency for these welfare results, some studies have looked at the welfare effects when the market equilibrium is not socially efficient. The focus of these studies has been on inefficiencies in the product market. Arkolakis et al. (2019) and Feenstra (2018) study the welfare effects in the presence of demand systems that provide variable mark-ups and show that the reallocation of resources towards firms with larger mark-ups tends to reduce the gains from trade relative to a case with constant mark-ups (Arkolakis et al., 2019), while a reduction in average mark-ups tends to affect welfare positively (Feenstra, 2018). The size of the two effects depends on assumptions regarding demand and the firm size distribution.

These studies assume that differences in measured productivity are driven by exogenous differences in marginal factor requirements, reflecting true differences in social efficiencies across firms. This is where we deviate. The recent literature on the sources of productivity differences and on the matching of heterogenous workers with firms points to differences in the quality of a firm's workforce as an important determinant for its productivity (e.g. Irarrazabal, Moxnes and Ulltveit-Moe, 2013; survey by Grossman, 2013). And this matching process is subject to many market imperfections that can lead to socially inefficient allocations of labor across firm types (Greenwald, 1986; Bloom and Van Reenen, 2011; Manning, 2011). A key problem in hiring talented workers is how to identify talent, and a large literature in labor and personnel economics has identified compensation policy as an appropriate (and widely used) tool to induce self-selection by quality (see Lazear and Shaw, 2007, Oyer and Schaefer, 2011, Bandiera et al., 2015, for recent overviews and empirical evidence). The focus of this paper is to study the role of these labor market imperfections for the endogenous formation of productivity differences across firms, and for the welfare effects of trade, with a particular focus on the trade-induced reallocations across firm types.

Our analysis is conducted in a version of the Yeaple (2005) model of endogenous firm heterogeneity in which information asymmetries give rise to labor market inefficiencies. In our model, workers' ability is private information on the labor market. Firms can adopt high-tech technologies for which high ability workers have a comparative advantage, or they can use an old technology in which worker ability is less crucial. To implement the high-tech technology, firms must invest in a human resources screening technology to identify the quality of workers.

In this setting, firms that invest in the high-tech technology have two advantages relative

to those that do not. First, their choice of technology lowers their marginal cost as it does in the original Yeaple (2005) paper. Second, their information advantage in the labor market confers onto them a form of market power. Because they select the best workers, adopting firms pay high wages, yet in equilibrium these wages do not fully compensate workers for their ability so that adopting firms have lower effective labor costs than non-adopting firms.[1] These lower effective labor costs allow these firms to grow and to expand into non-core activities, such as more product lines (as in extension 5.1) or exporting to more distant markets (as in appendix 7.2).

In equilibrium, the labor market recruiting advantage of large firms induces them to consume too many resources relative to the social optimum. This is because the marginal (most skilled) worker in small firms is paid a wage that reflects the average productivity of the labor pool and not her (higher) individual productivity. We show that an employment subsidy to small firms can achieve the socially optimal level of output by allocating more skilled workers to small firms.[2] Many industrialized countries do, in fact, have policies in place that favor small, non-exporting businesses. They are oftened viewed with scepticism and criticized as being the source of misallocations themselves. Our results suggest that these policies may deserve a re-evaluation in the light of the distortions described here. This policy implication of subsidizing small firms, however, is exactly opposite to the implications of models where the inefficiency is in the product market. Nocco, Ottaviano and Salto (2017) point out that in a framework with ex ante firm heterogeneity and variable mark-ups, optimal trade policy should promote the sales of large (low cost) firms. This shows once again the need to know why some firms "excel" and other do not before drawing policy conclusions.

Our model has important implications for the welfare effects of a trade liberalization. In the absence of a corrective subsidy to smaller firms, the impact of trade liberalization on aggregate welfare is ambiguous. On the one hand, a trade liberalization directly raises welfare through cheaper access to foreign produced goods as in standard models. On the other hand, a trade liberalization leads to a reallocation of resources from small to large firms and so worsens the market imperfection. This constitutes a first order negative welfare effect that can dominate the positive traditional gains from trade. Whatever the aggregate welfare impact, our model predicts that trade liberalization worsens income inequality as the

[1] In many models that generate the result that productive firms pay higher wages, labor institutions require firms to share rents with workers (e.g. Egger and Kreickemeier, 2009). Here, exactly the opposite obtains. Information asymmetries give employers an edge in the labor market and so workers share rents on their skill with firms.

[2] In our baseline model, we consider a scenario in which firms can enter freely as either small or large firms. We show in extension 5.3 that if firms are intrinsically heterogeneous that the result continues to hold. Rather than there being too many large firms and too few small firms, large firms are too large and small firms are too small relative to the social optimum.

real income of high ability workers rises and the real income of low ability workers falls.

The key assumptions of our paper that skilled workers have a comparative advantage using low marginal cost technologies and that firms make managerial investments in human resource management have received growing support in the empirical literature that uses matched employer-employee data. For instance, Bender et al. (2016) consider detailed employer-employee data from Germany. They show that average employee ability is higher for firms using advanced management practices and that a substantial portion of the productivity advantages of these firms can be attributed to their use of better workers.[3] Further, the authors directly document on-going selection by higher productivity firms of better-than-average employees,[4] writing

> ....better managed firms are able to build up a superior stock of employees through 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 lay off or fire the highest skilled workers in the period between 2004 and 2009. (p. 3)

As in Yeaple (2005), our model predicts wage stratification across firms within industries as workers select into firms that have adopted different technologies and that wage dispersion across firms should rise in response to trade liberalization. Unlike Yeaple (2005), our model also predicts greater wage dispersion within large, exporting firms than within small, non-exporters as it is these larger firms that have the informational advantage that allows them to compete for talent in the work force. Recent research by Becker et al. (2017) confirms that this is indeed the case among German firms: as firms become larger residual wage dispersion within firms grows.

Our paper is also related to Helpman, Itskhoki, and Redding (2010). In their paper, firms cannot observe worker ability and need to engage in costly screening to receive a noisy signal about their ability. Firms have an incentive to screen because the productivity of individual workers depends on the average ability of the workforce, so workers with a very low ability have a negative marginal product. In their framework, workers and firms are matched randomly and rents from search frictions are divided by symmetric bargaining.

[3] Friedrich (2017) uses matched employer-employee data for Belgium to show that high productivity firms invest in identifying more talented managers and then subsequently invest more heavily in their human capital accumulation. He models this empirical phenomenon as stemming from internal labor markets that arise from asymmetric learning and firm-specific human capital.

[4] Using similar German data, Card et al. (2013) establish that a significant portion of rising inequality among German workers can be attributed to increasing plant-level productivity heterogeneity and rising assortativeness in the assignment of workers to establishments.

As a result, there is endogenous ex-post heterogeneity in worker composition across firms, but the allocation of resources is socially optimal. As in Helpman, Itskhoki, and Redding (2010), we emphasize the imperfect observability of workers' abilities and the need for costly screening to overcome this imperfection. Unlike Helpman, Itskhoki, and Redding (2010), our framework exhibts asymmetric information and self-selection that leads to an inefficient matching of workers and firms.

Our paper contributes to the literature that explores how market imperfections in the presence of heterogeneous firms may affect the welfare impact of trade. Much of the recent literature has focused primarily on the product market by investigating the role of international trade on the reallocation of resources across firms that charge different mark-ups over their marginal cost (for example, see Arkolakis et al., 2019, Nocco, Ottaviano and Salto, 2017, and Edmond, Midrigan and Xu, 2015). In these settings, the key resource problem is that the most efficient firms are too small from a social point of view because they charge the highest mark-ups. In our setting, there is too much entry of large firms that stems from their artificially low cost of labor.[5]

Finally, our framework also provides useful insights with respect to the measurement of misallocations in the spirit of Hsieh and Klenow (2009). We show that in the social optimum, measures of total factor productivity - revenue (TFP-R) are not equalized across firms, not even if they are adjusted for average input quality. Thus, differences in TFP-R measurements, whether adjusted for input quality or not, are neither a necessary nor a sufficient condition for a misallocation.

The remainder of this paper is organized into five 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 some extensions to our basic framework before concluding.

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.

[5] Costinot, Rodriguez-Clare and Werning (2016) show that in a firm heterogeneity model a social planner can improve a country's welfare by raising tariffs on the most efficient exporters while leaving marginal exporters untaxed. This result is fundamentally different than ours as there is no rationale in their setting for subsidizing little firms.

2.1 Key Assumptions

2.1.1 Demand

On the demand side, we are not making any new or specific assumption but follow Krugman (1980). Consumers derive utility from the consumption of horizontally differentiated varieties. The utility function of a consumer is CES:

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

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

$$x(i) = EP^{\sigma-1} p(i)^{-\sigma}, \tag{2}$$

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

$$P \equiv \left(\int_{i \in \Omega} p(i)^{1-\sigma} di \right)^{\frac{1}{1-\sigma}}. \tag{3}$$

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 productivity) z. The distribution of skills in the economy is described by the 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 "Extensions" sections we discuss firm specific unit labor requirements as well as variety specific unit labor requirements in the context of multiproduct firms. 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 individual workers depends on the skills of these workers and on the technology used by the firm. There are two technologies available. In one technology, call it

low-tech, skills of workers are proportionate to their effective supply of labor $a(z)$. In this case a worker with skill z has an effective supply of labor of $a_L(z) = z$. In the other technology, call it high-tech, a worker with skill z has an effective supply of labor of $a_H(z) > z$, where $a_H(\underline{z}) = a_L(\underline{z}) = \underline{z}$, $a'_H(\underline{z}) \geq 1$ and $a''_H(z|\, z > \underline{z}) > 0$. Thus, a worker with a higher skill has an absolute advantage in both technologies, and a comparative advantage in the high-tech technology. This is essentially the same assumption as in Yeaple (2005). Since $a_L(z) = z$, we can drop the index H in $a_H(z)$.

Since the high-tech technology is superior to the low-tech technology, firms would always prefer to use the high-tech technology. However, we assume that the high-tech technology requires knowledge of the true productivity of workers, and this information is not available to all firms. In the absence of a screening technology, firms do not observe the productivity of any given worker. It is this information asymmetry that gives rise to the market imperfection. A screening technology exists but is only available to a firm if it incurs a higher fixed cost f_h (in units of management). We will refer to firms that invest in the screening technology as high-tech firms (h) and firms that do not as low-tech firms (l) and assume that

$$f_h > f_l. \tag{4}$$

One can think of this screening technology as an investment in a human resource staff that can accurately assess productivity. Firms that have incurred the higher fixed costs f_h can immediately evaluate the productivity of all workers while firms that have not acquired the screening technology can never observe the productivity of any individual worker. There are no learning effects. Thus, a firm that has invested in the screening technology knows the productivity of its workers and can use the more advanced technology. A firm that has not invested in this screening technology must use the less advanced technology.[6]

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 know only the distribution of productivity in the population, $G(z)$, which is common knowledge. We relax the latter assumption in an extension.

This is a one shot game that occurs in four stages. All agents have rational expectations and perfect foresight.

[6] Our narrative here suggests that firms pay a higher fixed cost for the screening technology and receive the more advanced technology as a side effect. But the two benefits (screening and technology) really come as package and our modelling is isomorphic to the alternative interpretation were firms pay the fixed costs for the advanced technology and this technology allows them to screen workers.

