DISCUSSION PAPER SERIES

No. 9341

HETEROGENEOUS WORKERS AND INTERNATIONAL TRADE

Gene Grossman

INTERNATIONAL TRADE AND REGIONAL ECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: www.cepr.org/pubs/dps/DP9341.asp

ISSN 0265-8003

HETEROGENEOUS WORKERS AND INTERNATIONAL TRADE

Gene Grossman, Princeton University and CEPR

Discussion Paper No. 9341
February 2013

Centre for Economic Policy Research
77 Bastwick Street, London EC1V 3PZ, UK
Tel: (44 20) 7183 8801, Fax: (44 20) 7183 8820
Email: cepr@cepr.org, Website: www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programme in **INTERNATIONAL TRADE AND REGIONAL ECONOMICS**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

CEPR Discussion Paper No. 9341

February 2013

ABSTRACT

Heterogeneous Workers and International Trade*

In this paper, I survey the recent theoretical literature that incorporates heterogeneous labor into models of international trade. The models with heterogeneous labor have been used to study how talent dispersion can be a source of comparative advantage, how the opening of trade affects the full distribution of wages, and how trade affects industry productivity and efficiency via its impact on sorting and matching in the labor market. Some of the most recent contributions also introduce labor market frictions to study the effects of trade on structural unemployment and on mismatch between workers and firms.

JEL Classification: F11 and F16
Keywords: heterogeneous labor, international trade, matching, productivity, sorting and wage distribution

Gene Grossman
Princeton University
International Economics Section
300 Fisher Hall, Prospect Avenue
Princeton, NJ08544-1021
USA

Email: grossman@princeton.edu

*Based on the Bernhard Harms Prize Lecture delivered at the Kiel Institute of World Economics on October 31, 2012. Part of the work on this lecture was completed when I was visiting STICERD at the London School of Economics and Centre de Recerca en Economia Internacional (CREI) in Barcelona. I thank both institutions for their hospitality and support. I am grateful to Arnaud Costinot, Elhanan Helpman, Stephen Redding, and Esteban Rossi-Hansberg for their comments and suggestions.

Submitted 01 February 2013

1 Introduction

Traditionally, trade theory has addressed the causes and consequences of international trade using models featuring broad factor aggregates. Distinctions typically are drawn between capital and labor or between skilled and unskilled labor, with land and other natural resources also distinguished in some applications. In the familiar theories, such as the Heckscher-Ohlin model, differences in aggregate factor endowments drive comparative advantage, as some countries are assumed to be relatively better endowed with capital versus labor or with skilled versus unskilled labor. The Stolper-Samuelson theorem and the Ricardo-Viner comparative statics speak to the distributional effects of trade and trade liberalization in terms of profits versus wages or the skill premium for highly-educated versus less-educated workers. When these models are used to address the efficiency of resource allocation, again the aggregates feature prominently in discussions of employment versus unemployment or the sectoral misallocation of labor.

Recently, trade economists have gained access to an array of firm-level data and, in particular, data that link workers to their employers. The matched employer-employee data allow researchers to gather information about the career experiences of individual workers including their education, years on the job, demographic characteristics and wage history, and also information about the firms for which they work, including their longevity, investments, employment levels and composition, industries of operation, and their output and export experiences. With these rich data, it becomes possible to study the sources of comparative advantage and the income distributional impacts of trade at much finer levels of detail than ever before. For example, one can ask whether and how a country's *distribution* of factors distinguished by quality or skills affects its pattern of specialization and trade. One can ask how globalization affects compensation not only in terms of the skill premium, but over the entire spectrum of the wage distribution. And one can ask how trade affects a country's productivity by influencing the matching of heterogeneous workers and heterogeneous firms and the formation of diverse production teams.

In this paper, I will survey an emerging theoretical literature that introduces worker heterogeneity into models of international trade. These are papers that typically assume a continuum of different types of labor. They ask, how do these heterogeneous workers *sort* to the different industries in the economy and how do they *match* in their chosen industry of employment with other workers, with firms that differ in technologies, and with other factors of production. They also ask how the opening of trade affects the wage distribution and, in some cases, how the sorting and matching of workers compares to the efficient allocation of resources.

A word on terminology is in order. Throughout this survey, I use "sorting" to refer to the allocation of resources *between* sectors of the economy and "matching" to refer to the allocation of resources *within* sectors. This terminology is not standard in the literature where, in fact, the words sorting and matching are often used interchangeably. In some general, theoretical sense there is no conceptual difference between a worker's decision about the industry in which to seek employment and the choice of firm within an industry. Firms have intangible assets that allow them to produce output and a particular combination of assets can make a firm especially productive

within its chosen industry or relatively productive in one line of business compared to another. Yet, I find it useful when thinking about international trade to distinguish resource allocation between and within industries. Many policies (and most trade policies) are levied at the industry level and apply to all firms in an industry. So, when considering international trade, we are used to thinking about the determinants of resource allocation and the sources of comparative advantage at the industry level. Moreover, much of the data on outputs and wages is reported at the industry level. So, for example, we know about the aggregate output of firms in particular industries but not the aggregate output of firms of certain types, and we often hear observations about between-industry and within-industry wage inequality. While there is no strong conceptual distinction between sorting to industries and matching within industries, the determinants of each draw on slightly different parts of the trade literature and they speak to slightly different features of the data.

I begin in Section 2 by introducing a production function that relates output of some good in some country to the *number* of workers and the *type* of workers employed by a firm there, to the type of the firm or of the other factors with whom the workers are matched, and to the quantities of any other inputs that may be used by the firm. This production function is sufficiently general to nest assumptions about technology in all of the papers that I review here and it provides an organizing principle for the subsequent discussion.

Complementarities are important for the sorting and matching of heterogeneous labor. Workers will sort to industries and choose occupations whose requirements are complementary to their skills and they will match with other workers and with technologies and other factors that enhance their own productivity. In Section 3, I provide a brief digression on supermodularity and log supermodularity, two mathematical concepts that capture the notions of complementarity that feature most prominently in the literature.

The literature survey begins in Section 4. There, I describe models that feature sorting of workers to industries but no matching of workers within industries. In that section, I discuss papers that assume constant returns to scale in the number of workers (i.e., linear production technologies), whereas in Section 5 I consider models that introduce diminishing returns to the quantity of labor. Matching does not occur in the papers reviewed in Sections 4 and 5, either because labor is the only factor of production (Section 4) or because all firms in an industry are assumed to be similar to one another and all other productive factors are assumed to be homogeneous (Section 5).

Section 6, in contrast, describes models with matching but no sorting. Here, workers match with firms that use different technologies or with managers that have different abilities. The authors of these papers abstract from issues of sorting by assuming that the economy comprises a single sector. Finally, in Section 7, I consider an economy with multiple industries and heterogeneous firms within each industry. Then, the heterogeneous workers potentially sort to the different sectors and they match with particular managers or other factors within each sector. Sorting and matching are interdependent choices here, inasmuch as a worker's choice of sector depends on what other types of factors he will be able to match with in each industry and the matching options for workers in a sector depend on the sorting decisions made by these other factors.

Most of the literature until now has studied economies with frictionless labor markets. However, the efficiency issues that arise in conjunction with the matching and sorting of heterogeneous workers become especially interesting when search frictions preclude instantaneous matching with heterogeneous firms and when imperfect information about worker types prevents firms from hiring the types of workers that might be most appropriate for their technologies and other attributes. In Section 8, I describe a few recent papers that incorporate either search frictions and unemployment or imperfect information and mismatch.

This survey focuses on the theoretical literature. However, before beginning that task, I set the stage by briefly describing a few empirical observations that have motivated some of the recent theorizing. This is not intended to serve as a comprehensive or systematic review of the empirical literature, but rather to provide a bit of context for the modeling efforts that I shall describe. I offer the following five motivating observations.

(i) Skill dispersion can be a source of comparative advantage

Bombardini et al. (2012) use micro data from the International Adult Literacy Survey (IALS) to document the existence of substantial cross-country differences in skill dispersion. The IALS provides internationally comparable measures of worker-related skills for 19 countries. The authors show that cross-country differences in the standard deviation of scores on standardized exams are 1.6 times as large as are the cross-country difference in mean scores. They proceed to establish that differences in bilateral trade flows by industry are correlated not only with cross-country differences in the means of IALS scores but also with differences in the variances. Using data for the year 2000 on trade flows between 19 exporters and 145 importers in 63 industries, they find that both the mean and the standard deviation of test scores influence the relative exports of two manufactured goods to a given market. In particular, countries with more (residual) skill dispersion tend to specialize in industries that exhibit weaker complementarities between workers skills. This evidence motivates the development of models in which the distribution of worker types is an independent determinant of specialization and perhaps models in which imperfect observability of worker type generates mismatch that affects relative industry productivity.

(ii) Wages vary widely across workers in the same occupation and sector

If worker heterogeneity matters for the effects of trade on income distribution, then we should expect to find that wage inequality exists not only between occupations and sectors but also *within* those occupations and sectors. Helpman et al. (2012) use matched employer-employee data for Brazil to document the importance of within industry and occupation wage inequality. When workers are distinguished by five occupation groups (professional and managerial, skilled white collar, unskilled white collar, skilled blue collar and unskilled blue collar) and by twelve industry categories, the within component accounts for more than two-thirds of the wage inequality that existed in Brazil in 1990 and more than two-thirds of the change in inequality from 1986 through 1995. Even at a much finer level of disaggregation, with 348 occupation categories and 283 industries, the within-cell component accounts for 52% of Brazilian wage inequality in 1990 and 54% of the

decadal change. Helpman et al. (2012) further report that residual wage inequality—the inequality that remains after accounting for observable differences in worker characteristics—explains about half of the overall level and growth of wage inequality in Brazil from 1986-1995. Residual wage inequality largely is found within occupation and sector categories. Obviously, the trends in such inequality cannot be understood using models that assume identical workers.

(iii) Recent trends in wage inequality are not captured well by the skill premium

Many papers have documented a substantial widening of the wage distribution in the United States and elsewhere during the 1980's, whether measured by educational or occupational grouping, by age, or by experience. Residual wage inequality within demographic and skill groups also increased at the time. Given the broad pattern of increased wage inequality—reinforced by similar changes in non-wage compensation—it seemed appealing to summarize the evolution of wage inequality using a simple statistic like the skill premium.

Autor et al. (2008) report that subsequent trends for the 1990's and early 2000's are more subtle. Whereas inequality has continued to grow in the upper half of the male wage distribution in the period after 1990, that among workers in the lower half of the distribution has narrowed during this later period. The authors describe the evolution of the U.S. wage distribution in recent years as being one of "polarization," with income gains at the top and the bottom of the distribution relative to those in the middle. They find polarization not only for the overall distribution of wages, but also for residual wage inequality and for earnings among workers at different education levels. Needless to say, polarization cannot be explained in a setting such as the Heckscher-Ohlin model that incorporates only two types of workers.

(iv) *Exporters pay a wage premium, but ...*

Many studies beginning with Bernard and Jensen (1995) have found that firms that export pay higher average wages than firms that serve only their domestic market, even after controlling for observable firm characteristics. Helpman et al. (2012) is a recent example of a paper that establishes an exporter premium (in Brazilian wages) after controlling for firm size and some observable worker characteristics. Premia such as those that have been identified could be an indication that exporting firms earn rents that they share with their workforce or simply that the composition of workers in exporting firms is different from that in firms that serve only the domestic market. Schank et al. (2007) provide some evidence in favor of the latter interpretation using matched employer-employee data for Germany. When they include person characteristics and person fixed effects to control for "worker quality" in their wage equations, they find that the estimated coefficient on a variable measuring the share of a firm's output that is exported shrinks substantially compared to a regression without these controls, and that export status (a dichotomous variable reflecting whether a firm exports or not) does not seem to affect the wages that a firm pays at all. These findings imply that unobserved worker characteristics are positively correlated with export behavior, as would be expected in a labor market with positive assortative matching.

(v) *Trade may affect the matching process, especially in export-oriented industries*

Labor economists have been asking for some time whether the more productive workers tend to work for the more productive firms. Beginning with the influential work of Abowd et al. (1999), a common methodology employed in the search for evidence of positive assortative matching (PAM) in labor markets uses matched employer-employee data to correlate the estimated fixed effect in a regression of employee wages with the estimated fixed effect in a regression of the wages firms pay. The idea is that workers who receive a high wage in all their jobs beyond what can be attributed to their observable characteristics are especially productive workers and firms that pay high wages to all workers beyond what would be predicted based on their observable characteristics are especially productive firms. Although the results of these many studies are mixed (see Postel-Vinay and Robin (2006) for a survey), by and large the evidence has been disappointing for those who believe that complementarities between worker skills and firm technologies are an important property of modern production processes.

Several authors have questioned whether the Abowd et al. (1999) methodology is up to the task up measuring PAM. For example, Lopes de Melo (2009) argues that there is a severe downward bias in using correlations between fixed effects to measure PAM.[1] The downward bias he suggests reflects non-monotonicities in the relationship between worker quality and wages caused by variation in firms' bargaining power. Productive firms may have stronger bargaining power than less productive firms, in which case they may pay lower wages to any given type of worker in the presence of costly search and labor market frictions. The productive firms may indeed hire the better workers, yet this may not be revealed by the wages they pay. Using an alternative measure that does not rely on fixed effects, Bagger and Lentz (2012) find significant evidence of PAM in matched employer-employee data for Denmark.

Davidson et al. (2010) is another recent study that finds evidence of PAM, in this case using data on Swedish workers and firms for the period 1995 to 2005. Interestingly, from the point of view of this survey, Davidson et al. (2010) proceed to investigate whether the degree of matching they find in a given industry and year is influenced by the extent of the industry's openness to trade. For openness, they use foreign industry tariff rates to reflect Swedish firms' access to foreign markets and Swedish tariff rates to reflect competition from imports. When they divide their sample into export-oriented industries and import-competing industries, they find that increased foreign openness strengthens the forces for PAM in Swedish export industries. In import-competing industries, by contrast, there is little evidence of a significant influence of (Swedish) trade openness on the degree of matching.

Taken together, these observations illustrate the many interesting lessons that can be gleaned from detailed information about firms and their workers. They suggest the subtle mechanisms of matching and sorting that operate in any labor market, mechanisms that only can be studied in models that allow for worker heterogeneity.

[1] See also Cahuc et al. (2006) and Eeckhout and Kircher (2011) on the bias that can result from using worker fixed effects in wage regressions to measure unobserved quality in the presence of search frictions in the labor market.

2 Modeling Worker Heterogeneity

Workers differ in a multitude of ways. A list of personal attributes that could be relevant for labor market outcomes would include health, gender, childhood background, family circumstances, years of schooling, years of job experience, reasoning ability, artistic ability, communication skills, organizational skills, and many others. Most of these attributes have several dimensions and they interact with job characteristics and industry of employment to determine a worker's productivity and pay.

In contrast to the richness of the actual differences between workers in the economy, the theoretical literature on labor heterogeneity deals with a much simplified abstraction. For purposes of surveying this literature, it will be enough to index workers by a variable q_L that describes a single dimension of variation.[2] It is typical to assume—as I will here—that workers with "more" of q_L (i.e., a higher index) are potentially more productive in all economic activities. In recognition of this assumption, I refer to q_L as a worker's "ability." I do allow the contribution that ability makes to a worker's productivity to vary across industries and occupations.

I posit a production function for the output of good i in country k that allows for (i) cross-country differences in technology, (ii) cross-industry differences in the contribution of ability to productivity, (iii) complementarities between worker ability and the quality or ability of any other factors or workers with whom a given worker is matched, (iv) constant or diminishing returns to the number of workers, given the quantities of any other inputs with which these workers are combined, and (v) cross-industry differences in the extent of any diminishing returns, which reflect differences in sectoral labor intensities. Specifically, I write

$$x_i^k = \psi_i^k \left(q_L, q_H\right) f_i \left(\ell\right) \ , \tag{1}$$

where x_i^k is the output of good i in country k when ℓ workers with ability q_L are combined with one unit of some other factor (if any) of quality or ability q_H and I assume $f_i'\left(\ell\right) \leq 0$. If there is a second factor, it might be physical capital or managerial time, or it might be another worker that is needed in combination with the worker in question to perform other of the various tasks required of a production team. The function $\psi_i^k \left(q_L, q_H\right)$ captures the contribution of ability to productivity in country k and industry i, as well as the complementarity or substitutability between factors or team members in the production process. The possibility that $f_i \left(\ell\right)$ is a declining function captures the potential diminishing returns to one factor, holding the quantity of another input fixed. These diminishing returns can reflect the scarcity of a manager's time, congestion on machines, or the diminishing productivity of doing more of one task without doing more of others. The industry subscript on f_i allows for the possibility that factor intensities differ across industries, as is common in multi-factor and multi-sector models of international trade.[3]

[2] Equivalently, there might be many dimensions of variation provided that these dimensions can be summarized in a single sufficient statistic for purposes of predicting labor market outcomes.

[3] A more general specification of the production technology would be to write $x_i^k = g_i^k \left(q_L, q_H, \ell, h\right)$, where h is the quantity of the second factor whose quality is q_H. This would allow, for example, for the possibility that the

Note that the production function (1) embeds the familiar, neoclassical production function with constant returns to scale that applies with homogeneous factors. If all types of a factor are equally productive, $\psi_i^k(q_L, q_H)$ is just a constant, in which case $x_i^k = \psi_i^k f_i(\ell)$ is output per unit capital (or per manager) in industry i and country k. In this case, we would typically assume that $f_i'(\ell) < 0$, in recognition of diminishing returns to labor for a given quantity of capital or managers. Of course, if labor is the only input, then constant returns would imply a linear production function, so we would write $x_i^k = \psi_i^k \ell$.

3 A Short Digression on Complementarity

Complementarity features prominently in the literature on the sorting and matching of heterogeneous labor. Two types of complementarity are captured in the literature. First, there may be complementarity between the attributes of a worker and the occupation or sector in which he or she is employed. A more able worker, though absolutely more productive in all activities, may be relatively more productive in certain occupations depending on the job requirements. For example, some occupations or industries may use production methods that are more technologically complex and therefore may reward greater technical sophistication on the part of their employees. More able workers may be only slightly more productive than their less able counterparts in a sector that uses a simple technology, but considerably more productive in one that uses a complex technology. Then, we would say that there is complementarity between the type (ability) of the worker and the type (technological complexity) of the industry or job. Second, but related, there may be complementarity between the type of a worker and the type of any other factor or team member with whom the worker is paired. In any given industry or occupation, a more able worker may be more productive than a less able worker when matched with a given manager or machine, but the more able worker may be especially productive compared to the less able worker when the manager or machine also is of high quality.

Mathematically, the concept of complementarity can be captured by the property of supermodularity. A function of two variables $F(x, y)$ is *supermodular* if

$$F(x_2, y_2) + F(x_1, y_1) \geq F(x_2, y_1) + F(x_1, y_2) \tag{2}$$

for all pairs $x_2 > x_1$ and $y_2 > y_1$ in the domain of $F(\cdot)$. By rearranging terms, this inequality is equivalent to $F(x_2, y_2) - F(x_1, y_2) \geq F(x_2, y_1) - F(x_1, y_1)$; i.e., the extra "output" generated by increasing the first input from x_1 to x_2 is greater if the input is combined with the larger "other input" y_2 than if it is combined with the smaller y_1. If the function $F(\cdot)$ is twice differentiable, this is equivalent to the cross-partial derivative, $F_{12}(x, y)$, being positive for all x and y.

A stronger notion of complementarity is log supermodularity.[4] A function $F(x, y)$ is *log super-*

diminishing returns to the quantity of labor depends on the quality of labor that is employed. I do not write the technology this way, because little of the literature that I shall review has anything to say about this more general case. But see Eeckhout and Kircher (2012), who study the the conditions for PAM in this more general environment.

[4] Actually, log supermodularity need not be a stronger requirement than supermodularity for functions that are

modular if

$$F(x_2, y_2) F(x_1, y_1) \geq F(x_2, y_1) F(x_1, y_2) \tag{3}$$

for all pairs $x_2 > x_1$ and $y_2 > y_1$ in the domain of $F(\cdot)$. If $F(x, y) > 0$ for all (x, y), then this is equivalent to requiring that the function $\log F(x, y)$ be supermodular. This time, we can rewrite the inequality as $F(x_2, y_2)/F(x_1, y_2) \geq F(x_2, y_1)/F(x_1, y_1)$, which means that output rises *relatively* more when the first input is increased from x_1 to x_2 if that input is combined with a larger other input y_2 than if it is combined with the smaller y_1.

For some purposes, it will also be useful to define a property of a function that characterizes strong substitutability between the arguments. A function $F(x, y)$ is *submodular* if

$$F(x_2, y_1) + F(x_1, y_2) \geq F(x_2, y_2) + F(x_1, y_1) \tag{4}$$

for all pairs $x_2 > x_1$ and $y_2 > y_1$ in the domain of $F(\cdot)$. This is, of course, just the opposite of the inequality that defines supermodularity in (2). A production function might be submodular if the output depends mostly on the greater of the two inputs. Then, it would be inefficient to "waste" a good input by combining it with another good input, when the input could do nearly as well with a poorer partner and the better "other input" could be saved for combination with a lesser partner.

4 Sorting without Matching: Constant Returns to Labor

I begin by discussing the sorting of workers to industries in a setting in which there is no matching of workers with other factors or teammates and no issue of diminishing returns to the quantity of labor. This sorting problem has been studied in its most general form in Costinot (2009). Costinot generalizes prior work by Ruffin (1988), Matsuyama (1992), and Bougheas and Riezman (2007).[5]

Consider a perfectly competitive economy and suppose that output in sector i and country k is a linear function of the number of workers after adjustment is made for each worker's sector-specific productivity. That is, write aggregate output in sector i and country k as

$$X_i^k = \int_{q \in Q_i^k} \psi_i^k(q) \phi^k(q) \bar{L}^k dq$$

where Q_i^k is the set of worker types that sorts to industry i in country k in the competitive equilibrium, $\phi^k(q)$ is the probability density of workers of type q in country k, and $\bar{L}^k$ is the total mass of workers in country k. Notice that workers of type q have productivity $\psi_i^k(q)$ in industry i in country k and that aggregate output in the industry is just the sum of what the individual

not monotone. But since the functions we consider here are all monotonically increasing in their arguments, we can consider log supermodularity as the more stringent condition.