In stage 1, firms enter and decide whether they want to pay f_h and acquire the screening technology or pay $f_l < f_h$ without the screening technology. This determines their type: Type-h firms pay f_h, type-l firms pay f_l. There is a continuum of firms of both types and their masses will be denoted by n_j ($j \in \{h, l\}$).

Once firms have made their entry and screening technology investments, two labor markets open. Firms that have made the screening investment, $j = h$, operate in one labor market while firms that have not made the screening investment, $j = l$, operate in the other. Let the set of workers that ultimately choose to be in labor market j be denoted as Z_j.

We refer to the labor market associated with firms $j = h$ as the "frictionless" labor market because all information regarding workers in that labor market is known by all firms. Perfect competition implies that the wage of worker 1 relative to worker 2 with skills z_1, $z_2 \in Z_h$ and productivities $a(z_1)$ and $a(z_2)$ satisfy the no arbitrage condition $w_1/w_2 = a(z_1)/a(z_2)$.

We refer to the labor market associated with firms $j = l$ as the "frictional" labor market because individual worker productivies, $z \in Z_l$, are known only to the workers. The inability of firms $j = l$ to verify workers' productivities requires that there must be a single wage $w = w_S$ for all $z \in Z_l$.

In stage 2, workers choose whether to enter the frictionless or the frictional labor market. They make this choice with perfect foresight regarding the wage they would receive in each labor market.

In stage 3, both frictionless and frictional labor markets clear. 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 (2) and a market structure of monopolistic competition, the profit-maximizing price of firm j is a constant mark-up over its marginal costs:

$$p(c_j) = \frac{\sigma}{\sigma - 1} c_j, \tag{5}$$

where j denotes firm type $j \in \{h, l\}$. Since all firms have access to the same technology, 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 labor markets, their factor costs c_j may be different, hence the subscipt j.

In order to simplify notation we define

$$A \equiv (\sigma - 1)^{\sigma-1} \sigma^{-\sigma} E P^{(\sigma-1)}. \tag{6}$$

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 (2), (5) and (6), output of firm j can be written as

$$x(c_j) = (\sigma - 1) A c_j^{-\sigma}, \tag{7}$$

and revenues are

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

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

$$\pi(c_j) = A c_j^{1-\sigma} - f_j. \tag{9}$$

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 $j \in \{h, l\}$ is

$$n_j x_j = \tilde{L}_j, \tag{10}$$

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 $j \in \{h, l\}$ firms are atomistic and take wages as given.

Market clearing of the numéraire factor (management M) requires that

$$n_h f_h + n_l f_l = M. \tag{11}$$

It is implied in general equilibrium.

2.2.3 Worker Sorting

Workers can observe whether a firm has invested in the screening technology or not. Thus, they can decide whether they want to apply for a job in a high-tech firm (h) or in a low-tech firm (l) by choosing the respective labor pool. There are no differences in non-pecuniary job returns, so this decision is entirely based on differences in wages.

The labor market of high-tech firms (h) is perfectly competitive. After screening, the true productivity of workers is known by all firms in this labor market segment, and they can pay a wage to individual workers based on this worker's true productivity. Anticipating correctly the effective wage c_h determined in stage 4, firms of type-h pay

$$w_h(z) = c_h a(z). \tag{12}$$

The labor market of low-tech firms (l) is only imperfectly competitive. Firms in this labor market segment have not acquired the screening technology and hence never know the true productivity of their workers. But they do know the distribution of productivities in their labor market pool. 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:[7]

$$w_l = c_l \mathbb{E}_l(z), \tag{13}$$

where c_l is the effective wage rate in this labor market segment.

Given that wages differ between these two types of firms, each worker can decide whether he or she wants to apply for a job in the frictionless labor market of high-tech firms or in the frictional labor market of low-tech firms. The wage of a worker with productivity z is thus

$$w = \max\{c_l \mathbb{E}_l(z); c_h a(z)\}. \tag{14}$$

The following proposition describes the sorting outcome:

Proposition 1 (Sorting) *In an economy that features both high- and low-tech 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 high-tech firms, and workers with $z < \tilde{z}$ will choose to work for*

[7] Alternatively, one can simply assume that being technologically able to pay a piece rate wage requires some investment, so that firms that pay the higher fixed costs f_h are able to pay a piece rate wage and firms that only pay f_l are not. This microfoundation for the wage setting scheme does not rely on any assumption about screening.

low-tech firms. The critical $\tilde{z}$ is determined by

$$c_l \bar{z}_l(\tilde{z}) = c_h a(\tilde{z}), \tag{15}$$

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

Proof. Assume a $\tilde{z}$ exists, so that $\mathbb{E}_l(z) = \int_{\underline{z}}^{\tilde{z}} z dG(z) / G(\tilde{z}) = \bar{z}_l(\tilde{z})$. Then rewrite condition (15) as $\bar{z}_l(\tilde{z}) / a(\tilde{z}) = c_h / c_l$. Using L'Hôpital's rule, we can determine the limits of $\bar{z}_l(\tilde{z}) / a(\tilde{z})$ as $\tilde{z}$ approaches the boundaries of the support: $\lim_{\tilde{z} \to \underline{z}} [\bar{z}_l(\tilde{z}) / a(\tilde{z})] = 1$ and $\lim_{\tilde{z} \to \infty} [\bar{z}_l(\tilde{z}) / a(\tilde{z})] = 0$. Since $\bar{z}_l(\tilde{z}) / a(\tilde{z})$ is differentiable, this proves existence of (at least) one equilibrium with $\underline{z} < \tilde{z} < \infty$ for $c_h < c_l$. Furthermore, this equilibrium implies sorting where the most productive workers work for type-h firms and the least productive work for type-l firms: $c_h a(z) > c_l \bar{z}_l(\tilde{z})$ for $z > \tilde{z}$ and $c_h a(z) < c_l \bar{z}_l(\tilde{z})$ for $z < \tilde{z}$. This equilibrium is stable if for $\zeta < \tilde{z}$, $c_l \bar{z}_l(\zeta) > c_h a(\zeta)$, and for $\zeta > \tilde{z}$, $c_l \bar{z}_l(\zeta) < c_h a(\zeta)$. Thus, stability implies that $\bar{z}_l(\zeta) / a(\zeta)$ is decreasing in ζ at $\zeta = \tilde{z}$ and requires that

$$\frac{\tilde{z} g(\tilde{z})}{G(\tilde{z})} \frac{[\tilde{z} - \bar{z}_l(\tilde{z})]}{\bar{z}_l(\tilde{z})} < \frac{a'(\tilde{z}) \tilde{z}}{a(\tilde{z})}. \tag{16}$$

Since $\bar{z}_l(\tilde{z}) / a(\tilde{z})$ is decreasing globally (from 1 to 0), at least one stable equilibrium must exist. This equilibrium is unique if $\bar{z}_l(\tilde{z}) / a(\tilde{z})$ is monotonically decreasing.[8] ■

[FIGURE 1 here]

In Figure 1 we illustrate the equilibrium and its stability graphically. For illustrative purposes, the function $\bar{z}_l(\zeta) / a(\zeta)$ is not monotonic. Clearly, if $c_l \bar{z}_l(\zeta) > c_h a(\zeta)$, a worker with skill ζ earns higher wages in type-l firms than in type-h firms. Thus, if ζ was a sorting cutoff, this would not be an equilibrium because the marginal worker would want to work for type-l firms, leading to an increase in this cutoff. Therefore, a stable equilibrium requires that the $\bar{z}_l(\zeta) / a(\zeta)$-function intersects c_h / c_l 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 (16) holds.

One important implication of the sorting equilibrium is that

$$c_h = \frac{\bar{z}_l(\tilde{z})}{a(\tilde{z})} c_l < c_l. \tag{17}$$

Thus, high-tech firms that have invested in the screening technology pay a lower effective wage rate (in efficiency units) than low-tech firms with no access to the screening technology.

[8] If worker skills are Pareto distributed, inequality (16) always holds.

This has to hold in equilibrium because the productivity of the marginal worker is discretely higher than the average productivity of all workers with a lower productivity: $\tilde{z} > \bar{z}_l(\tilde{z})$. Therefore, low-tech firms have to pay a premium on the effective wage rate of high-tech firms in order to compensate their above-average workers for pooling them with below-average workers.

Note that the difference in technologies enlarges the wage differences in the two labor market segments, but is not a necessary condition for the labor market segmentation.

Corollary 1 *The difference in technologies between type-h and type-l firms is neither necessary nor sufficient for the sorting equilibrium.*

Proof. If $a_H(z) = a_L(z) = z$ equation (17) reduces to $c_h = [\bar{z}_l(\tilde{z})/\tilde{z}]\, c_l$, where the term $\bar{z}_l(\tilde{z})/\tilde{z}$ is larger than the term $\bar{z}_l(\tilde{z})/a(\tilde{z})$ in proposition 1 but behaves identically at the limits. ■

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.

In a sorting equilibrium, we can now also determine the effective supplies of labor $\tilde{L}_j$ for the two types of firms from (10):

$$\tilde{L}_l = LG(\tilde{z})\,\bar{z}_l(\tilde{z}) \qquad \text{and} \qquad \tilde{L}_h = L[1 - G(\tilde{z})]\,\bar{a}_h(\tilde{z}), \tag{18}$$

where $\bar{a}_h(\tilde{z}) \equiv \int_{\tilde{z}}^{\infty} a(z)\, dG(z) / [1 - G(\tilde{z})]$.

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 (9), this implies that

$$Ac_j^{1-\sigma} = f_j. \tag{19}$$

Taking ratios for $j = h, l$ we obtain

$$\frac{c_h}{c_l} = \left(\frac{f_l}{f_h}\right)^{\frac{1}{\sigma-1}}. \tag{20}$$

For convenience, we define

$$\Phi \equiv \left(\frac{f_l}{f_h}\right)^{\frac{1}{\sigma-1}}. \tag{21}$$

Now combining (15) and (20), and using the definition in (21), we obtain

$$\frac{\bar{z}_l(\tilde{z})}{a(\tilde{z})} = \Phi. \tag{22}$$

We can now prove the following proposition:

Proposition 2 (Co-existence) *In a free entry equilibrium, both types of firms (high-tech and low-tech firms) will co-exist.*

Proof. First note from (4) that high-tech firms have higher fixed costs. Therefore, a necessary condition for co-existence with free entry is that $c_h < c_l$, which is met [see (17)]. Second, we can show that an equilibrium with only one type of firm is inconsistent with free entry: If $\tilde{z} \to \infty$ (no high-tech firms), $\lim_{\tilde{z}\to\infty} c_h = c_l \lim_{\tilde{z}\to\infty} [\bar{z}_l(\tilde{z})/a(\tilde{z})] = 0$ and $\lim_{\tilde{z}\to\infty} \pi_h = +\infty$. Hence, high-tech firms must exist. If $\tilde{z} \to \underline{z}$ (no low-tech firms), $\lim_{\tilde{z}\to\underline{z}} c_h = c_l \lim_{\tilde{z}\to\underline{z}} [\bar{z}_l(\tilde{z})/a(\tilde{z})] = c_l$ and $\pi_l > \pi_h$ (because $f_l < f_h$). Hence, low-tech firms must exist. Inequality (4) is a necessary condition for co-existence. ■

Proposition 2 is at the core of our theory. It shows how firms with different cost structures can arise endogenously from ex ante identical firms due to labor market imperfections and different strategies to deal with them. Co-existence of low-tech firms and high-tech firms is only possible because the screening technology leads to sorting and allows firms to segment labor markets. In the frictional labor market, firms pay a wage based on the average productivity of workers in this labor market segment. Such a wage scheme implies an implicit transfer of rents from the more productive workers in this segment to the less productive workers. In the frictionless labor market, firms pay a wage based on the true productivity of workers so that no transfer takes place between workers. In the frictionless labor market, the implicit transfer is from workers to firms: Because high-skilled workers want to avoid being pooled with low-skilled workers for their wages, they are willing to accept a discount on their wages for not being pooled. This discount on wages is why high-tech firms can co-exist despite their higher fixed costs for screening.

The differences in the wage schemes between the two labor markets have important implications for the allocative efficiency of resources. In the frictional labor market, rents go from relatively high productive workers to less productive workers. But since all workers are paid the same wage, this does not affect the allocation of resources. In the frictionless labor market, the rents are transferred from (high productivity) workers to firms because firms are

paying a lower effective wage. This induces them to expand and implies a misallocation of resources. This will be important in the welfare analysis.

Our framework has a number of interesting implications that are important for empirical work or for welfare analysis. We present them here as corollaries of propositions 1 and 2:

Corollary 2 (Size) *High-tech firms have higher sales than low-tech firms.*

Proof. It follows directly from $c_h < c_l$ that $x(c_h) > x(c_l)$. ■

High-tech firms have higher sales because they have lower marginal production costs.

Corollary 3 (Productivity) *High-tech firms are more productive than low-tech firms as measured by revenue per worker.*

Proof. Using (5), (10), and (18), revenues per worker in high-tech firms φ_h can be expressed as

$$\varphi_h \equiv \frac{p(c_h)\, x(c_h)}{L\left[1 - G(\tilde{z})\right]/n_h} = \frac{\sigma}{\sigma - 1} c_h \bar{a}_h(\tilde{z}) \tag{23}$$

Similarly, revenues per worker in low-tech firms φ_l can be expressed as

$$\varphi_l \equiv \frac{p(c_l)\, x(c_l)}{LG(\tilde{z})/n_l} = \frac{\sigma}{\sigma - 1} c_l \bar{z}_l(\tilde{z}) \tag{24}$$

Then, using the sorting condition (15), the ratio of the two productivity measures can be expressed as

$$\frac{\varphi_h}{\varphi_l} = \frac{\bar{a}_h(\tilde{z})}{a(\tilde{z})} > 1 \tag{25}$$

■

Note that this result depends entirely on the sorting of workers, and not on the assumption of different technologies across firms. The difference in revenues per workers also shows up when there is only one homogenous technology available and $a(z) = z$ also holds for type-h firms. In this case, $\bar{a}_h(\tilde{z}) \equiv \int_{\tilde{z}}^{\infty} z dG(z) / [1 - G(\tilde{z})]$ and $\varphi_h/\varphi_l = \bar{a}_h(\tilde{z})/\tilde{z} > 1$. Therefore, high-tech firms generate higher revenues per worker because they employ more productive workers and not because they use a different technology.

Corollary 4 (Wages) *High-tech firms pay higher average wages.*

Proof. Low-tech firms pay a flat wage of $w_l = c_l \bar{z}_l$. High-tech firms pay wages based on individual productivities. The average wage in high-tech firms is $\bar{w}_h \equiv \int_{\tilde{z}}^{\infty} w(z)\, dG(z) / [1 - G(\tilde{z})] = c_h \bar{a}_h(\tilde{z})$. Again using (15), the relative average wage in high-tech firms is

$$\frac{\bar{w}_h}{w_l} = \frac{\bar{a}_h}{a(\tilde{z})} \left(= \frac{\varphi_h}{\varphi_l} \right) > 1 \tag{26}$$

■

High-tech 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 the frictional labor market and work for low-tech firms. They receive a flat wage given by $w_l = c_l \bar{z}_l$. Above $\tilde{z}$, workers decide to go on the frictionless labor market, work for high-tech firms and receive a wage $w_h(z) = c_h a(z)$. This figure also illustrates nicely why a sorting equilibrium implies that the effective wage c_l in the frictional labor market has to be larger than the effective wage c_h in the frictionless labor market. If low-tech firms paid the same effective wage as high-tech firms, $w_l = c_h \bar{z}_l$, then the wage for workers with above-average productivity $z > \bar{z}_l$ would be discretely lower in low-tech firms than in high-tech firms $[c_h \bar{z}_l < c_h a(z)$ for all $z \in (\bar{z}_l, \tilde{z})]$. Consequently, this could not be a sorting equilibrium. Instead, low-tech firms have to pay a premium on the effective wage rate, $c_l > c_h$, in order to compensate their above-average workers for pooling them with below-average worker, so that $c_l \bar{z}_l = c_h a(\tilde{z})$. Put differently, high-tech 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 ability with firms rather than the other way around as is common in much of the literature on "fair wages."

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:

$$E \equiv L \int_{\underline{z}}^{\infty} w(z)\, dG(z) + M = L \left[c_l \int_{\underline{z}}^{\tilde{z}} z dG(z) + c_h \int_{\tilde{z}}^{\infty} a(z)\, dG(z) \right] + M. \tag{27}$$

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

$$E = \frac{\sigma}{\sigma - 1} L \left\{ c_l G(\tilde{z}) \bar{z}_l(\tilde{z}) + c_h [1 - G(\tilde{z})] \bar{a}_h(\tilde{z}) \right\} = \sigma M. \tag{28}$$

With E determined, and A pinned down by (19), the price index P can be derived easily from (6).

3 Welfare Implications of Labor Market Imperfections

This section analyzes the welfare implications of the information advantage of high-tech firms. We begin by solving for the optimal allocation of labor to firms as chosen by a social planner that wishes to maximize aggregate real income. We then show that this is less labor than is allocated in a market equilibrium because too many workers self-select into the high-tech labor pool. We conclude the section by showing that the market imperfection can be improved by a subsidy to employment at small firms.

Given (1), aggregate welfare W can be expressed as

$$W = \frac{E}{P} = \frac{\sigma}{\sigma - 1} \frac{\bar{w}}{P} L, \tag{29}$$

where

$$\frac{\bar{w}}{P} = \int_{\underline{z}}^{\tilde{z}} \frac{c_l}{P} \bar{z}_l(\tilde{z}) \, dG(z) + \int_{\tilde{z}}^{\infty} \frac{c_h}{P} a(z) \, dG(z) \tag{30}$$

is the average real wage that consists of the average real wage in low-tech firms (the first term) and the average real wage in high-tech firms (the second term), weighed with the respective employment shares.

Using the free entry conditions (19) for $j \in \{h, l\}$, and substituting (6) for A, (28) for E, and (21) for Φ, welfare can be expressed as

$$W = \Xi \left(\int_{\underline{z}}^{\tilde{z}} z dG(z) + \Phi \int_{\tilde{z}}^{\infty} a(z) \, dG(z) \right), \tag{31}$$

where

$$\Xi \equiv \left(\frac{M}{f_l} \right)^{\frac{1}{\sigma - 1}} L \tag{32}$$

is a constant.

Note that for this expression of welfare we have not used the sorting condition (15) to pin down $\tilde{z}$. But we have used the zero profit conditions (19). Thus, maximization of (31) with respect to $\tilde{z}$ yields the constrained social optimum (as defined in Dixit and Stiglitz, 1977).

We obtain

$$\frac{dW}{d\tilde{z}} = \Xi \left[\tilde{z} - \Phi a\left(\tilde{z}\right)\right] g\left(\tilde{z}\right) \tag{33}$$

and the constrained social optimum is at $dW/d\tilde{z} = 0$, or

$$\frac{\tilde{z}^*}{a\left(\tilde{z}^*\right)} = \Phi. \tag{34}$$

This leads us to proposition 3:

Proposition 3 (Social Optimum) *A socially optimal allocation of labor $\tilde{z}^* \in (\underline{z}, \infty)$ exists that satisfies $\tilde{z}^* = \Phi a\left(\tilde{z}^*\right)$.*

Proof. The first derivative of (31) evaluated at $\tilde{z} = \tilde{z}^*$ yields (34). The second derivative is negative by (16). Since $\Phi \in (0,1)$ (because $f_h > f_l$), $\lim_{\tilde{z}^* \to \underline{z}} \tilde{z}^*/a\left(\tilde{z}^*\right) = 1$ and $\lim_{\tilde{z}^* \to \infty} \tilde{z}^*/a\left(\tilde{z}^*\right) = 0$, we have $\tilde{z}^* \in (\underline{z}, \infty)$. ■

Intuitively, the social optimum requires that the marginal productivity of labor and management are equalized across firm types. For the marginal productivity of labor, this implies that[9]

$$\frac{\tilde{z}^*}{a\left(\tilde{z}^*\right)} = \left(\frac{x_l^*}{x_h^*}\right)^{\frac{1}{\sigma}}. \tag{35}$$

For the marginal productivity of management, this implies that[10]

$$\left(\frac{x_l^*}{x_h^*}\right)^{\frac{\sigma-1}{\sigma}} = \frac{f_l}{f_h}. \tag{36}$$

Together, the two conditions yield (34).

Now that we know the socially optimal allocation of labor we can compare the sorting equilibrium to the social optimum. This leads to proposition 4:

Proposition 4 (Sorting Efficiency) *The sorting equilibrium leads to a socially inefficient allocation of labor across firm types. Compared to the social optimum, employment in high-tech firms is too high in the sorting equilibrium.*

Proof. The social optimum in (34) requires that $\tilde{z}^*/a\left(\tilde{z}^*\right) = \Phi$. The sorting equilibrium in (22) yields $\bar{z}_l\left(\tilde{z}\right)/a\left(\tilde{z}\right) = \Phi$. When we combine the two, we obtain $\tilde{z}^*/a\left(\tilde{z}^*\right) = \bar{z}_l\left(\tilde{z}\right)/a\left(\tilde{z}\right)$, and, thus, $\tilde{z} < \tilde{z}^*$. ■

[9] The marginal productivity of labor is $dX/dL_l = \left(X/x_l\right)^{\frac{1}{\sigma}} \tilde{z}$ in low-tech firms and $dX/dL_h = \left(X/x_h\right)^{\frac{1}{\sigma}} a\left(\tilde{z}\right)$ in high-tech firms, where $X \equiv UL$, $L_l = G\left(\tilde{z}\right) L$ and $L_h = \left[1 - G\left(\tilde{z}\right)\right] L$.

[10] The marginal productivity of management in firms of type j is $dX/dM_j = \sigma\left(\sigma - 1\right)^{-1} X^{\frac{1}{\sigma}} x_j^{\frac{\sigma-1}{\sigma}}/f_j$, where $M_j = n_j f_j$.

When we compare the social optimum in (34) with the market (sorting) equilibrium in (22) it is immediately obvious that $\tilde{z} < \tilde{z}^*$ because $\bar{z}(\tilde{z}) < \tilde{z} \ \forall \ \tilde{z} \in (0, \infty)$. Therefore, the market equilibrium allocates too little employment to low-tech firms, $LG(\tilde{z}) < LG(\tilde{z}^*)$, and too much employment to high-tech firms, $L[1 - G(\tilde{z})] > L[1 - G(\tilde{z}^*)]$. The social optimum and the market equilibrium are illustrated in Figure 3.