[5] Costinot allows for an arbitrary number (or a continuum) of workers, industries and countries. Ruffin assumes there are two countries that share common productivity for a given worker type in a given industry. Bougheas and Riezman focus on an economy with two sectors in which all workers are equally productive in one of them. Finally, Matsuyama analyzes a small country with two sectors in which newborn agents can choose their industry of lifetime employment after learning their individual comparative advantage. He studies the dynamics of resource allocation as workers die and are replaced by new generations.

workers who are employed there produce "on their own". In this setting, a worker of type q in country k can produce output with value $p_i\psi_i^k(q)$ by working in industry i. Competitive firms in industry i will bid this amount for the worker's services. In equilibrium, each worker will sort to whichever sector offers him the highest wage.

Costinot first describes the equilibrium sorting pattern for two special cases. In a setting with only Ricardian technological differences across countries,

$$\frac{\psi_{i_2}^k(q)}{\psi_{i_1}^k(q)} = \frac{\tilde{\psi}_{i_2}^k}{\tilde{\psi}_{i_1}^k} \text{ for all } q \text{ and all } (i_1, i_2).$$

That is, the relative productivity of a worker in country k in any two different industries is the same, regardless of the worker's type. Here, the more able workers in a country are proportionally more productive than their less able counterparts in all sectors of the economy. Yet, (all) workers in a country may be relatively more productive in some industry than another compared to the relative productivity of workers located in some other country.

In such a setting, if a worker with ability q_1 prefers to work in industry i_1 than in industry i_2—because he is offered a higher wage there—then so does a worker with some different ability level q_2. At given prices, all workers in a country sort to the same industry, or to multiple industries only if the relative price exactly compensates for the relative productivity differential. The sorting in different countries will vary depending on the international pattern of technological comparative advantage.

Costinot shows for this case that, if countries and industries can be ordered in such a way that productivity as a function of k and i is log supermodular (that is, if high-index countries are relatively more productive in high-index industries), then aggregate output also will be a log supermodular function of k and i. The high-index countries will produce relatively more in the high-index industries and will export these goods in a trade equilibrium. Trade has no effect on the distribution of income in any country, because the relative productivity of any two workers in country k depends only on their relative ability levels and not on their ultimate industry of employment.

In Costinot's second special case, countries differ from one another in their endowment distributions and workers differ in their relative productivities in different industries, albeit in a similar way across countries. If there are any cross-country differences at all in technologies, they are Hicks-neutral with respect to worker-type and industry. More formally, suppose that

$$\frac{\psi_{i_2}^k(q)}{\psi_{i_1}^k(q)} = \frac{\tilde{\psi}_{i_2}(q)}{\tilde{\psi}_{i_1}(q)} \text{ for all } q \text{ and all } (i_1, i_2).$$

Here, the relative productivity of a worker in two different industries i_1 and i_2 depends on the worker's ability q, but it does not depend on the country in which that worker is located. Accordingly, workers of a given type q will face the same relative wage opportunities in any pair of industries irrespective of where these workers live, and hence they will make the same employment

choices.

The implied pattern of specialization is reminiscent of the Heckscher-Ohlin theorem. Suppose that industries can be ordered such that productivity is a log supermodular function of i and worker ability q; i.e., more-able workers are relatively more productive in higher-index industries. And suppose that countries can be ordered so that the probability density of workers is a log supermodular function of k and q; i.e., more-able workers are relatively more abundant in high-index countries.[6] Then aggregate output will be a log supermodular function of i and k; i.e., a country with a higher index will produce relatively more of a good with a higher index among any pair of goods. This reflects the fact that sorting in any pair of countries is the same and so relative output reflects the relative numbers in the countries' endowments of different types of labor.

Costinot and Vogel (2010) study a world economy comprising two countries that have similar technologies but different distributions of worker types, as in Costinot's second special case.[7] They focus on the effects of trade on income distribution, a question that Costinot (2009) does not consider. In a variant of the model that might describe "North-South trade", let the North be relatively well endowed with the more able workers among any pair of workers types. Then the integrated world economy has a relatively greater number of the less able workers among any pair of worker types than does the Northern economy in autarky, but it has a relatively smaller number of the less able of any pair of worker types than does the Southern economy in autarky. Assuming that consumers worldwide have identical, constant-elasticity-of-substitution preferences across the many goods produced by these economies, the authors show that, in a trade equilibrium, Northern workers of any given ability level find employment in a higher-indexed industry than where they work in autarky, and that Southern workers of any given ability level find employment in a lower-indexed industry than where they work in autarky. Intuitively, the high-ability workers are relatively more scarce in the integrated world economy than they are in the autarky equilibrium in North, so they are more "needed" to produce in industries that are complementary to their skills. The upshot is that trade causes the relative wage of the more able worker in any pair of workers to rise in the North and to fall in the South. The more able Northern workers see their scarcity rise as a result of integration with the South, whereas the less able Southern workers see their scarcity rise as a result of integration with the North.

Now consider a variant that might describe certain instances of "North-North trade" in which the two countries differ in the *diversity* of their distributions of worker ability but not necessarily in their mean ability levels. Let one country, say "West," have a relatively greater endowment of the less able workers than "East" for any pair of worker types with indexes q_1 and q_2 both less than some q', but a relatively greater endowment of the more able workers for any pair of worker types with indexes q_3 and q_4 both greater than q'. In other words, West has more workers in the tails of the distribution compared to East, with relatively many of the very high ability workers

[6] Let $V(q,k)$ be the supply of workers of type q in country k. Then if V is log supermodular, $V(q_1,k_1)/V(q_0,k_1) > V(q_1,k_0)/V(q_0,k_0)$ whenever $q_1 > q_0$ and $k_1 > k_0$.

[7] Actually, Costinot and Vogel allow for cross-country productivity differences provided they are Hicks-neutral; i.e.,proportionately the same for all worker types and industries.

but also relatively many of the very low ability workers. In this case, it is the extreme workers in West who find that their relatively scarcity increases as a result of trade with East. Every Western worker with ability $q > q'$ sorts to a higher-indexed sector in the trade equilibrium than in autarky while every Western worker with with ability $q < q'$ sorts to a lower-indexed sector in the trade equilibrium than in autarky. As a result, the relative wage of the more able worker of any pair rises in the West, if both workers in the pair have ability above q', while the relative wage of the more able worker of any pair falls in West, if both members of the pair have ability less q'; i.e., trade causes a *polarization* of the income distribution. The distributional effects of trade are just the opposite in East.

Returning to the analysis of sorting in Costinot (2009), consider the more general case in which relative productivity in two industries is common neither across worker types in a given country nor across countries for a given worker type. Costinot considers the case in which productivity is a log supermodular function of industry i, country k, and ability q.[8] He examines the pattern of specialization when the distribution of worker types is a log supermodular function of k and q; i.e., the higher-indexed country in any pair has a relatively larger endowment of the more able of any pair of worker types. This, he shows, is not sufficient to allow a complete description of the sorting pattern. In particular, the higher-index country in a pair may produce relatively more in some industry i_1 compared to i_2, even if $i_1 < i_2$. All that we know in general is that the higher-index country in a pair produces relatively more in the *highest*-index industry and relatively less in the *lowest*-index industry among all the industries in the world economy. In other words, if cross-country differences in relative sectoral productivities and cross-industry differences in relative worker productivities (in conjunction with differences in type distributions across countries) are both sources of comparative advantage, then the pattern of international specialization is clear only for the two most extreme sectors.

Ohnsorge and Trefler (2007) study a similar economy in which there again are no diminishing returns to labor and workers differ in their relative productivity across sectors. They begin with the intriguing possibility that each worker is endowed with a pair of attributes, q_a and q_b. Of course, a worker brings both of his attributes to any industry in which he accepts employment. A worker's productivity depends on this pair of attributes and the industry of application, but not on the country in which the worker is located. This would seem a more complicated problem than Costinot's, because the pair of attributes that describe a worker may contribute differently to productivity in the different sectors. However, Ohnsorge and Trefler assume that $\psi_i(q_a, q_b)$ is homogeneous of degree one, that is, it can be written as

$$\psi_i(q_a, q_b) = q_a \tilde{\psi}_i(q_b/q_a).$$

[8] A function of three arguments is log supermodular if it is log supermodular as a function of any pair of arguments, holding the third argument constant. In our setting, this would mean that higher index workers are relatively more productive in higher index industries in every country, higher index workers are relatively more productive in higher index countries in every industry, and higher index countries have a relatively better technology in higher index industries for every worker type.

With this assumption, the relative wage that any worker can earn in any pair of industries depends only on his relative endowment of the two attributes, q_b/q_a, and not on his absolute endowment of either one. All workers with the same relative endowment make the same job choices, regardless of their locations and regardless of their absolute abilities as captured here by q_a. It follows that sorting to industries is guided once again by a unidimensional ordering of worker types. If productivity is a log supermodular function of industry and relative attribute endowment q_b/q_a, then the worker that has a higher q_b/q_a among any pair will sort to the industry with the higher index in the pair.

One can examine in this setting how the distribution of attributes across the labor forces of two countries affects the pattern of specialization and the effects of trade on the wage distribution. Suppose that productivity is a log supermodular function of i and q_b/q_a; i.e., a high relative endowment of attribute b compared to attribute a confers a comparative advantage to a worker in industries with higher indexes. Then, for example, if $\log q_a$ and $\log q_b/q_a$ are spread among workers according to a bivariate normal distribution, and if the means and variances of these variables are the same in the two countries, the country with the higher correlation between $\log q_a$ and $\log q_b/q_a$ will produce relatively more of the higher-indexed goods. The explanation for this should be clear. The sorting of workers to industries reflects the relative attribute endowment of the workers and is common in the two countries. Then the country with the higher correlation between $\log q_a$ and $\log q_b/q_a$ has, on average, more *absolutely* productive workers among those that have a *comparative* productivity advantage in the high-index industries.

The Ohnsorge-Trefler model features both between-industry and within-industry wage inequality. The inequality across industries reflects, as in Costinot and Vogel (2010), the different relative productivity of workers with different q_b/q_a in the various sectors. The within-industry inequality reflects the fact that different workers who sort to the same industry (because they have the same q_b/q_a) have different absolute productivity levels. Looking across industries, the country with the higher correlation between absolute and comparative advantage has a steeper wage profile and therefore more between-industry wage inequality. However, for given q_b/q_a, a higher correlation between $\log q_a$ and $\log q_b/q_a$ implies a lower variance of q_a. Hence, the high-correlation country has lower within industry wage inequality.

5 Sorting without Matching: Diminishing Returns to Labor

In the last section, I described sorting to industries based on worker comparative advantage in a setting in which each individual worker can make his employment decision independently from the decisions of others. A worker's potential productivity in an activity did not depend on his pairing with other workers or other factors of production, nor did it depend on the number of other workers who choose the same sector of employment. Accordingly, the workers only need information about their own relative strengths in different occupations and information about the market prices of final outputs in order to decide where to seek and accept employment. In this section, I continue with the assumption that a worker's productivity depends only on his own attributes and not on

the attributes of others in the firm or of the factors with which he is combined. But now I introduce diminishing returns to the quantity of labor. These diminishing returns may reflect congestion on capital equipment or increasing scarcity of a manager's time. For concreteness, I adopt the latter terminology. In order to postpone the discussion of matching between workers and managers, I follow Grossman et al. (2013) in beginning with the simpler (if unrealistic) case in which all managers are similar.

Consider a two-sector world economy in which managers are mobile across the sectors.[9] Suppose that the potential output in sector i that can be produced by a single manager and ℓ workers with ability q is given by $x_i = \psi_i(q)\,\ell^{\gamma_i}$. In this technology—which is common to the two countries—the extent of diminishing returns to the quantity of labor per manager is reflected in the parameter $\gamma_i < 1$. These diminishing returns can differ across sectors as they will if the managerial input is relatively more important in one sector than the other. This captures, of course, the cross-industry differences in factor intensities that are common in models of trade based on factor endowments.

In this setting, firms in each industry must choose how many managers to hire, what type of workers to hire, and how many workers to combine with each manager. In the competitive equilibrium, as usual, the size of the individual firm (i.e., the number of managers in a particular firm) is not determined. The equilibrium salary of managers will be such that all active firms make zero profits and each individual firm is indifferent as to the number of managers it employs. However, the aggregate employment of managers in each sector is uniquely determined in the general equilibrium, as is the number of workers per manager.

An individual firm may opt to hire a particular type of labor, depending on the wage schedule it faces. Moreover, the ability level of workers determines a firm's productivity and so a firm that hires more able workers may decide to combine a greater number of workers with each manager that it employs. A firm's profit maximization problem can be solved in two steps, by first writing the optimal number of workers per manager as a function of the ability levels of those workers and their market wage (also a function of ability) and then by optimizing over the choice of q. The resulting first-order condition that reflects both optimal type and optimal number can be written as

$$\frac{\varepsilon_{\psi_i}(q)}{\gamma_i} = \varepsilon_w(q), \tag{5}$$

where $\varepsilon_{\psi_i}(q)$ is the elasticity of the productivity function ψ_i for sector i with respect to ability and $\varepsilon_w(q)$ is the elasticity of the wage schedule with respect to ability. The left-hand side of (5) reflects the marginal benefit to the firm in industry i (expressed as an elasticity) from upgrading the quality of its workforce, taking into account both the direct effect on productivity and the indirect effect that results from its adjustment in the size of its work unit. More able workers are more productive (as reflected by ε_{ψ_i}) but they also are hired in greater numbers, the more so in labor-intensive industries. Essentially, a firm faces a choice between hiring *better* workers or *more* workers to produce its desired output per manager and the trade-off is resolved when the firm

[9] Mussa (1982) has a similar model, except that his second factor—which he calls capital—is specific to an industry.

equates the ratio of the elasticity of output with respect to ability (given ℓ) relative to the elasticity of output with respect to quantity (given q) with the elasticity of the wage schedule.[10]

It might seem from equation (5) that all firms in sector i will hire the same type of worker by choosing the q that maximizes profits. However, were that to occur, many types of workers would find themselves unemployed. Of course, such unemployment does not arise in a competitive equilibrium with flexible wages and flexible input ratios. Instead, the wage schedule adjusts in equilibrium to ensure that (5) is satisfied for the entire range of workers that sorts to sector i. In short, the wage of a worker with ability q who works in sector i is

$$w(q) = w_i [\psi_i(q)]^{1/\gamma_i}, \text{ for } q \in Q_i, \tag{6}$$

where w_i is a scalar that anchors wages in the sector. Facing this wage schedule, the firms operating in sector i are indifferent among the various worker types that sort there, and so all will be hired by some firm or another.

The wage schedule reveals the equilibrium sorting pattern in both countries. If $\varepsilon_{\psi_1}(q)/\gamma_1 > \varepsilon_{\psi_2}(q)/\gamma_2$ for all types q in the labor force, then the more able workers with $q > q^*$ for some endogenous q^* will sort to industry 1 and the less able workers with $q < q^*$ will sort to industry 2. With any other pattern of sorting, the firms in industry 1 could find some high-ability workers in industry 2 whom they would prefer to hire than their own employees. Similarly, the firms in industry 2 could find some low-ability workers in industry 1 that would be more profitable as hires than their own employees. As Grossman et al. discuss, if the ranking of the sectors in terms of $\varepsilon_{\psi_i}(q)/\gamma_i$ reverses in the domain of worker types for the economy, then the sorting pattern will be more complex with, perhaps, high and low-ability workers employed in one sector, and middle-ability workers employed in the other.[11]

The pattern of trade reflects both the countries' aggregate relative abundance of managers versus workers and the distributions of worker types in each country. Suppose, for example, that the two countries share the same probability density functions for q but that $\bar{H}^A/\bar{L}^A > \bar{H}^B/\bar{L}^B$, where $\bar{H}^k$ is the number of managers in country k. Then the distributions of worker types generates no comparative advantage for either country, but country A, with its relative abundance of managers, has a comparative advantage in the manager-intensive sector. That is, if $\gamma_2 > \gamma_1$ and the countries have identical and homothetic preferences, country A will be relatively specialized in producing good

[10] Note that (6) represents a generalization of the wage equation (11) in Costinot and Vogel (2010). In my notation, let $\psi(q,i)$ be the productivity of a worker with ability q in industry i. Then, Costinot and Vogel prove in their Lemma 2 that

$$\varepsilon_\psi [q, \iota(q)] = \varepsilon_w(q)$$

where $\iota(q)$ is the equilibrium sector of employment of a worker with ability q and $\varepsilon_\psi \equiv q(\partial\psi/\partial q)/\psi$. Their model has constant returns to labor, so implicitly $\gamma(i) = 1$. Therefore, their Lemma 2 also implies that wages rise with an elasticity equal to the elasticity of productivity with respect to ability divided by the elasticity of output with respect to quantity (which, in their case, is equal to one).

[11] Of course, if $\gamma_1 = \gamma_2$, then the sorting pattern reflects a comparison of ε_{ψ_1} and ε_{ψ_2}. The better workers will sort to industry 1 if and only if productivity is a log supermodular function of ability and industry. This observation represents a generalization of the findings of Costinot (2009) to the case with diminishing returns when factor intensities are the same in all industries.

1 and will export this good in the trade equilibrium. Alternatively, if $\bar{H}^A/\bar{L}^A = \bar{H}^B/\bar{L}^B$ but the distribution of talent is shifted proportionately to the right in country A relative to country B, then country A will produce relatively more in the industry where ability contributes more elastically to productivity. That is, if $\varepsilon_{\psi_i}(q) > \varepsilon_{\psi_j}(q)$, country A will have a comparative advantage in producing good i relative to good j. Notice that this need not be the good produced by the country's most able workers.

How does trade affect the wage distribution here? The opening of trade generates no change in any within-industry distribution inasmuch as (6) pins down the relative wages of any pair of workers employed in the same industry as a function of their respective ability levels and technological considerations. However, trade does affect the relative wages of workers employed in different sectors, as movements in relative price induce changes in the equilibrium wage anchors, w_1 and w_2. Grossman et al. show that an increase in the relative price of the exported good increases the relative wage of any worker employed in the export industry compared to any worker employed in the import-competing sector. If the two industries do not differ much in labor intensity, so that $\gamma_1 \approx \gamma_2$, then workers who remain in the import-competing industry will see their real wage fall as a result of an increase in the price of the export good, while those initially in the export industry will benefit from such a price change. Alternatively, if $\varepsilon_{\psi_1}(q)/\gamma_1 \approx \varepsilon_{\psi_2}(q)/\gamma_2$, then all workers in the economy will gain from an increase in the relative price of exports if and only if the export sector is more labor intensive than the import-competing sector. These findings can be understood with reference to the well-known results for the Heckscher-Ohlin and Ricardo-Viner models for the case of homogeneous labor. If factor intensities are similar in the two industries, there are no Stolper-Samuelson forces to benefit workers at the expense of managers, or vice versa. What remains is the fact that some workers have comparative advantage in the export industry as a reflection of their (high or low) ability levels, while others have comparative advantage in the import-competing industry. These comparative advantages create partial sector specificity, as in Mussa (1982) or Grossman (1983). On the other hand, if the incentives for sorting are muted, because $\varepsilon_{\psi_1}(q)/\gamma_1 \approx \varepsilon_{\psi_2}(q)/\gamma_2$, then there are no strong forces to tie a worker's fate to his industry of employment. Instead, all workers gain in real terms when an increase in the price of the labor-intensive good generates an economy-wide increase in the demand for labor. The workers in the export sector will gain more, unless the forces that induce intersectoral sorting are absent entirely.

6 Matching without Sorting

I turn next to models that examine the matching of heterogeneous workers with managers or technologies. The simplest such models abstract from the sorting of workers to industries by assuming that the economy has only one sector. In the case of Antràs et al. (2006), the world economy is competitive and firms everywhere produce a homogeneous good. In such a setting, there is no "trade" (except to balance factor payments) and the focus instead is on the formation of international production teams. In the case of Sampson (2012), the world economy produces

symmetrically differentiated products under conditions of monopolistic competition. Then, there is intra-industry trade but, of course, no inter-industry trade. In both cases, the authors show conditions for PAM and analyze how globalization affects the wage distribution by changing the equilibrium matching of workers to managers or firms.

Antràs et al. (2006) develop an elegant model of production hierarchies based on Garicano and Rossi-Hansberg (2006). In their model, every individual—who is endowed with one unit of time but has different ability or "knowledge"—faces an occupational choice between becoming a worker or a manager. If an individual chooses to be a worker, he will have the opportunity to use his unit of time to produce a unit of output of the homogeneous final good. However, first he must solve a "problem" that arises in the course of production. A worker of ability (or with knowledge) q_L can solve the problem that confronts him with probability q_L. If the worker cannot solve the problem, he can pass it along to his manager. The manager presumably is more able (or has greater knowledge) than the workers on his team, so the original problem is solved with probability $q_H > q_L$ when the manager has ability q_H. However, each problem that is brought to the attention of the manager absorbs a fraction δ of his time endowment. The authors ask, which individuals choose to be managers and which choose to be workers, what production units form in equilibrium, and how does integration with another economy that has a different distribution of ability levels affect team formation and the resulting distribution of income?