[FIGURE 3 here]

The misallocation of labor is due to the fact that workers in low-tech firms receive a wage based on the average productivity of the entire labor pool in low-tech firms, and not on the individual productivity of the marginal worker in that pool. And since the productivity of the average worker in the low-tech labor pool is lower than the productivity of the marginal worker $[\bar{z}_l(\zeta) < \zeta \forall \zeta \in (0, \infty)]$, relative wages in low-tech firms are too low compared to the social optimum $[\bar{w}_h / w_l(\tilde{z}) > \bar{w}_h / w_l(\tilde{z}^*)]$. As a consequence, fewer workers self-select into the low-tech labor pool than what is socially desirable and the mispricing of labor manifests itself in too little employment in low-tech firms. Put differently, while high-tech firms have desirable characteristics (larger economies of scale, higher labor productivity, higher wages, see corollaries 2 to 4), from a social perspective they collectively produce too much. In this sense, there is "too much of a good thing."

It is also interesting to note that the overemployment in high-tech firms is due to too much entry of high-tech firms, and not due to a distortion in their sizes. In fact, we can prove the following corollary:

Corollary 5 *Relative firm sizes are socially optimal in the market equilibrium.*

Proof. By combining equations (7) and (20) we obtain for the relative sizes of firms in the market equilibrium

$$\frac{x_l}{x_h} = \left(\frac{f_l}{f_h}\right)^{\frac{\sigma}{\sigma-1}}. \tag{37}$$

This is the same as (36) and it proves that $x_l / x_h = x_l^* / x_h^*$. ■

This corollary shows that the mispricing of wages in low-tech firms in the market equilibrium leads to too much entry of high-tech firms compared to the social optimum. Thus, the distortion is in the measure of firms, but not in the relative sizes of firms. This insight has an important implication. If relative sizes x_l / x_h are not distorted, the ratio of effective wages $c_h / c_l = (x_l / x_h)^{1/\sigma}$ is also not distorted and reflects true shadow prices. In fact, firms are behaving optimally, workers make distorted decisions. The (non-effective) wages that workers face do not reflect social opportunity costs and make too many of them join the high-tech labor pool.

[FIGURE 4 here]

Figure 4 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 high-tech firms is given by $w(z) = c_h a(z)$ (the thick black line). Workers in high-tech firms are correctly valued so the social valuation is equal to the market valuation. The market valuation of a worker with skill z in low-tech firms is independent of z and given by $w_l = c_l \bar{z}_l(\tilde{z})$ (the thick blue line). The intersection of the two gives the market allocation $\tilde{z}$. However, the social valuation of workers in low-tech firms is not aligned with the market valuation. From a social point of view, workers with skill z in low-tech firms should be valued at $c_l z$ (the thick red line). The socially optimal allocation $\tilde{z}^*$ is then given by the intersection of the two social valuations $c_h a(z)$ and $c_l z$. All workers with skill levels $z < \tilde{z}$ are optimally allocated. They work for low-tech firms, and their social valuation in low-tech firms is higher than the social valuation in high-tech firms: $c_l z > c_h a(z)$. All workers with skill levels $z > \tilde{z}^*$ are also optimally allocated. They work for high-tech firms ($z > \tilde{z}$), and their social valuation in high-tech firms is higher: $c_h a(z) > c_l z$. It is workers with skill levels $z \in (\tilde{z}, \tilde{z}^*)$ that are misallocated. These workers work for high-tech firms ($z > \tilde{z}$), but their social valuation is higher in low-tech firms $[c_l z > c_h a(z)]$. 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 low-tech firms. Since $\tilde{z} < \tilde{z}^*$, it follows that $\partial W / \partial \tilde{z} (\tilde{z} < \tilde{z}^*) > 0$, and a reallocation of labor from high-tech to low-tech firms (an increase in $\tilde{z}$) increases welfare. To see how such a subsidy can increase welfare assume that the government can subsidize employment in low-tech firms and finance this subsidy with a non-distorting per capita tax on income. This changes equations (19) (for $j = l$) and (27):

$$A\left[c_l(1-s)\right]^{1-\sigma} = f_l, \tag{38}$$

$$E = \left[c_l(1-s)\int_{\underline{z}}^{\tilde{z}} z dG(z) + c_h \int_{\tilde{z}}^{\infty} a(z)\, dG(z)\right] L + M = \sigma M, \tag{39}$$

where s is the subsidy rate, $c_l(1-s)$ are after subsidy effective labor costs in low-tech firms, and $s c_l \bar{z}_l(\tilde{z}) G(\tilde{z}) L$ is the total subsidy paid.

The allocation of labor in (22) changes to

$$\frac{\bar{z}_l(\tilde{z})}{a(\tilde{z})} = \Phi(1-s). \tag{40}$$

Since $\bar{z}_l(\tilde{z}) / a(\tilde{z})$ is decreasing in $\tilde{z}$ by (16), an increase in s increases $\tilde{z}$ and the subsidy is

effective in raising employment in low-tech firms.

Since the effective real wage in low-tech firms continues to be pinned down by the free entry condition, $c_l (1-s) P^{-1} = (\sigma - 1) \sigma^{-1} \Xi L^{-1}$, the expression for welfare, equation (31) is unchanged. Thus, the subsidy affects welfare only through the allocation of labor $\tilde{z}$, $dW/ds = (\partial W/\partial \tilde{z})(d\tilde{z}/ds)$, and the optimal subsidy s^* is where $\partial W/\partial \tilde{z} = 0$. We can calculate this optimal subsidy as the percentage difference between the average and the marginal productivity in the low-tech labor pool, evaluated at the optimal allocation:

$$s^* = \frac{\tilde{z}^* - \bar{z}_l(\tilde{z}^*)}{\tilde{z}^*}. \tag{41}$$

Proposition 5 (Subsidy) *There exists an optimal subsidy rate on employment in low-tech 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, low-tech firms. These firms are too small to cover the costs of screening workers, and as a consequence need to pool their workers and pay a wage based on the average productivity of their work force. This strategy allows them to survive, but it creates a misallocation of labor due to the fact that the marginal worker has a higher productivity than the average worker. As a consequence, the employment share of small, low-tech firms is too small compared to the social optimum, and a subsidy on employment in SPF can be welfare improving.[11]

The corrective subsidy raises welfare because it induces a reallocation of workers from high-tech to low-tech firms through the exit of high-tech firms and the entry of low-tech firms. The workers that switch from the high-tech pool to the low-tech pool in response to the subsidy are more skilled than the workers previously working for low-tech firms. As a consequence, 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, Skans and Vikström, 2018).

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

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

4 Open Economy

Let us now consider international trade in an open economy setting with two identical countries. 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.

Profits in the domestic market continue to be given by (9). Profits in the export market are given by

$$\pi_j^x = A\left(\tau c_j\right)^{1-\sigma} - f^x, \tag{42}$$

and total profits of an exporting firm are

$$\pi_j^t = A\left(1+\tau^{1-\sigma}\right) c_j^{1-\sigma} - f_j - f^x, \tag{43}$$

where the index t refers to trade.

Using (19) for the domestic market, profits of exporting can be expressed as $\pi_j^x = \tau^{1-\sigma}\left(f_j - \tau^{\sigma-1} f^x\right)$. Depending on the value of trade costs $\tau^{\sigma-1} f^x$ relative to domestic fixed costs f_h and f_l, we can distinguish between three cases:

- $f_l < f_h < \tau^{\sigma-1} f^x$: No exporting
- $f_l < \tau^{\sigma-1} f^x < f_h$: Exporting only by high-tech firms (Melitz condition)[12]
- $\tau^{\sigma-1} f^x < f_l < f_h$: All firms export

Free entry ensures that profits of exporting firms are also driven down to zero

$$A\left(1+\tau^{1-\sigma}\right) c_j^{1-\sigma} = f_j + f^x. \tag{44}$$

If firms of one type do not export, their free entry condition continues to be given by (19).

By combining free entry for $j \in \{l, h\}$ with the sorting condition in (15) we can determine the market allocation of labor in the open economy:

$$\frac{\bar{z}_l\left(\tilde{z}_t\right)}{a\left(\tilde{z}_t\right)} = \Phi_t, \tag{45}$$

where Φ_t depends on the selection into exporting, and thus ultimately on the value of trade

[12] We refer to this case as the Melitz (2003) condition, because trade costs are higher than (low-tech) manufacturing fixed costs and the equilibrium exhibits sorting into exporting where only some (high-tech) firms export.

costs:

$$\Phi_t^{\sigma-1} \equiv \begin{cases} \frac{f_l\left(1+\tau^{1-\sigma}\right)}{f_h+f^x} & \text{if } f_l < \tau^{\sigma-1} f^x < f_h \\ \frac{f_l+f^x}{f_h+f^x} & \text{if } \tau^{\sigma-1} f^x < f_l < f_h \end{cases} \tag{46}$$

An inspection of (46) reveals that $\Phi_t \in (0,1)$ and, hence, $\underline{z} < \tilde{z}_t < \infty$. Thus, both types of firms co-exist and proposition 2 continues to hold.

By comparing Φ_t in (46) with Φ in (21) we can establish the following proposition:

Proposition 6 (Sorting Threshold) *Trade leads to a fall in the threshold value for sorting, $\tilde{z}_t < \tilde{z}$.*

Proof. From (22) and (45) we know that $\tilde{z}_t < \tilde{z}$ implies that $\Phi_t > \Phi$. If $f_l < \tau^{\sigma-1} f^x$, $\Phi_t > \Phi$ implies that $f_l\left(1+\tau^{1-\sigma}\right)/\left(f_h+f^x\right) > f_l/f_h$, or $f_h > \tau^{\sigma-1} f^x$. If $\tau^{\sigma-1} f^x < f_l$, $\Phi_t > \Phi$ implies that $\left(f_l+f^x\right)/\left(f_h+f^x\right) > f_l/f_h$, or $f_h > f_l$. Both hold. ■

An immediate corollary of proposition 6 is:

Corollary 6 (Employment) *Since $\tilde{z}$ falls, trade raises the employment share in high-tech firms $L\left[1-G\left(\tilde{z}\right)\right]$.*

Proposition 6 and corollary 6 show that a switch from autarky to trade pushes ressources out of low-tech firms and into high-tech firms. The reason for this reallocation of ressources depends on the trade regime. If $f_l < \tau^{\sigma-1} f^x < f_h$, only high-tech firms export. Consequently, only high-tech firms benefit from trade, and expand. The logic behind these reallocations is essentially the same as in Melitz (2003). If $\tau^{\sigma-1} f^x < f_l < f_h$, all firms export, and all firms can benefit from the opening to trade. However, fixed costs in both types of firms increase by the same absolute magnitude (by f^x), so the increase is relatively smaller for high-tech firms than for low-tech firms $\left[\left(f_l+f^x\right)/f_l > \left(f_h+f^x\right)/f_h\right]$. As a consequence, high-tech firms expand.