We can write the productivity of a production team here in terms of our earlier notation. A manager with ability q_H can supervise $1/\delta\,(1-q_L)$ workers of ability q_L, considering that a fraction $1-q_L$ of the problems that these workers face will be passed on to the manager and that each such problem uses a fraction δ of the manager's time. Considering the manager's time constraint, the output of a production team with a manager of ability q_H and workers of ability q_L is given by

$$\psi\,(q_H, q_L) = \frac{q_H}{\delta\,(1-q_L)}. \tag{7}$$

The productivity function (7) is supermodular, which induces positive assortative matching between managers and workers. PAM means that the more able of any pair of managers always is teamed with the more able group of workers. PAM derives from the complementarity between manager ability and worker ability that is implicit in (7). The contribution of the manager's ability to team output depends on how many opportunities the unit has to generate output. More workers provide more production opportunities. And the manager can support a greater number of workers, and thereby confront a greater number of production opportunities, when the workers can solve a greater fraction of the problems on their own. It follows that the contribution of a manager's ability to his unit's output increases with the ability of his team of workers.

Antràs et al. (2006) establish the existence of a "threshold equilibrium" for this economy. In this equilibrium, the most able individuals with ability greater than some q^* choose to become managers while the less able individuals with ability less than q^* choose to become workers. The complementarity between worker and manager abilities, together with the resulting pattern of PAM,

generates an earnings schedule (of worker's wage or manager's salary) that is a strictly convex function of q. Now suppose that "North" integrates with "South," where the former country has a uniform distribution of abilities q on support $[0, 1]$ and the latter has a uniform distribution of abilities q on support $[0, q^S_{\max}]$, for some $q^S_{\max} < 1$. In the integrated world economy, firms can combine a manager located in one country with workers situated in the other.

The integrated equilibrium with global production teams features fewer Southern managers and more Southern workers than does the autarky equilibrium in the South. This is because some of the Southerners that choose to become managers in autarky opt instead to become workers in international production teams when they can join with more able Northern managers. The matches formed by all of the original Southern workers are upgraded as the result of globalization, because the management pool improves with the addition of the able Northern managers and the exit of the least able of the former Southern managers. Meanwhile, the matches of the remaining Southern managers deteriorate inasmuch as these managers must compete for workers with the group of talented Northerners. Considering that the return to ability for a worker increases with the ability of the manager with whom he is teamed and similarly the return to ability for a manager increases with the ability of the workers that report to him, the authors find a steepening of the wage profile for the initial Southern workers and a flattening of the salary profile for the original Southern managers. The steepening of the wage profile together with the addition to the worker pool of some former managers implies an increase in within-worker wage inequality in the South.

In the North, some middle-ability individuals who opt to be workers in the autarky equilibrium choose instead the alternative occupation in the integrated equilibrium in order to serve as managers for the larger pool of less-able workers. It follows that a group of the lowest ability Northern workers will see the quality of their manager downgraded as the threshold for choosing management as an occupation falls. The best Northern workers, in contrast, find better matches in the integrated equilibrium than in autarky, since some of their most able competitors now opt out of the worker pool in order to manage international production teams. Meanwhile, the least able Northern managers in the autarky equilibrium find themselves matched with better workers in the integrated equilibrium, but the most able Northern managers may see their matches deteriorate. The upshot is that the return to ability falls for low-ability Northern workers, but rises for high-ability Northern workers. The authors show that overall wage inequality can rise or fall in the North as a result of integration with the South.

Sampson (2012) studies matching between heterogeneous firms and heterogeneous workers in an economy in which the firms have access to different technologies. The better the technology, the more productive is the firm, but productivity also depends on the ability q_L of the workers that the firm hires. In our notation, a firm with technology q_H that hires ℓ workers with ability q_L produces

$$x = \psi(q_H, q_L)\,\ell$$

units of output, where $\psi(q_H, q_L)$ is assumed to be log supermodular to capture complementarity between technology and worker type. Notice that the production technology is linear in the number

of workers. But Sampson assumes that the firms produce differentiated products and therefore face downward sloping demands. It follows that the firms perceive diminishing marginal *revenue* from additional employment.

Here, the log supermodularity of $\psi(q_H, q_L)$ induces PAM between workers and firms; i.e., the best workers are hired by the firms with the best technologies. Facing demands of the form $Zp_j^{-\sigma}$, where p_j is the price of variety j and σ is the elasticity of substitution between any pair of goods, the optimal employment for a firm with technology q_H that hires workers of ability q_L and pays them the competitive wage $w(q_L)$ is

$$\ell(q_H, q_L) = \left(\frac{\sigma - 1}{\sigma}\right)^{\sigma} Z\psi(q_H, q_L)^{\sigma-1} w(q_L)^{-\sigma}.$$

Using this expression for optimal employment in the firm's profit function, Sampson relates the optimal match to the elasticity of the wage schedule, namely

$$\varepsilon_{\psi L}[m(q_L), q_L] = \varepsilon_w(q_L) \tag{8}$$

where $m(q_L)$ is the type of the firm that hires workers of ability q_L, $\varepsilon_w(q_L)$ is the elasticity of the wage schedule, and $\varepsilon_{\psi L} = q_L[\partial\psi/\partial q_L]/\psi$ is the (partial) elasticity of productivity with respect to worker ability. Notice two differences between (8) and (5), which is the analogous expression for optimal hiring in a competitive model with homogeneous managers and firms. First, the left-hand side of (8) is not divided by an output elasticity with respect to labor quantity, which reflects the fact that, in Sampson's model, there are no diminishing returns to the number of workers. Second, the elasticity of productivity in Sampson's setting depends not only on the worker's type, but also on the type of the firm that employs him, with such matching determined in the general equilibrium.

The distribution of technologies across firms and the wage schedule together govern the aggregate demands for workers at the various ability levels. In equilibrium, these demands must match the exogenous supplies. Let $\bar{H}$ and $\bar{L}$ represent the masses of firms and workers in equilibrium (the former endogenous and the latter exogenous) and let $\phi_F(q_F)$ represent the probability density function that describes the distribution of factor types for $F = H, L$. Then labor market clearing for every ability level requires

$$m'(q_L) = \frac{\bar{L}\phi_L(q_L)}{\bar{H}\phi_H(q_H)} \frac{1}{\ell(q_H, q_L)}. \tag{9}$$

Together (8) and (9) comprise a pair of differential equations that determine the equilibrium wage schedule and the equilibrium matching schedule, once the mass and distribution of firm types is known.

From this, Sampson makes an interesting observation. The differential equation (8) can be

solved for wages in terms of the equilibrium matches, which gives

$$w\left(q_L\right) = \exp\left[-\int_{q_{L\min}}^{q_{L\max}} \frac{1}{\psi\left[m\left(q_L\right), q_L\right]} \frac{\partial \psi\left[m\left(q_L\right), q_L\right]}{\partial q_L} dq_L\right].$$

It follows that, once the matching function is known, so is the wage schedule, and thus the former is a sufficient statistic for the latter. Moreover, the complementarity between firm type and worker ability implies that output is more responsive to worker ability in firms with better technologies. Therefore, if a shock causes a subset of workers to match with better firms than before, the wage schedule becomes steeper and inequality rises among this set of workers. This is true also in Costinot and Vogel (2010) and in Antràs et al. (2006), where in place of "better firms" we would say "higher-indexed industries" in the former and "better managers" in the latter.

Sampson goes on to consider an economy in which firms draw their technologies q_H from a truncated Pareto distribution upon paying a sunk entry cost, as in Melitz (2003). After entering and learning its type, a firm must pay a fixed cost in order to produce and an additional fixed cost for the opportunity to export. The mass of entrants is such that expected profits are zero. In this setting—as is well known from Melitz—it may be that only the firms that draw the best technologies choose to export. In such an equilibrium, the opening of trade allows all workers to match with better quality firms than in autarky inasmuch as the firms with the worst technologies exit the market when confronted by competition from imports. The technological upgrading by workers benefits especially the most able among them, and so there is a pervasive increase in wage inequality. Finally, the combination of PAM and the fact that only the firms with the best technologies choose to export generates an exporter wage premium. Here, the premium arises solely from the different composition of workers in firms that export compared to that in firms that sell only domestically.

In the Melitz model and its extension by Sampson to a setting with heterogeneous labor, the distribution of technologies among firms is exogenous. Alternatively, one might imagine situations in which firms can choose their technologies by, for example, opting to install certain types of machines. Yeaple (2005) was the first to consider such endogenous technology choice in a model with heterogeneous workers and trade.[12] He assumed that firms that produce differentiated products can choose one of two machine types and one of two worker types. Better machines cost more to install. Better workers have a comparative advantage in operating the better machines. There are both fixed and variable costs of trade between two identical countries.[13]

In this setting, the distribution of technologies adapts to the distribution of workers. The firms that hire the less able workers with $q_L < q_L^*$ install the cheaper and less productive machines. Firms that hire the more able workers install better machines. The latter firms produce more and pay higher wages. Moreover, for a range of parameter values, the high-tech firms choose to export, while

[12] See also Bustos (2011).

[13] In Yeaple (2005), there is also a homogeneous good, so there is both matching within the differentiated products sector and sorting between sectors. Here, I focus on the implications of his model for matching, and so I omit the homogeneous good from my discussion.

the low-tech firms do not. Thus, again, we find an exporter wage premium. Yeaple shows that a reduction in variable trade costs increases pay at the top of the wage distribution but reduces pay at the bottom, thereby increasing inequality. Sampson (2012) extends this one-sector variant of the Yeaple model to include a continuous choice of technology with an entry cost that is an increasing function of machine quality q_H. He shows that the endogenous distribution of technologies reflects the exogenous distribution of labor, so that if the latter is continuous the former will be so as well. Under the assumption that $\psi(q_H, q_L)$ is log supermodular, the equilibrium features PAM. An opening to trade causes firms that hire the more able workers to upgrade their technologies. The return to ability rises for exporters and wage inequality among those who work in firms that export increases. In contrast, trade has no effect on the wage distribution among workers employed by firms that sell only domestically, but the wage of anyone who works in an export firm rises relative to that of anyone who works in a firm that sells only on the domestic market.

7 Matching and Sorting

Until now, I have discussed sorting and matching in isolation. But, clearly, there can be important interactions between the two. If workers of a certain type are most productive when they participate in production units with some particular composition, then the forces that influence the sorting of other factors must also influence the sorting of workers. And if other factors sort in a certain way, then the matching possibilities for workers will be constrained by the choices made by these others. Moreover, in a world with both matching and sorting, the effects of trade on the wage distribution reflect both the changes in the demand for labor that result as export sectors expand and import-competing sectors contract and the change in the returns to ability that result as workers reorganize into new production units. In this section, I discuss some recent research that features both the formation of multi-factor or multi-worker production teams and the allocation of labor to one of several industries that differ in their uses of worker attributes.

Grossman et al. (2013) consider an economy with two goods and two inelastically-supplied factors in which both workers and managers are heterogeneous in their abilities. The output of good i that results when h managers of ability q_H are combined with ℓ workers of ability q_L is

$$x_i = h\psi_i(q_H, q_L)(\ell/h)^{\gamma_i}.$$

Country c is endowed with a density $\bar{H}^c\phi_H^c(q_H)$ of managers of ability q_H and a density $\bar{L}^c\phi_L^c(q_L)$ of workers of ability q_L.