The reallocation of labor towards high-tech firms also affects relative (real) wages. Using (12), (13) and (15), the relative wage of a worker in a high-tech firm relative to a fellow worker in a low-tech firm is

$$\frac{w_h\left(z\right)}{w_l} = \frac{a\left(z\right)}{a\left(\tilde{z}\right)}, \tag{47}$$

and this is clearly decreasing in $\tilde{z}$. Thus, as trade reduces $\tilde{z}$, it raises relative wages in high-tech firms and lowers relative wages in low-tech firms. The intuition is rather straightforward: Since high-tech firms are hiring and low-tech firms are firing, the change in relative wages simply reflects relative demand for labor in the two types of firms.

Now let us turn to a comparative static analysis of a reduction in variable trade costs. For this we assume that the Melitz condition holds, $f_l < \tau^{\sigma-1} f^x < f_h$, so that Φ_t is given by $\Phi_t^{\sigma-1} = f_l\left(1+\tau^{1-\sigma}\right)/\left(f_h+f^x\right)$, as stated in (46), and $d\ln\Phi_t/d\ln\tau = -\left(1+\tau^{\sigma-1}\right)^{-1} < 0$.

First, we confirm that a continuous trade liberalization has the same effect on the allocation of labor as a discrete trade liberalization. From (45) and (16) we have

$$\frac{d\ln\tilde{z}_t}{d\ln\tau} = \left(1+\tau^{\sigma-1}\right)^{-1}\left[\frac{a'\left(\tilde{z}_t\right)\tilde{z}_t}{a\left(\tilde{z}_t\right)} - \frac{\tilde{z}_t - \bar{z}_l\left(\tilde{z}_t\right)}{\bar{z}_l\left(\tilde{z}_t\right)}\frac{\tilde{z}_t g\left(\tilde{z}_t\right)}{G\left(\tilde{z}_t\right)}\right]^{-1} > 0, \tag{48}$$

so that a reduction in variable trade costs increases the employment share in high-tech firms.

Next, we analyze the effect on real wages. Real wages move in opposite directions in the two types of firms. They are given by

$$\frac{w_l}{P} = \bar{z}_l\left(\tilde{z}_t\right)\frac{\sigma-1}{\sigma}\frac{\Xi}{L} \qquad \frac{w_h\left(z\right)}{P} = a\left(z\right)\Phi_t\frac{\sigma-1}{\sigma}\frac{\Xi}{L} \tag{49}$$

where Ξ is defined as in (32). Relative changes can be calculated as

$$\frac{d\ln\left(w_l/P\right)}{d\ln\tau} = \frac{g\left(\tilde{z}_t\right)\tilde{z}_t}{G\left(\tilde{z}_t\right)}\frac{\tilde{z}_t - \bar{z}_l\left(\tilde{z}_t\right)}{\bar{z}_l\left(\tilde{z}_t\right)}\frac{d\ln\tilde{z}_t}{d\ln\tau} > 0 \tag{50}$$

$$\frac{d\ln\left[w_h\left(z\right)/P\right]}{d\ln\tau} = -\left(1+\tau^{\sigma-1}\right)^{-1} < 0 \tag{51}$$

If the Melitz condition holds, only high-tech firms are exporting and only high-tech firms are benefiting from the reduction in variable trade costs. They expand, their employment share increases, and the increase in demand for labor translates into higher effective real wages (c_h/P increases). Since non-effective real wages, $w_h\left(z\right)/P = a\left(z\right)c_h/P$, depend on the productivity of individual workers, $a\left(z\right)$, and not on the composition of the workforce, they increase by the same percentage. Low-tech firms, on the other side, are not exporting, and their effective real wages are pinned down by the zero profit condition in the domestic market ($c_l/P = \left(\sigma-1\right)\sigma^{-1}\Xi/L$). Hence, effective real wages in low-tech firms are unaffected by changes in variable trade costs. The change in non-effective real wages is driven entire by the change in the composition of the workforce in the low-tech labor pool. Since $\tilde{z}$ falls, the most productive workers in this pool move to the high-tech labor pool, so that the average productivity of workers in the low-tech labor pool falls. As a consequence, their real wages also fall. This is summarized in the following proposition:

Proposition 7 (Real Wages) *Trade raises real wages in high-tech firms/for high-skill workers and lowers real wages in low-tech firms/for low-skill workers.*

Finally, let us turn to welfare. Welfare can be expressed as in (31):

$$W = \Xi\left(\int_{\underline{z}}^{\tilde{z}_t} z dG\left(z\right) + \Phi_t\int_{\tilde{z}_t}^{\infty} a\left(z\right)dG\left(z\right)\right) \tag{52}$$

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

$$\frac{d\ln W}{d\ln\tau}=\left(\frac{\partial\ln W}{\partial\ln\Phi_t}+\frac{\partial\ln W}{\partial\ln\tilde{z}_t}\frac{d\ln\tilde{z}_t}{d\ln\Phi_t}\right)\frac{d\ln\Phi_t}{d\ln\tau} \tag{53}$$

The first effect,

$$\frac{\partial\ln W}{\partial\ln\Phi_t}\frac{d\ln\Phi_t}{d\ln\tau}=-\frac{1-\lambda\left(\tilde{z}_t\right)}{1+\tau^{\sigma-1}}<0, \tag{54}$$

is the direct effect of a reduction in trade costs. Here, $\lambda\left(\tilde{z}_t\right)$ is defined as the labor share in low-tech firms,

$$\lambda\left(\tilde{z}_t\right)\equiv\frac{c_l\int_{\underline{z}}^{\tilde{z}_t}zdG\left(z\right)}{c_l\int_{\underline{z}}^{\tilde{z}_t}zdG\left(z\right)+c_h\int_{\tilde{z}_t}^{\infty}a\left(z\right)dG\left(z\right)}\in\left[0,1\right], \tag{55}$$

where $\lambda\left(\underline{z}\right)=0$ and $\lim_{\tilde{z}_t\to\infty}\lambda\left(\tilde{z}_t\right)=1$. The direct effect is negative, indicating that a reduction in variable trade costs tends to raise welfare. It depends on the labor share in (exporting) high-tech firms, $1-\lambda\left(\tilde{z}_t\right)$, and on the share of trade costs in total costs, $\tau^{1-\sigma}/\left(1+\tau^{1-\sigma}\right)=1+\tau^{\sigma-1}$. The larger these two shares are, the larger is the impact of a reduction in variable trade costs on total costs, and thus the reduction in the price index of consumption.

The second effect,

$$\frac{\partial\ln W}{\partial\ln\tilde{z}_t}\frac{d\ln\tilde{z}_t}{d\ln\Phi_t}\frac{d\ln\Phi_t}{d\ln\tau}=\frac{\lambda\left(\tilde{z}_t\right)\theta\left(\tilde{z}_t\right)}{\left[\frac{a'(\tilde{z}_t)\tilde{z}_t}{a(\tilde{z}_t)}-\theta\left(\tilde{z}_t\right)\right]\left(1+\tau^{\sigma-1}\right)}>0, \tag{56}$$

is the indirect effect working through the reallocation of labor $\tilde{z}_t$, where $d\ln\tilde{z}_t/d\ln\Phi_t$ is calculated from (45). Here, $\theta\left(\tilde{z}_t\right)$ is a measure of the extent of the deviation of the market equilibrium from the social optimum that consists of the relative difference between $\tilde{z}_t$ and $\bar{z}_l\left(\tilde{z}_t\right)$ as well as the relative inverse Mills ratio at $\tilde{z}_t$:

$$\theta\left(\tilde{z}_t\right)\equiv\frac{\tilde{z}_t-\bar{z}_l\left(\tilde{z}_t\right)}{\bar{z}_l\left(\tilde{z}_t\right)}\frac{g\left(\tilde{z}_t\right)\tilde{z}_t}{G\left(\tilde{z}_t\right)}>0. \tag{57}$$

The 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. There is already too much labor in high-tech firms so that $\partial\ln W/\partial\ln\tilde{z}_t>0$ (see figure 3) and an additional boost of employment in high-tech firms through trade liberalization exacerbates the misallocation. This effect depends on the labor share in low-tech firms, $\lambda\left(\tilde{z}_t\right)$, on the size of the misallocation, $\theta\left(\tilde{z}_t\right)$, as well as on the extend of the reallocation of labor $d\ln\tilde{z}_t/d\ln\Phi_t$.

Given that the two effects are counteracting and weighed by the share of labor in the two

types of firms, we can establish the following proposition:

Proposition 8 (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 low-tech firms have a high share in employment, and the welfare increasing effect dominates when trade costs are low and high-tech firms have a high share in emploment.*

Proof. To prove that the welfare reducing effect dominates when trade costs are high we substitute (45) into (52) and show that $W/\Xi = \int_{\underline{z}}^{\tilde{z}_t} zdG(z) + \bar{z}_l(\tilde{z}_t)\, a(\tilde{z}_t)^{-1} \int_{\tilde{z}_t}^{\infty} a(z)\, dG(z)$ approaches its upper limit from below. Define the upper limit (where all labor is employed in type-l firms) as $W_l \equiv \lim_{\tilde{z}_t \to \infty} W(\tilde{z}_t) = \Xi \int_{\underline{z}}^{\infty} zdG(z)$. Then note that $W(\tilde{z}_t) < W_l$ implies that

$$\frac{1}{a(\tilde{z}_t)} \frac{\bar{a}_h(\tilde{z}_t)}{\bar{z}_h(\tilde{z}_t)} < \frac{1}{\bar{z}_l(\tilde{z}_t)}, \tag{58}$$

where $\bar{z}_h(\tilde{z}_t) \equiv \int_{\tilde{z}_t}^{\infty} zdG(z) / [1 - G(\tilde{z}_t)]$. The upper limit on the right hand side of this inequality is $\lim_{\tilde{z}_t \to \infty} [1/\bar{z}_l(\tilde{z}_t)] = 1/\int_{\underline{z}}^{\infty} zdG(z) > 0$. The upper limit on the left hand side is $\lim_{\tilde{z}_t \to \infty} a(\tilde{z}_t)^{-1} \bar{a}_h(\tilde{z}_t) / \bar{z}_h(\tilde{z}_t) = 0$.[13] Hence, inequality (58) holds for sufficiently high values of $\tilde{z}_t$, and the indirect effects dominates the direct effect. ■

There are two explanations for the ambiguity of the 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 high-tech firms and decrease in low-tech firms. The aggregate effect depends on the share of the two types of firms in employment. That is why real income rises when high-tech firms have a large employment share in the economy and falls when low-tech firms have a large emplyoment share.

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 imperfection. The marginal productivity of labor in low-tech firms is discretely higher than in high-tech firms: $dX/dL_h = [\bar{z}_l(\tilde{z}_t)/\tilde{z}_t](dX/dL_l)$ which implies $dX/dL_h < dX/dL_l$ [see (35)

[13] Note that this limit can be expressed as $\lim_{\chi \to \infty} \left(\lim_{\tilde{z}_t \to \chi} a(\tilde{z}_t)^{-1} \int_{\tilde{z}_t}^{\chi} a(z)\, dG(z) / \int_{\tilde{z}_t}^{\chi} zdG(z) \right) = \lim_{\chi \to \infty} \left(a(\chi)^{-1} \times a(\chi)/\chi \right) = \lim_{\chi \to \infty} (1/\chi) = 0$, where $\lim_{\tilde{z}_t \to \chi} \int_{\tilde{z}_t}^{\chi} a(z)\, dG(z) / \int_{\tilde{z}_t}^{\chi} zdG(z) = a(\chi)/\chi$ by L'Hôpital's rule.

and the footnote there]. By reallocating the marginal worker to high-tech 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 $\bar{z}_l(\tilde{z}_t)$ 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 low-tech 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 7 (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 either serve as robustness checks of our main result with respect to specific assumptions of our framework or illustrate how a little more structure is able to generate additional propositions that may be particularly useful when taking this framework to data.