The optimal choice of worker type by a firm in industry i requires that

$$\frac{q_L\psi_{iL}[m(q_L), q_L]}{\gamma_i\psi_i[m(q_L), q_L]} = \varepsilon_w(q_L) \tag{10}$$

where $\psi_{iL} \equiv \partial\psi_i(q_H, q_L)/\partial q_L$ is the marginal contribution of worker ability to productivity in industry i (given the ability of the worker's manager) and $m(q_L)$ is the ability of the manager that

is paired with workers of type q_L. Similarly, the optimal choice of manager implies

$$\frac{m(q_L)\,\psi_{iH}[m(q_L),q_L]}{(1-\gamma_i)\,\psi_i[m(q_L),q_L]} = \varepsilon_r[m(q_L)], \tag{11}$$

where $\psi_{iH} \equiv \partial\psi_i(q_H,q_L)/\partial q_H$ is the marginal contribution of manager ability to productivity in industry i, $r(q_H)$ is the salary schedule as a function of managerial ability, and $\varepsilon_r(q_H)$ is the elasticity of that schedule with respect to ability. Notice that (10) is analogous to (5), which applies when all managers are similar, except that now the elasticity of productivity with respect to worker ability can depend on the type of the manager with whom the workers are matched. Also, since the elasticity of output in sector i with respect to the number of managers is $1-\gamma_i$, (11) has a similar interpretation to (10), equating the ratio of the elasticity of productivity with respect to ability to the elasticity of output with respect to quantity (on the left-hand side) with the elasticity of the factor price with respect to ability (on the right-hand side).

Consider first, as Grossman et al. do, the special case that arises when

$$\psi_i(q_H,q_L) = q_H^{\beta_i} q_L^{\alpha_i}. \tag{12}$$

In this "Cobb-Douglas" case, the elasticity of productivity in sector i with respect to managerial ability is a constant β_i, independent of the composition of the production unit in which the manager participates, and the elasticity of productivity with respect to worker ability is α_i, independent of the ability of the manager in charge. Then the left-hand side of (10) is the constant α_i/γ_i and the left-hand side of (11) is the constant $\beta_i/(1-\gamma_i)$. It follows that a range of workers can be absorbed into industry i if and only if the wage schedule has a constant elasticity equal to α_i/γ_i over that range, and a range of managers can be absorbed into industry i if and only if the salary schedule has a constant elasticity equal to $\beta_i/(1-\gamma_i)$ over that range. Given such constant-elasticity wage and salary schedules, firms will earn the same profits no matter which type of manager they team with which type of workers among those that sort to the industry. In other words, the matching of managers and workers is indeterminate in this case. The sorting pattern, in contrast, is clear enough; high-ability workers sort to sector 1 if and only if $\alpha_1/\gamma_1 > \alpha_2/\gamma_2$, whereas high-ability managers sort to sector 1 if and only if $\beta_1/(1-\gamma_1) > \beta_2/(1-\gamma_2)$.

The determinants of the trade pattern in the case of Cobb-Douglas productivity are similar to those for a two-factor economy with heterogeneous workers and homogeneous managers. If two countries share the same distributions of worker and manager abilities and if one country has a relatively larger number of managers compared to workers than the other, then the country that is relatively abundant in managers produces relatively more of the good that uses managers more intensively in the production process; i.e., the good i with the smaller γ_i. If, instead, the two countries have the same relative numbers of managers versus workers and identical distributions of manager ability but one has more able workers in the sense of a proportional rightward shift in distribution, then the country with the better distribution of workers produces relatively more in the industry in which ability contributes more elastically to productivity; i.e., the good i with

the larger α_i. Finally, if the two countries have the same relative numbers of managers versus workers and identical distributions of worker ability but one has more able managers in the sense of a proportional rightward shift in distribution, then the country with more able managers will produce relatively more of the good for which manager ability contributes more elastically to productivity; i.e., the good i with the larger β_i.

With Cobb-Douglas synergies between heterogeneous workers and heterogeneous managers, the effects of trade on income distribution also are similar to those in an economy with homogeneous managers and a more general productivity function. First, trade has no effect on within-industry wage distributions, because wages in sector i must rise with a constant elasticity α_i/γ_i as a function of worker ability in order that firms are indifferent among the types that sort there. The relative wage of any two workers employed in sector i is equal to their relatively ability level raised to the power α_i/γ_i, irrespective of the country's openness to trade. The between-industry wage distribution again reflects a mix of Stolper-Samuelson and Ricardo-Viner forces. If $\alpha_1/\gamma_1 \approx \alpha_2/\gamma_2$ and $\beta_1/(1-\gamma_1) \approx \beta_2/(1-\gamma_2)$, then an opening of trade boosts the real wage of all workers in the country that exports the labor-intensive good and reduces the real wage of all workers in the country the imports that good. The assumptions that $\alpha_1/\gamma_1 \approx \alpha_2/\gamma_2$ and $\beta_1/(1-\gamma_1) \approx \beta_2/(1-\gamma_2)$ mute the forces for sorting across industries, leaving only the the aggregate effect of trade on the demand for labor versus managers. In contrast, when $\gamma_1 \approx \gamma_2$, the opening of trade raises the real wage of all workers initially in a country's export sector and reduces the real wage of workers who remain in the import-competing sector. In this case, the similarity of factor intensities in the two sectors means that relative demands for the factors are unchanged as one sector expands and the other contracts. But workers with different ability levels have different comparative advantage in the two industries and those with a relative advantage in the expanding sector gain while those with a relative advantage in the contracting sector lose.

Grossman et al. proceed to examine the determinants and effects of trade in an environment with stronger complementarities between worker ability and manager ability than are implied by the Cobb-Douglas form for $\psi_i(q_H, q_L)$. In particular, they consider a setting in which $\psi_i(q_H, q_L)$ is *strictly* log supermodular for $i = 1, 2$. In equilibrium, the best workers with ability greater than some q_L^* and the best managers with ability greater than some q_H^* may sort to sector 1 while the remaining workers and managers sort to sector 2. Alternatively, the best workers may sort to sector 1, while the best managers sort to sector 2. The authors provide sufficient conditions for each of these outcomes. Still other sorting patterns are possible; for example, the workers that sort to industry 1 may include the most and least able among those in the workforce but not workers with intermediate ability levels. In any case, there is always positive assortative matching between the workers and the managers employed in a given industry.

Suppose that the equilibrium sorting pattern is characterized by a pair of cutoff ability levels, q_L^* and q_H^*, such that a worker sorts to industry 1 if and only if his ability exceeds q_L^* and a manager sorts to either sector 1 or sector 2 (as the case may be) if and only if his ability exceeds q_H^*. Grossman et al. refer to an equilibrium in which the best types of both factors sort to industry 1 as an HH/LL

equilibrium and one in which the best workers sort to industry 1 while the best managers sort to industry 2 as an HL/LH equilibrium. In either case, as they show, if the distribution of abilities of each factor is the same in two countries then the country that has relatively more workers compared to managers exports the labor-intensive good.

The setting with two heterogeneous factors and the strong complementarities that are associated with a log supermodular productivity function is one in which trade affects both the within-industry and between-industry distributions of income. In an HL/LH equilibrium, the opening of trade draws more workers and more managers into the export sector. Consider, for example, a country that exports good 1, the good produced by the industry that attracts the best workers (and the worst managers). An increase in the price of good 1 attracts marginal workers to the industry that are less able than those employed their initially, while it draws marginal managers to the industry that are more able than those employed there initially. As a result, workers initially employed in industry 1 find improved matches after the opening of trade, whereas managers initially employed there see their matches deteriorate. The complementarities between worker and manager ability imply that the improved matching benefits especially the workers of greatest ability. Therefore, the opening of trade induces a steepening of the wage profile in the export industry and a spread in the wage distribution there. Meanwhile, in the import-competing sector, the workers that depart after an opening of trade are more able than those that remain in the industry, whereas the managers that leave are less able than those that remain behind. As a result, those workers who remain in the industry are able to find better matches than before, while the managers that stay behind suffer the opposite fate. Here, too, the wage schedule steepens and the distribution of wages becomes more unequal. Managers' salaries become more compressed in both industries due to the deterioration in the quality of their matches.

The effects of trade on between-industry distribution are reminiscent of those described before. Trade may benefit all workers in the economy if the export industry is labor intensive. But even if export are labor intensive, the Stolper-Samuelson forces can be overturned for those types of workers that are relatively most productive in the import-competing industry. And now there is a third consideration. The return to worker ability rises in an industry if and only if the matching with managers improves. Improved matching in a sector benefits especially those workers at the top end of the industry's ability (and wage) distribution.

Grossman and Maggi (2000) study sorting and matching in an economy with heterogenous production teams. They suppose that production in each of two industries requires teams of workers performing different tasks. For simplicity, let there be two such tasks in each sector, so that teams everywhere comprise pairs of workers. In sector i, the output of a team depends on the abilities of the worker who perform each of the tasks. If task a is performed by a worker of ability q_H and task b is performed by a worker of ability q_L, then the team output when it operates in industry i is given by $x_i = \psi_i(q_H, q_L)$.

Grossman and Maggi assume that the tasks in an industry contribute symmetrically to productivity, in the sense that $\psi_i(q_H, q_L) = \psi_i(q_L, q_H)$. They also assume constant returns to worker

ability in both sectors; i.e., $\psi_i(q_H, q_L) = q_H \psi_i(q_H/q_L, 1)$ for $i = 1, 2$. They study trade between two countries that have symmetric distributions of worker ability with similar mean ability levels but a difference in diversity. Diversity is defined here in terms of a single crossing of the cumulative distribution functions at some q'; the country with the more diverse distribution has a greater fraction of its workers with ability less than q if and only if $q < q'$. The countries are otherwise similar and in particular neither enjoys a comparative advantage in producing either of the goods.

Suppose first that both $\psi_1(q_H, q_L)$ and $\psi_2(q_H, q_L)$ are supermodular functions; i.e., that the two tasks are complementary in producing output in both industries, albeit possibly to differing degrees. When the tasks are complementary, efficiency requires "self-matching"; in every production team, the ability of the individual performing task a should be matched as closely as possible to that of the individual performing task b. With self-matching in both sectors and constant returns to ability, a difference in diversity of the labor force creates no comparative advantage in either sector. Since the authors assume an absence of technological comparative advantage, it follows that two countries with different distributions of talent do not trade.

But now suppose that whereas $\psi_1(q_H, q_L)$ is supermodular, $\psi_2(q_H, q_L)$ instead is submodular. That is, the two tasks needed for producing good 2 are substitutable in the sense that it is better to have one performed very well and the other poorly than to have both performed at an intermediate level. With a submodular technology, efficiency dictates "maximal cross-matching"; among the resources that are devoted to industry 2, the most output is produced when the most able worker is teamed with the least able worker, the second most able worker is teamed with the second worst worker, and so on. In the competitive equilibrium that results for any relative price, a country devotes its most and least able workers to production in industry 2, with cross-matching there, and its workers of intermediate ability to industry 1, where self-matching prevails. The country that has the more diverse labor force has more workers in the tails of the distribution and these workers, when combined in production teams, are especially productive in industry 2. It follows that diversity breeds comparative advantage in sectors with submodular production technologies relative to those with supermodular technologies.

Grossman and Maggi discuss the distributional implications of trade between countries that differ in diversity. In the country with the more diverse work force, trade benefits the least able workers, harms a set of workers with intermediate levels of ability, and may benefit or harm the most able workers. The ambiguity for these latter workers reflects the fact that an increase in the price of good 2 redistributes income to workers who are best suited for employment in that sector, but also to the less able worker amongst any pair of workers producing with the submodular technology.

8 Sorting and Matching with Labor Market Frictions

Most of the existing literature that studies interactions between worker heterogeneity and international trade neglects frictions in the labor market. In the papers that I have described thus far, the

authors have assumed that firms can readily observe the abilities and other attributes of all workers they might potentially hire, that these firms can fill their vacancies immediately and costlessly, that workers can switch between jobs without spells of search and unemployment, and that a worker's decision to accept an offer imposes no externalities on others in the labor market. Of course, labor markets are plagued by such frictions, which impede the efficient sorting of labor to industries and the efficient matching of workers to heterogeneous firms and with others in the formation of production teams. In this section, I describe a few recent papers that seek to understand how trade affects matching and sorting, and consequently the distribution of income, in settings with labor market frictions of various sorts.

Grossman (2004) studies an economy without search frictions or unemployment in which firms are nonetheless hampered by their inability to observe all attributes that contribute to a worker's productivity and workers by their inability to observe the output or profits realized by their employers. In such a setting, employment contracting is bound to be imperfect, because firms cannot link compensation to ability, nor can they offer piece rates or bonus schemes that reward an individual's productivity. The imperfect contracts that can be offered provide for a fixed wage irrespective of ability and performance and thus they create the possibility that workers will make inefficient choices of occupation and industry.

Suppose there are two countries and two industries. One industry requires team production, with teams of size two that perform complementary tasks. Output of a pair with abilities q_a and q_b is given by $\psi(q_a, q_b)$, where the productivity function $\psi(\cdot)$ is supermodular, symmetric, homogeneous of degree one, and has an elasticity of substitution less than or equal to one. The output produced by a team is observable only to the one designated as "owner", who is the residual claimant among the pair and pays the other a fixed wage. In the second industry, output is linear in ability and workers can produce alone. Ability levels are distributed differently and exogenously in the two countries, which are otherwise similar in regard to tastes and technology.

The fixed-wage contracts that are paid in the sector with team production induce adverse selection. An owner's expected profit in that sector depends on the wage he offers and the average output produced by workers who seek positions as employees, but the wage that appeals to a worker who can produce an average level of output might not appeal to one that can produce much more than average. These more-able individuals can choose instead to become owner-managers of firms in the industry with team production, or else to work as individuals in the sector with solo production.

Grossman shows that the equilibrium has a unique and determinate pattern of industry sorting and occupational choice. For given prices, a set of the least-able individuals in any economy opts to become employees and collect a fixed wage, a set of intermediate-ability individuals chooses to become owner-managers in the sector with team production, while the most-able individuals sort to the industry with individual production and performance pay. The sizes of the first two sets must be equal to allow for pairings within each production team. The size of the third set ensures that the most able (and risk neutral) owner-manager earns the same expected profit in the sector

with team production as he would earn on his own in the sector with solo production. This, in turn, depends on the relative price of the two goods. The sorting and occupational choice reflects the fact that greater ability earns no marginal return among workers, it earns positive but less than complete marginal return among managers, and it earns a proportional return in the sector with individual production. The equilibrium employment pattern is inefficient (compared to a hypothetical economy with perfect labor contracting) for two reasons. First, workers and managers are not perfectly self-matched in the sector with team production, as is optimal for a productivity function that is symmetric and supermodular. Second, there are too few individuals in the sector with team production, because those that opt for the industry with performance pay do not take into account that their choice imparts an externality on those in the sector with team production by depressing the talent pool there. Were the marginal individual in the sector with solo production to choose instead to work in the sector with team production, his own expected income would be the same, but the expected income of others in the industry would be higher.

At any given relative price, an individual with some (high) ability level q faces a greater incentive to join the industry with solo production in the country that has the larger share of low-ability workers. That is, if the countries have distributions of ability with the same mean but one has a greater spread than the other, then an individual with ability q will anticipate a less favorable pool of potential partners and a smaller expected profit from team production in the country with the larger lower tail of abilities than in the country with the more tightly distributed talent levels. Accordingly, there exists a set of ability levels such that every individual with ability in this set opts for team production in the country with the tighter distribution of talents and for solo production in the country with greater talent diversity. These differences in the incentives for occupational choice can create an opportunity for trade between otherwise similar countries. For example, if the countries both have uniform distributions of ability with the same mean but different supports, the country with the greater spread of talents exports the good that is produced by individuals and imports the good that is produced by teams.

The trade that results from differences in the distributions of ability in the face of imperfect labor contracting has notable consequences. First, this trade causes a deterioration in the income distribution in the country that imports the goods produced by teams. On the margin, an increase in the relative price of the good generated by individual production induces the most talented individuals to abandon the sector with team production, thereby degrading the talent pool for those that remain behind. This degradation of the talent pool in the sector with team production reduces average productivity there, and therefore depresses wages for employees and profits for owner-managers. Meanwhile, the change in relative price implies that the compensation for those at the top of the ability distribution rises. Moreover, the departure of the most able individuals from the sector with team production exacerbates the aforementioned distortion in the allocation of talent. Since there are fewer workers in the sector with team production than is socially optimal to begin with, further specialization in the sector with individual production in the country with the more diverse talent pool might reduce national income even as it worsens the distribution of

income there.

Bombardini et al. (2011) also study an economy in which firms are unable to observe perfectly the ability levels of their potential hires. Their model bears a family resemblance to Grossman and Maggi (2000) inasmuch as production requires teams of workers and industries differ in the substitutability or complementarity of employee's talents. Whereas Grossman and Maggi predict that otherwise similar countries that differ only in their distributions of ability will not trade when productivity functions in all industries are supermodular, Bombardini et al. (2011) draw a different conclusion in the face of imperfect matching and sorting. They assume a search environment such that firms in every sector sample randomly from the country's talent distribution. As a consequence, perfect matching of workers with similar ability levels is not possible and efficient sorting of abilities to industries does not obtain. The inefficient matches that result have implications for relative outputs across sectors. In particular, the random matches are particularly costly in terms of expected output in industries where the complementarities between worker talents are greatest. In a setting with many countries and many goods, the authors derive sufficient conditions under which a country with a lesser dispersion of ability levels exports the goods that are characterized by greater complementarities in the team production processes.

Davidson et al. (2008) focus more narrowly on how globalization affects the mismatches that can result from search frictions. Their starting point is a model of endogenous technology choice and endogenous selection into exporting, as in Yeaple (2005), and a model of the search process based on Albrecht and Vroman (2002). In any country, there are fixed supplies of two types of workers, those with high ability, H, and those with low ability, L. A firm requires a single worker to produce output and combines its worker with a chosen quantity of other inputs, which the authors call "capital." The firm has a choice of two technologies to produce the homogeneous final good. The more advanced technology A can only be operated by the high-ability workers, whereas the more basic technology B can be operated by workers of either type. The more able workers are, of course, assumed to be more productive than the less able workers, but especially so if they operate the more advanced technology. In particular, output of a firm that employs a technology of type $q_T = \{A, B\}$, and worker of type $q_W = \{H, L\}$, and k units of capital is $x = \psi(q_T, q_W) k^{\alpha}$, where $\psi(A, H) > \psi(B, H) > \psi(B, L)$ and $\psi(B, L) = 0$. Each firm also faces a choice whether to sell its output on the domestic market or on the foreign market.[14] If it sells domestically, it pays a flow cost c_d to maintain a local distribution network whereas if it exports, it pays the higher flow cost, $c_f > c_d$. The prices in the two markets are p_d and p_f, where the former is endogenously determined to equate supply and demand and the latter is taken as exogenous.

Upon entry, a firm chooses its technology and posts a vacancy. The vacancy entails a flow cost and generates a Poisson arrival of randomly selected job seekers. The arrival rate of workers for any firm with a posted vacancy depends on the ratio of the aggregate number of vacancies to the number of unemployed workers. When a worker and firm meet, the worker decides whether to

[14] Here, unlike in Yeaple (2005) or Melitz (2003), except in a knife-edge case, every firm prefers to sell all of its output in one market or the other. Therefore, domestic firms do not export, and export firms do not sell domestically.

accept employment with the firm and, if so, the two sides bargain over the surplus. Both sides rationally anticipate whether the firm will subsequently sell to the domestic market or the foreign market. Each existing job terminates with some fixed and exogenous flow probability.

The authors note that two types of equilibria are possible in their model. In a "cross-skill matching" equilibrium, every high-ability worker accepts the first job that comes along. If he happens to match with a firm that has chosen an advanced technology, the match is a good one thanks to the complementarity between technology and ability, and the worker earns a high wage. If, instead, he matches first with a firm that has chosen the basic technology, the match is less good and the wage is lower. But for parameters that satisfy the conditions for this type of equilibrium, it is not worthwhile for the worker to remain unemployed and wait for a better match. In an "ex-post segmentation" equilibrium, by contrast, the wage gap between what the H worker can earn in a B firm versus an A firm is sufficiently large that waiting is the preferred strategy. It is possible that an equilibrium of either type exists for the some parameters, each supported by the workers' expectations of what others will do.

Davidson et al. focus mainly on the equilibrium with cross-skill matching, where the acceptance of basic jobs by the more able workers represents a sort of mismatch between worker and firm types.[15] Note that the ramifications of this mismatch are not limited to the high-ability workers, as the filling of vacancies by these workers in firms with basic technologies affects also the employment prospects for the less able workers.

The authors first show that the firms that choose the advanced technology and match with the high-ability workers face the strongest incentives to sell to the foreign market. The firms that choose the basic technology and match with the less-able workers face the weakest such incentives. The incentives facing the mismatched firms are somewhere in between. Thus, there are foreign prices and cost parameters such that only firms with advanced technology and high-ability workers export, such that all firms with high-ability workers export, and such that firms with any of the three combinations of technology and worker export. Now consider the effects of a reduction in the flow cost of serving the foreign market. If the economy starts in a cross-matching skill equilibrium and remains in one after the fall in c_f, then industry productivity is enhanced by the reallocation of market shares in favor of firms using the advanced technology, and the frequency of mismatch between high-ability workers and firms that use the basic technology falls. The reduction in trade costs increases the surplus created by matches between advanced firms and high-ability workers, and so the H workers employed by A firms gain from liberalization of this sort. Since the outside opportunities for high-ability workers that happen to match with the other firms improve, the wages of H workers rise even among those employed by firms using the basic technology. Meanwhile, the shrinking percentage of firms that use the basic technology implies fewer job opportunities for the less able workers and, with their diminished bargaining power these workers might lose from globalization.