5.1 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, Redding, and Schott (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

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

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

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

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

- $f_l < f_h \leq f^p$: All firms are single-product firms
- $f_l \leq f^p < f_h$: High-tech firms are multiproduct firms, low-tech firms are single-product firms
- $f^p < f_l < f_h$: All firms are multiproduct firms

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

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

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

$$\omega_h > \omega_l \tag{63}$$

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

$$\frac{\bar{z}_l(\tilde{z}_{mpf})}{a(\tilde{z}_{mpf})} = \Phi_{mpf}. \tag{64}$$

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

$$\Phi_{mpf}^{\sigma-1} \equiv \left(\frac{\int_0^{\omega_h} \alpha(\omega)^{1-\sigma} d\omega}{\int_0^{\omega_l} \alpha(\omega)^{1-\sigma} d\omega} \right) \left(\frac{\omega_l f^p + f_l}{\omega_h f^p + f_h} \right) < 1 \tag{65}$$

Since $\omega_h > \omega_l$ and $f_h > f_l$ implies that $\Phi_{mpf} < 1$, $\tilde{z}_{mpf} \in (\underline{z}, \infty)$ and both types of firms coexist. Furthermore, (15) implies that $c_h < c_l$. 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 9 (Multi-product Firms) *High-tech firms produce and export more products.*

High-tech firms pay a lower effective wage rate than low-tech firms ($c_h < c_l$). 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 the screening technology is applicable in all divisions within the firm, so that by adding products to their product range they can lower the fixed costs per product.

In the appendix we provide an additional extension to multiple export destinations that is methodologically very similar to this extension here.

5.2 Multiple Screening Technologies

In this extension we illustrate that our assumption of just two technologies, 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 screening technologies that lead to a higher degree of firm heterogeneity.

Suppose there are N screening technologies $j \in \{0, ..., N\}$ with labor productivity $a_j(z)$, where $a_0(z) = z$, $a_{j+1}(z) > a_j(z)$ $\forall z > \underline{z}$, $a'_{j+1}(z) > a'_j(z) \geq 1$, $a''_j(z | z > \underline{z}) > 0$. A screening technology j allows firms to screen 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. Technologies are ranked by the skills they can observe ($\breve{z}_j$), so that $\breve{z}_{j+1} > \breve{z}_j$. Technology $j = 0$ is essentially our low-tech technology in previous sections and implies no screening ($\breve{z}_0 = \underline{z}$). A better screening technology (with a higher $\breve{z}_j$) is more expensive and requires strictly higher fixed costs, so that $f_{j+1} > f_j > 0$.[14]

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

We consider only efficient screening technologies so that no screening technology exists that is characterized by the same level of observable skills but higher fixed costs.

Given that a firm with screening technology j (type-j firm) can only observe skills up to level $\breve{z}_j$, it can only pay a wage based on the worker's true productivity up to this level. Above this level, it has to pay a wage based on the expected productivity.[15] Hence, the wage paid by a type-j firm to workers with skill z depends on the skill of these workers:

$$w_j(z) = c_j \times \begin{cases} a_j(z) & \text{for} \quad \tilde{z}_j < z < \breve{z}_j < \tilde{z}_{j+1} \\ \bar{a}_j(\breve{z}_j, \tilde{z}_{j+1}) & \text{for} \quad \tilde{z}_j < \breve{z}_j < z < \tilde{z}_{j+1} \end{cases}, \tag{66}$$

where $\bar{a}_j(z_1, z_2) \equiv \int_{z_1}^{z_2} a_j(z)\, dG(z) / [G(z_2) - G(z_1)]$. 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 5 here]

The profile of wages in type-j firms is illustrated in Figure 5. 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: $w_j(z) = c_j a_j(z)$. Above $\breve{z}_j$, these firms only know the average productivity of their workforce because they use a technology that does not allow them to screen workers with these skills. As a consequence, the wage is based on $\bar{a}_j$: $w_j(z) = c_j \bar{a}_j(\breve{z}_j, \tilde{z}_{j+1})$. 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 $a_j(\breve{z}_j) < \bar{a}_j(\breve{z}_j, \tilde{z}_{j+1})$, i.e. that the productivity of the worker with the exact observable skill $\breve{z}_j$ is discretely lower than the average productivity of workers in the range $(\breve{z}_j, \tilde{z}_{j+1})$. This jump is ultimately due to the fact that we are assuming that firms know the distribution of skills in their labor pool and can therefore pay wages based on expected productivity. If firms did not know this distribution, they could only pay wages based on the highest skill they observe, i.e. $w_j(z : z \in (\breve{z}_j, \tilde{z}_{j+1})) = c_j a_j(\breve{z}_j)$. In this case there is no jump and the wage profile is a continuous function of skills. This alternative assumption has no qualitative effect on our results and even increases the inefficiency described earlier because $a_j(\breve{z}_j) < \bar{a}_j(\breve{z}_j, \tilde{z}_{j+1})$.

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

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

firms of type j and $j-1$, this sorting condition is

$$\frac{c_j}{c_{j-1}} = \frac{\bar{a}_{j-1}\left(\breve{z}_{j-1}, \tilde{z}_j\right)}{a_j\left(\tilde{z}_j\right)} \tag{67}$$

If this condition is fulfilled, workers with a marginal 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 $c_{j-1}\bar{a}_{j-1}\left(\breve{z}_{j-1}, \tilde{z}_j\right)$. In the type-$j$ labor pool their skill is below the threshold of observability, so firms with technology j will pay them a wage based on their true productivity of $c_j a_j\left(\tilde{z}_j\right)$. We assume that condition (16) holds locally, i.e. that $\bar{a}_{j-1}\left(\breve{z}_{j-1}, \zeta\right)/a_j\left(\zeta\right)$ is decreasing in ζ at $\zeta = \tilde{z}_j$.

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

$$\frac{\bar{a}_{j-1}\left(\breve{z}_{j-1}, \tilde{z}_j\right)}{a_j\left(\tilde{z}_j\right)} = \Phi_j, \tag{68}$$

where

$$\Phi_j^{\sigma-1} \equiv \frac{f_{j-1}}{f_j} < 1, \tag{69}$$

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

$$\frac{\bar{w}_j}{\bar{w}_{j-1}} = \frac{\varphi_j}{\varphi_{j-1}} = \frac{\bar{a}_{j-1}\left(\breve{z}_{j-1}, \tilde{z}_j\right)}{\bar{a}_{j-1}\left(\tilde{z}_j, \tilde{z}_{j+1}\right)} \frac{\bar{a}_j\left(\tilde{z}_j, \tilde{z}_{j+1}\right)}{a_j\left(\tilde{z}_j\right)} > 1 \tag{70}$$

Proposition 10 (Multiple Screening Technologies) *With multiple screening technologies, sorting leads to multiple labor pools where all efficient screening technologies can coexists. Firms with a better screening technology (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 screening technologies is similarily distorted as the equilibrium with only two technologies. 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 technology $j > 0$, employment in firms with a lower productivity is too low and in firms with a higher productivity 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 technologies 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,

$$\lim_{\tilde{z}_{j+1} \to \breve{z}_j} \bar{a}_j \left(\breve{z}_j, \tilde{z}_{j+1} \right) = a_j \left(\breve{z}_j \right), \tag{71}$$

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 screening technology, and effective wages are equalized across screening technologies. As a consequence, the inefficiency associated with pooling also disappears.

Proposition 11 (Continuous Screening Technologies) *With a continuous screening technology, there is no more pooling in the labor market and the inefficiency associated with pooling disappears.*

5.3 Exogenous Firm Heterogeneity

In section 3, we showed in our benchmark model with endogenous firm entry into h and l technologies that the distortion created by imperfect information manifested itself as excess worker sorting into the screening pool. The resulting labor supply misallocation meant that there were too many firms using the h technology from the perspective of a social planner but free entry had the implication that both h and l 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 $i \in \{h, l\}$, is exogenously fixed.[16] We will show that when firms are exogenously heterogeneous that the distortion created by imperfect information manifests itself as h-type firms that are too large relative to the social optimum and l-type firms that are too small. In this case, from a social point of view it would be better if bigger firms were smaller.

We will also show that naive measures of total factor productivity - revenue (TFP-R) that fail to account for variation in worker quality should be higher at h-type firms than at l-type firms. In this situation the reallocation of resources from low to high measured TFP-R firms would worsen the distortion. This result highlights the need to know the source of observed exceptional firm performance before drawing policy conclusions in general and in the need to properly control for worker composition in the calculation of TFP-R in specific.

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

$$x_i = \tilde{L}_j / n_j. \tag{72}$$

[16] 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 $n_h/(n_h + n_l)$.

Consequently, relative sizes are given by

$$\frac{x_l}{x_h} = \frac{n_h}{n_l} \frac{G(\tilde{z})\,\bar{z}_l(\tilde{z})}{[1 - G(\tilde{z})]\,\bar{a}_h(\tilde{z})} \tag{73}$$

Given (7) and using ζ to represent the ability of the cutoff worker, we show that the relative labor market equilibrium can be written

$$\frac{c_h}{c_l} = \left(\frac{n_h}{n_l} \frac{\int_{\underline{z}}^{\zeta} \zeta dG(\zeta)}{\int_{\zeta}^{\infty} a(\zeta) dG(\zeta)} \right)^{\frac{1}{\sigma}}. \tag{74}$$

Equation (74) is depicted in figure 6 as an increasing locus in $(c_h/c_l, \zeta)$ space ("Labor Market Equilibrium"). It is increasing because a higher cutoff level ζ implies a higher relative supply in the type-l labor pool and requires in equilibrium a lower relative effective wage in that pool to boost relative labor demand.

[Figure 6 here]

As in the benchmark model, the "Sorting Condition" is given by $c_h/c_l = \overline{z}_l(\zeta)/a(\zeta)$ and is decreasing by condition (16). In the market equilibrium, the cutoff skill is given by $\tilde{z}$ and the relative effective wage facing producers is $c_h/c_l = \overline{z}_l(\tilde{z})/a(\tilde{z})$.

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. $c_h^*/c_l^* = \zeta/a(\zeta)$. This condition is shown as a decreasing curve in figure 6. Crucially, this locus lies everywhere above the "Sorting Condition" curve [since $\zeta > \overline{z}_l(\zeta)$]. It follows immediately from the figure that too much labor is allocated to the h-type labor pool, i.e. $\tilde{z} < \tilde{z}^*$, and the effective cost of labor at h-type firms is too low relative to l-type firms, i.e. $c_h/c_l(\tilde{z}) < c_h^*/c_l^*(\tilde{z}^*)$. The latter result contrasts with the benchmark model where c_h/c_l is flat at its optimal level due to the technologically imposed free entry condition. With exogeneous firm heterogeneity, a low equilibrium $\tilde{z}$ reduces c_h/c_l in order to make h-type firms grow to absorb the excess labor and to induce l-type firms to shrink to conserve labor. The following proposition follows immediately from figure 6.

Proposition 12 (Optimal Firm Size) *Without free entry, high-tech 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.[17]

According to this proposition, a reallocation of labor from h-type firms to l-firms would raise aggregate welfare. We now show that our model has the implication that this requires that resources be reallocated away from firms that exhibit high TFP-R to firms that exhibit low TFP-R when TFP-R measures do not control for worker heterogenity. We show this by calculating TFP-R for each firm type when labor input is simply measured as the count of workers employed by each firm type. We then show that if labor inputs are measured appropriately in efficiency units that the TFP-R rankings of the two firm types switch.

As in section 2.2.4 let φ_i be the revenue per worker of the average i-type firm. In the one factor model considered here, this corresponds to measured TFP-R of firms of type i. As shown in corollary 3 and reproduced here for convenience, naive TFP-R measurement (i.e. does not control for ability composition) yields

$$\varphi_i = \frac{\sigma}{\sigma - 1} c_i \bar{a}_i (\tilde{z}) \tag{75}$$

These expressions show that TFP-R can be decomposed into three terms: mark-ups ($\frac{\sigma}{\sigma-1}$), variations in the cost (c_i) and TFP-Q [$\bar{a}_i (\tilde{z})$].[18] Input quality adjusted TFP-R measurement yields

$$\varphi_i^{cor} = \frac{\sigma}{\sigma - 1} c_i \frac{\bar{a}_i (\tilde{z})}{\bar{z}_i (\tilde{z})}, \tag{76}$$

where $\bar{z}_h (\tilde{z}) \equiv \int_{\tilde{z}}^{\overline{z}} z dG (z) / [1 - G (\tilde{z})]$. The appropriate correction requires measuring $\bar{z}_i (\tilde{z})$ for firms of type i. Matched employer-employee data makes this correction possible. Given such data, $\bar{z}_i (\tilde{z})$ could be approximated using the approach of Abowd, Kramarz, and Margolis (1999).[19]

The ratio of unadjusted and input quality adjusted TFP-R measurements is

$$\frac{\varphi_h}{\varphi_l} = \frac{c_h}{c_l} \frac{\bar{a}_h (\tilde{z})}{\bar{z}_l (\tilde{z})} \qquad \text{and} \qquad \frac{\varphi_h^{cor}}{\varphi_l^{cor}} = \frac{c_h}{c_l} \frac{\bar{a}_h (\tilde{z})}{\bar{z}_h (\tilde{z})}. \tag{77}$$

Based on these ratios we can establish the following proposition:

Proposition 13 (TFP-R at Social Optimum) *At the optimal allocation $\tilde{z}^*$, adjusted and non-adjusted TFP-R measurements are* not *equalized. TFP-R in type-h firms is always larger than in type-l firms.*

[17] Were we to introduce multiproduct firms as in section 5.1, h-type firms would also opt to manage an excessive number of product lines while l-type firms would be inefficiently proscribed in their product offerings.

[18] Note that $a_l (z) = z$ and hence $\bar{a}_l (\tilde{z}) = \bar{z}_l (\tilde{z})$.

[19] Alternatively, correcting TFP-R using wage data should remove variation in TFP-R across firms.

Proof. The social optimum implies $(c_h/c_l)^* = \tilde{z}^*/a(\tilde{z}^*)$. The ratio of adjusted TFP-R measurements is then $(\varphi_h^{cor}/\varphi_l^{cor})^* = [\tilde{z}^*/a(\tilde{z}^*)]\,\bar{a}_h(\tilde{z}^*)/\bar{z}_h(\tilde{z}^*) > 1$ (where $\bar{a}_h(\tilde{z}^*)/\bar{z}_h(\tilde{z}^*) > a(\tilde{z}^*)/\tilde{z}^*$ follows from $a', a'' > 0$). The ratio of unadjusted TFP-R measurements is $(\varphi_h/\varphi_l)^* = [\tilde{z}^*/a(\tilde{z}^*)]\,\bar{a}_h(\tilde{z}^*)/\bar{z}_l(\tilde{z}^*) > (\varphi_h^{cor}/\varphi_l^{cor})^* > 1$ (which follows immediately from $\bar{z}_h(\tilde{z}^*) > \bar{z}_l(\tilde{z}^*)$). ■

This proposition indicates that adjusted and non-adjusted TFP-R measurements are not equalized, even when the allocation of workers is socially optimal. This is important to recognize because it implies that the quantitative analyses of misallocation that are done in the spirit of Hsieh and Klenow (2009) may draw inappropriate conclusions in the presence of worker heterogeneity and adverse selection. In this case, differences in TFP-R measurements, whether adjusted for input quality or not, are neither a necessary nor a sufficient condition for a misallocation. The fact the even the adjusted TFP-R measurements are not equalized is due to the fact that there are increasing returns to skills in the high-tech technology ($a''_H(z) > 0$) and an adjustment by average input quality overadjusts low quality inputs and underadjusts high-quality inputs.

6 Conclusion

The share of employment in large exporting firms is rising, and international trade is playing an important role in this process. We show in this paper that to assess the implications of these reallocations requires a knowledge of exactly why these firms are large. If differences in firm size are driven by (exogenous) differences in social efficiencies across firms (as in most of the heterogeneous firms literature), then these reallocations may improve aggregate productivity and lead to a more efficient allocation of resources. However, if differences in firm size are driven by market failures based on factor market imperfections (as emphasized by a large labor market literature), then these reallocations can exacerbate the misallocation of resources and reduce overal welfare. Our study highlights the need to know why firms "excel" before drawing welfare conclusions and policy recommendations regarding cross firm reallocations of resources.

In our model ex post differences in firm size and productivity are generated by different strategies to overcome labor market imperfections by ex ante identical firms. The resulting industry structure exhibits many facts that are consistent with recent empirical findings: Larger firms are more productive, more likely to export, produce more products, pay higher wages and employ a more productive workforce. However, the resulting equilibrium is inefficient because too much labor is employed in large firms. This has two important implications: First, there is an incentive to subsidize small firms, and second, a reallocation of labor to-

wards large firms has a first order negative welfare effect that can even dominate traditional gains from trade.

In our framework the superior human resources management capabilities of large firms are a mixed blessing from the point of view of efficient resource allocation. On the one hand, because knowledge of workers' skills is necessary to use a technology adapted for skilled workers, human resource management capabilities allow skilled labor to be used more efficiently. On the other hand, the market power conferred on large, multiproduct firms artificially lowers their labor costs and induces too much entry of large, exporting firms. In such a world, subsidization of employment at small, non-export oriented firms is optimal and gains from trade liberalization can only be ensured given a proper subsidy.

We have analyzed only one type of factor market distortion that can give large firms an advantage relative to smaller firms. In an environment in which larger firms are better equipped to influence government policy, it is likely that there are other, perhaps more pernicious forces, that induce large firms to be too large from a social point of view. We hope that this will become a vibrant area of research.

7 Appendix

7.1 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 (21) changes to

$$\Phi_\alpha \equiv \frac{\alpha_l}{\alpha_h}\left(\frac{f_l}{f_h}\right)^{\frac{1}{\sigma-1}}. \tag{78}$$

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

$$\frac{\alpha_l}{\alpha_h} < \left(\frac{f_h}{f_l}\right)^{\frac{1}{\sigma-1}}. \tag{79}$$

Since $f_h > f_l$, $\alpha_l = \alpha_h$ is a sufficient condition for co-existence.

But co-existence 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:

$$(\sigma - 1)(\ln \alpha_l - \ln \alpha_h) \leq (\ln f_h - \ln f_l). \tag{80}$$

One interesting case is where the investment f_h can be interpreted as an innovation that does not only allow firms to use the high-tech technology, but also reduces unit labor requirements. Suppose

$$\alpha_j = f_j^{-\rho}. \tag{81}$$

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

$$\rho < (\sigma - 1)^{-1}. \tag{82}$$

7.2 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:

$$\pi_j^{mx} = 2\left(Ac_j^{1-\sigma}\int_0^{\delta_j}\tau(\delta)^{1-\sigma}\,d\delta - \delta_j f^x\right) \tag{83}$$

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

$$\frac{d\pi_j^{mx}}{d\delta_j} = 2\left(Ac_j^{1-\sigma}\tau(\delta_j)^{1-\sigma} - f^x\right) = 0. \tag{84}$$

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:

- $f_l < f_h < f^x$: No exporting
- $f_l < f^x < f_h$: Exporting only by high-tech firms
- $f^x < f_l < f_h$: All firms export

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

$$\pi_j^{mx} = A\left(1 + 2\int_0^{\delta_j} \tau(\delta)^{1-\sigma}\, d\delta\right) c_j^{1-\sigma} - 2\delta_j f^x - f_j, \tag{85}$$

where mx stands for $m$ultiple e$x$port destinations. Combining free entry and (84) yields

$$\tau(\delta_j)^{\sigma-1}\left(1 + 2\int_0^{\delta_j} \tau(\delta)^{1-\sigma}\, d\delta\right) - 2\delta_j = \frac{f_j}{f^x} \tag{86}$$

Since the left hand side of (86) is increasing in δ_j and the right hand side is increasing in f_j, $f_h > f_l$ implies

$$\delta_h > \delta_l. \tag{87}$$

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

$$\frac{\bar{z}_l(\tilde{z}_{mx})}{a(\tilde{z}_{mx})} = \Phi_{mx}. \tag{88}$$

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

$$\Phi_{mx}^{\sigma-1} \equiv \left(\frac{1 + 2\int_0^{\delta_h} \tau(\delta)^{1-\sigma}\, d\delta}{1 + 2\int_0^{\delta_l} \tau(\delta)^{1-\sigma}\, d\delta}\right)\left(\frac{2\delta_l f^x + f_l}{2\delta_h f^x + f_h}\right) < 1 \tag{89}$$

Since $\delta_h > \delta_l$ implies that $\Phi_{mx} < 1$, $\tilde{z}_{mx} \in (\underline{z}, \infty)$ and both types of firms coexist. Furthermore, (15) implies that $c_h < c_l$. Thus, the misallocation and the welfare effects are qualitatively identical.

Proposition 14 (Multiple Export Destinations) *High-tech firms export to more destinations and ship longer distances.*

References

[1] Abowd, John M., Francis Kramarz, and David N. Margolis. 1999. "High wage workers and high wage firms." *Econometrica* 67.2: 251-333.

[2] 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.

[3] 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." Working Paper

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

[5] Becker, Sascha, Hartmut Egger, Michael Koch, and Marc-Andreas Muendler. 2017. "Tasks, Occupations, and Wage Inequality in an Open Economy." Mimeo University of Bayreuth.

[6] Bender, Stefan, Nicholas Bloom, David Card, John Van Reneen, and Stephanie Wolter. 2016. "Management Practices, Workforce Selection and Productivity." NBER Working Paper 22101, NBER Working Paper Series: Cambridge.

[7] 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.

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

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

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

[11] 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.

[12] Card, David, Jörg Heining, and Patrick Kline. 2013. "Workplace heterogeneity and the rise of West German wage inequality." *The Quarterly Journal of Economics* 128(3): 967-1015.

[13] 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.

[14] Costinot, Arnaud, Andres Rodriguez-Clare, and Ivan Werning. 2016. "Micro to Macro: Optimal Trade Policy with Firm Heterogeneity." NBER Working Paper No. 21989, NBER Working Paper Series: Cambridge.

[15] Criscuolo, C., P. Gal, and C. Menon. 2014. "The Dynamics of Employment Growth: New Evidence from 18 Countries." OECD Science, Technology and Industry Policy Papers No. 14. OECD Publishing: Paris.

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

[17] Dhingra, Swati and John Morrow. 2016. "Monopolistic Competition and Optimum Product Diversity Under Firm Heterogeneity." *Journal of Political Economy* (forthcoming)

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

[19] 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.

[20] Egger, Hartmut, and Udo Kreickemeier. 2009. "Firm heterogeneity and the labor market effects of trade liberalization." *International Economic Review* 50.1: 187-216.

[21] 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..

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

[23] Friedrich, Benjamin. 2017. "Internal Labor Markets and the Competition for Managerial Talent." Mimeo Northwestern University.

[24] Greenwald, Bruce. 1986. "Adverse Selection in the Labour Market." *Review of Economic Studies* 53: 325-347.

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

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

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

[28] Hsieh, Chang-Tai, and Peter J. Klenow. 2009. "Misallocation and manufacturing TFP in China and India." *The Quarterly Journal of Economics* 124.4: 1403-1448.

[29] Hsieh, Chang-Tai and Peter J. Klenow. 2017. "The Reallocation Myth." Mimeo Stanford University.

[30] Hurst, Erik and Benjamin W. Pugsley. 2011. "What do small businesses do?" *Brookings Papers on Economic Activity* No. 2 (Fall): 73-142

[31] Irarrazabal, Alfonso, Andreas Moxnes and Karen Helene Ulltveit-Moe. 2013. "Heterogeneous Firms or Heterogeneous Workers? Implications for Exporter Premiums and the Gains from Trade." *Review of Economics and Statistics* 95(3): 839-849.

[32] Krueger, Alan B. 2018. "Reflections on Dwindling Worker Bargaining Power and Monetary Policy." Luncheon Address at the Jackson Hole Economic Symposium. August 24, 2018. https://www.kansascityfed.org/~/media/files/publicat/sympos/2018/papersandhandouts/824180824kruegerremarks.pdf?la=en

[33] Krugman, Paul. 1980. "Scale Economies, Product Differentiation, and the Pattern of Trade." *American Economic Review* 70: 950-959.

[34] 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.

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

[36] Manning, Alan. 2011. "Imperfect Competition in the Labor Market." In: Card, David and Orley Ashenfelter (eds.). Handbook of Labor Economics (Chapter 11). Volume 4b. Elsevier: 973-1041.

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

[38] Melitz, Marc J. and Stephen J. Redding. 2015. "New Trade Models, New Welfare Implications." *American Economic Review* 105(3): 1105-1146

[39] Mirrlees, James, Stuart Adam, Tim Besley, Richard Blundell, Stephen Bond, Robert Chote, Malcolm Gammie, Paul Johnson, Gareth Myles and James M. Poterba. 2011. "Small business taxation." in: Tax By Design: the Mirrlees Review. Oxford University Press: Oxford, pp. 451–469 (Chapter19).

[40] Nocco, Antonella, Gianmarco Ottaviano, and Matteo Salto. 2017. "Geography, Competition, and Optimal Multilateral Trade Policy." Mimeo London School of Economics.

[41] 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.

[42] Restuccia, Diego and Richard Rogerson. 2013. "Misallocation and productivity." *Review of Economic Dynamics* 16: 1-10.

[43] 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/

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

[45] The World Trade Organization (WTO). 2016. "The World Trade Report 2016: Levelling the trading fields for SMEs." WTO Publications: Geneva.

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

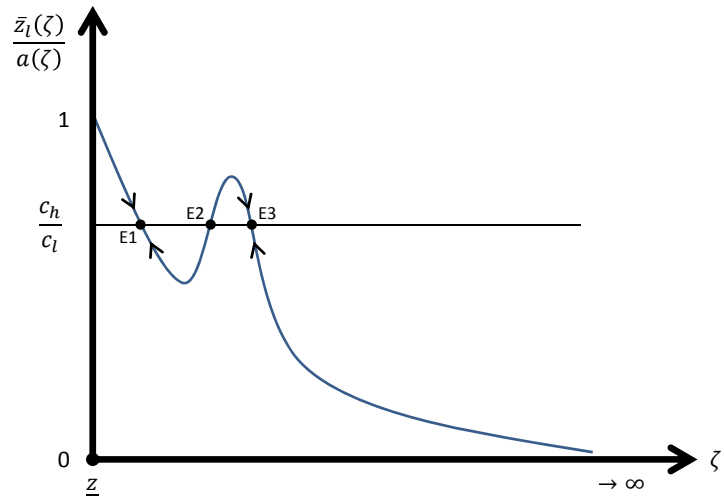

Figure 1: Stability of Sorting Equilibrium

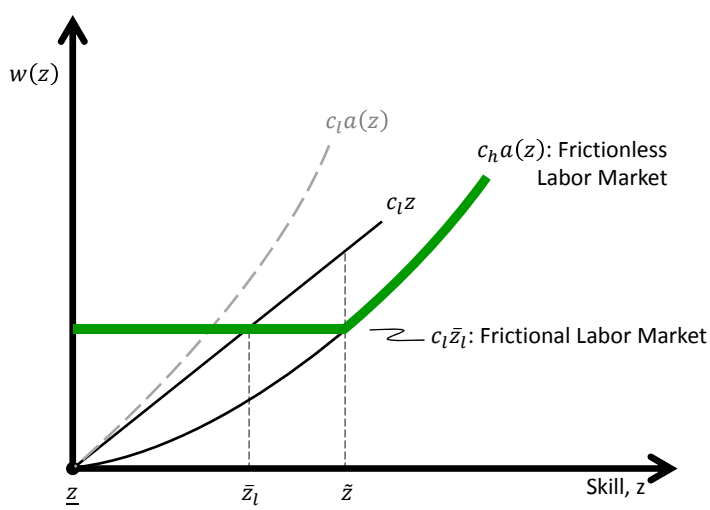

Figure 2: Hockey Stick Wage Profile

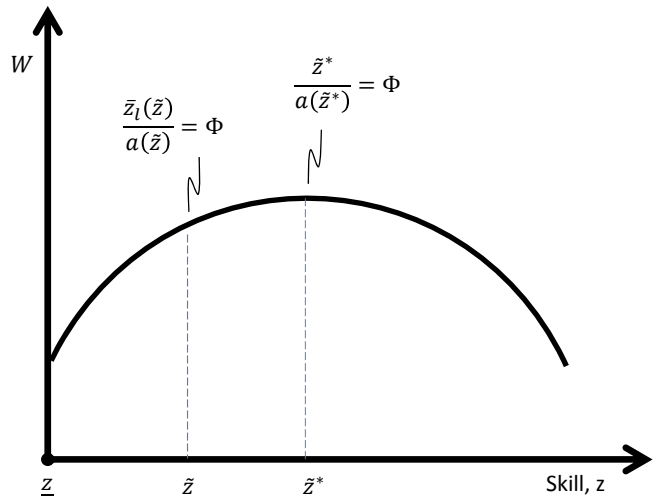

Figure 3: Sorting and Welfare

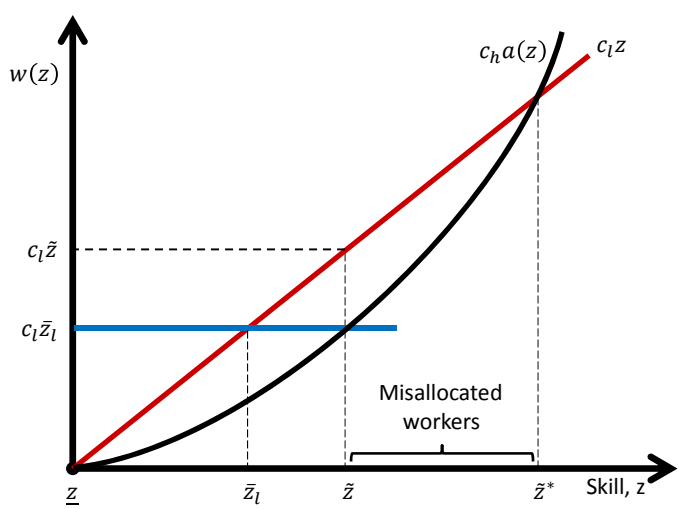

Figure 4: Allocation of Workers

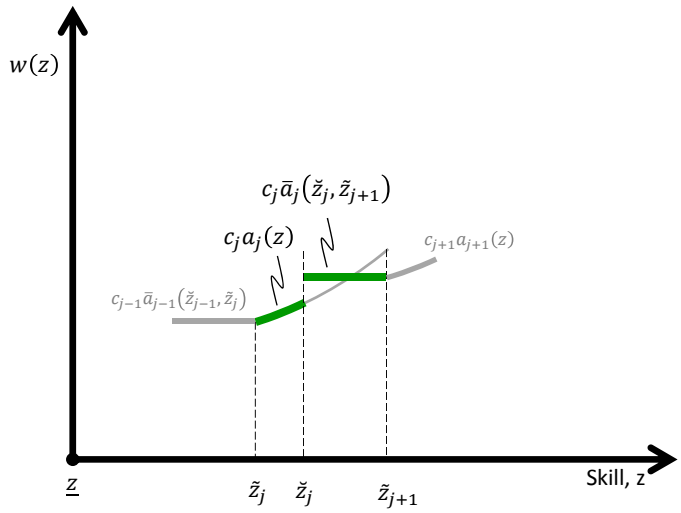

Figure 5: Multiple Screening Technologies

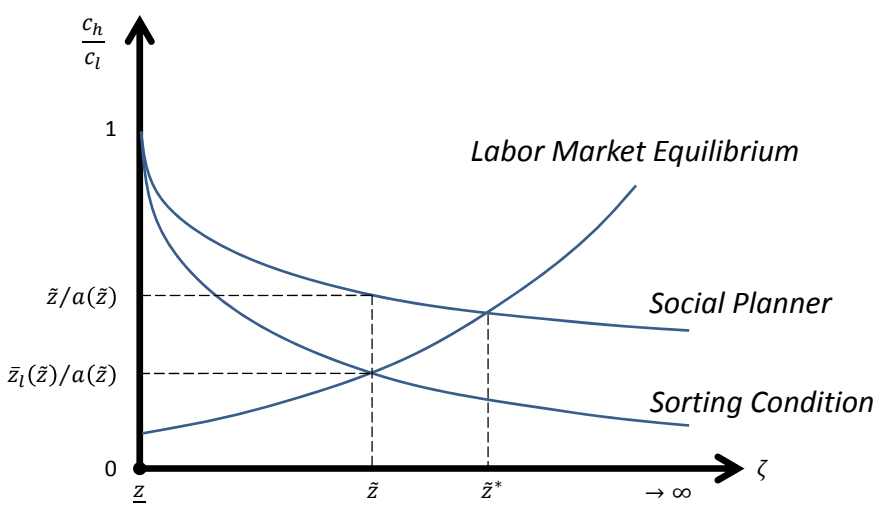

Figure 6: Exogenous Heterogeneity