[15] Given the costliness of search, it is not necessarily socially optimal for the high-ability workers to decline offers from firms with the basic technology. But the externalities associated with the employment decision suggest that cross-matching skill equilibrium will be inefficient for some parameter values.

A fall in trade costs can increase the profitability of firms that choose the advanced technology and export relative to firms that choose the basic technology and sell domestically to such an extent that the cross-matching skill equilibrium ceases to exist. In the equilibrium with ex-post segmentation that replaces it, there is no longer the possibility of mismatch between workers and technologies. In the event, the aggregate productivity gains for the economy can be quite substantial as the number of firms adopting the basic technology falls dramatically. It is interesting to note that the aggregate productivity gains are realized despite the fact that the weak firms that still enter suffer an expected loss in productivity from the fact that they no longer have any chance to hire the more able workers.

Davidson et al. also consider the impact of falling trade costs on productivity in import-competing industries. A fall in the cost of imports reduces the gap in revenues between firms that choose the advanced technology (in this case, to sell domestically) and firms that choose the basic technology. This makes it easier for firms using the basic technology to attract the more able workers. It follows that, if the economy begins in an equilibrium with ex-post segmentation in an import-competing industry and trade costs fall, the increased openness can cause a switch to an equilibrium with cross-skill matching. Alternatively, if the initial equilibrium already has cross-skill matching, then the frequency of mismatch increases with openness to trade. Accordingly, the model predicts that globalization generates a more efficient allocation of labor and a boost to productivity in export industries but the opposite outcomes in import-competing industries.

These predictions are more nuanced in an extended version of the model incorporating monopolistic competition and intra-industry trade that has been developed by Davidson and Matusz (2012). In this variant, firms initially draw their types from a continuous distribution, as in Melitz (2003), albeit with a fixed number of firms. The productivity function exhibits complementarity between worker and firm types. It follows that firms with a technology parameter that exceeds some critical value will only hire the high-ability workers, whereas firms with a less sophisticated (and less productive) technology will hire whichever worker they match with first. In this setting, a reduction in the fixed cost of exporting for domestic firms has a non-monotonic impact on the degree of mismatch in the labor market. When domestic firms face a high fixed cost of exporting, a fall in these costs reduces the critical value for technical sophistication beyond which firms are selective in their hiring. As a result, the frequency of mismatch falls. However, when fixed exporting costs are already low for domestic firms and fall further, the result is reversed. In contrast, a decline in the fixed cost of exporting for foreign firms, which implies an intensification of import competition, has an ambiguous effect on the frequency of mismatch between firms and workers. The increased competition from imports harms the more technologically sophisticated (and thus larger) firms more while they are producing, thereby shrinking the revenue gap between selective and non-selective firms. But selective firms spend a greater fraction of the their time than non-selective firms with unfilled vacancies—due to their greater selectivity—which reduces the profit gap between the two types of firms. The presence of these offsetting effects accounts for the ambiguous overall finding.

Perhaps the most complete analytical framework that has been developed to date for studying

resource allocation and income distribution in an open economy with labor market frictions is provided by Helpman et al. (2010). Their model includes imperfect information about worker ability, search frictions for workers and firms, matching of heterogeneous workers and heterogeneous firms, and equilibrium unemployment as a component of worker income variability.

The authors begin by describing a "sectoral equilibrium" in which workers and firms take the employment and wage opportunities outside the sector as given. The heterogeneous firms in the industry produce Dixit-Stiglitz differentiated products using heterogeneous workers as their only input. Worker productivity in a firm depends upon characteristics of the firm and on the ex post, match-specific realization of the worker's ability to serve the firm. In our notation, the output of a firm of type q_H that hires ℓ workers of *average* ability $\bar{q}_L$ is

$$x = \psi\left(\bar{q}_L, q_H\right) f\left(\ell\right) = q_H \bar{q}_L \ell^{\gamma}.$$

The supermodularity of the productivity function generates incentives for PAM. But neither a firm nor the worker know anything about the latter's match-specific ability until after a match has been consummated. At that point, the firm can pay a screening cost that allows it to determine whether or not a worker's ability in the firm exceeds $\tilde{q}_L$. The screening cost is assumed to be an increasing and concave function of the threshold level, $\tilde{q}_L$; i.e., it is more costly for the firm to distinguish a narrower set of more able workers, presumably because this requires a more subtle and complex test. A worker's ability in a given match is drawn from a Pareto distribution that is independent across workers and firms.

Firms first observe their type, then decide whether or not to bear the fixed cost of production, whether or not to bear the fixed cost of exporting, how many workers to sample for potential hire, and what threshold to use in screening them. Due to the presence of search frictions, the firm incurs a cost for each worker that it samples, which depends in familiar Diamond-Mortensen-Pissarides style on the "tightness" of the labor market; i.e., the ratio of the aggregate number of workers sampled to the number of workers seeking employment in the sector. If a worker passes a firm's screening test, the two sides bargain over a wage as in Stole and Zwiebel (1996a, 1996b) subject to their common understanding that the worker's match-specific ability exceeds $\tilde{q}_L$.

The more productive firms have a greater incentive to screen, thanks to the complementarities between worker ability and firm technology. Therefore, these higher-type firms screen more intensively in equilibrium, reject workers with match-specific ability below their higher threshold levels, and end up with a workforce of higher average ability. Moreover, as in Melitz (2003), it is the more productive firms that select into exporting. The export opportunities associated with the opening of trade enhance the incentives for the productive firms to screen their potential employees and to apply a more exacting standard in their hiring decisions. Therefore, in a trade equilibrium, the exporting firms ensure themselves of a workforce of higher average ability than non-exporters and they pay higher average wages.

Helpman et al. are interested in the connection between trade, income distribution, and unemployment. They show first that the opening of trade in a closed economy raises within-industry

wage inequality. Since the larger firms pay higher wages and the opening of trade raises revenues especially for the large firms that export, it follows that wage dispersion is greater in the open than in the closed economy. Surprisingly, however, the relationship between openness to trade and wage inequality is non-monotonic. Once trade costs are low enough that all firms export, an increase in these costs that causes some of the least productive firms to cease their export activity will reduce the wages paid by these firms relative to the more productive firms that continue to export,and thus wage inequality increases.

Next, the authors address the subtle interplay between wage inequality and structural unemployment. The unemployment rate reflects both the fraction of workers searching for jobs who are lucky enough to be sampled by some firm and the stringency of the screening tests they face conditional on finding a job opportunity. The more productive firms pay higher wages, but they also turn away a greater percentage of the workers they screen. When trade opens, there are two effects on unemployment. First, the more productive firms expand at the expense of the less productive, which in turn implies more stringent screening that contributes to higher unemployment. Second, the tightness of the labor market changes and with it the fraction of workers who are able to find job opportunities. Labor market tightness can increase or decrease as trade costs fall, and indeed so can the unemployment rate.

Helpman et al. conclude by showing how their model of industry outcomes can be embedded in a general equilibrium. They consider two alternative approaches, one that includes an "outside sector" where output is competitively produced and no labor market frictions exist and another that excludes any outside sector, so that the number of job seekers in the sector equals the economy's exogenous labor supply. Then, a worker's outside opportunity is determined endogenously. They show, for example, that in either case, if workers are risk neutral, their ex ante expected utility must rise with the opening of trade.

9 Concluding Remarks

Models of trade with heterogeneous labor afford the researcher the opportunity to study how fine details about the skill distribution affect trade patterns and how globalization effects the distribution of wages and unemployment across the complete spectrum of worker types. The papers that have been reviewed here provide a rich set of predictions. Some are extensions of familiar results that apply when labor is homogeneous. For example, when the different workers in a country have equal relative productivity in different activities, the trade pattern is determined by comparative technological differences across countries. When countries have similar distributions of worker types, then trade is governed by countries' relative abundance of different factors in terms of quantities. And when globalization causes a change in relative prices, there are Stolper-Samuelson forces that favor factors used intensively in expanding industries and Ricardo-Viner forces that favor factor types that have comparative advantage in those industries.

Others predictions are new. In some settings, diversity of factor types can be an independent

source of comparative advantage and trade between countries that have similar aggregate endowments of factors. This is true, for example, when workers' skills or abilities are complementary contributors to team production in one sector, but substitutable in another, or when the extent of complementarity between the abilities of team members differs across sectors and imperfect information impedes the formation of efficient production units. Integration also affects world resource allocation in production hierarchies when managers in one country can use their superior skill and knowledge to solve the production problems that confront workers in another.

In some cases, the models with heterogeneous labor predict that an opening of trade will affect within-industry income distribution. This is typically true when matching between firms (or managers) and workers takes place in an environment with strong complementarities between types and trade impacts the matches that result. In a country where trade improves the matches that workers achieve in an industry—resulting in these workers finding employment with either better managers or better technologies than before—then within-industry wage inequality will increase. Upgraded matching tends to benefit all workers in an industry, but especially those at the higher end of the ability and wage distribution. Also, models in the spirit of Melitz (2003) that predict that the larger and more productive firms are the ones that export also typically predict an exporter wage premium. This premium can come about as a result of workers' sharing in the rents earned by these better firms, but more typically they reflect endogenous differences in the composition of hires. The "better" firms may have superior managers or technologies that account for their greater productivity and these attributes of the manager or technology may dictate positive assortative matching with workers. In the event, the more productive firms that export will also be ones that hire the more skilled or higher-ability workers.

Particularly promising are a recent set of models that introduce labor market frictions into a world with heterogeneous workers. In such a setting, trade affects not only the distribution of income, but also the efficiency of resource allocation. In some cases that have been studied, globalization reduces the mismatch between technologies and worker types that results from the inability of firms to find costlessly and immediately their ideal workforces. But in other cases, trade may dampen the incentives for team production in a country where the number of individuals willing to lead production teams is inefficiently small. Much work remains to be done on modeling the search process and labor contracting before we will understand the subtle effects of openness on labor market performance and determine what, if any, policy remedies are indicated.

References

[1] Abowd, John M., Kramarz, Francis and Margolis, David N., 1999. "High Wage Workers and High Wage Firms," *Econometrica 67(2)*, 251–334.

[2] Albrecht, James and Vroman, Susan, 2002. "A Matching Model with Endogenous Skill Requirements," *International Economic Review 43(1),* 283-305.

[3] Antràs, Pol, Garicano, Luis and Rossi-Hansberg, Esteban, 2006. "Offshoring in a Knowledge Economy," *Quarterly Journal of Economics 121(1)*, 31-77.

[4] Autor, David H., Katz, Lawrence F., and Kearney, Melissa S., 2008. "Trends in U.S. Wage Inequality: Revising the Revisionists," *Review of Economics and Statistics 90(2)*, 300-23.

[5] Bagger, Jesper and Lentz, Rasmus, 2012. "An Empirical Model of Wage Dispersion with Sorting." Downloadable at https://sites.google.com/a/wisc.edu/rasmus-lentz/research.

[6] Bernard, Andrew B. and Jensen, J. Bradford, 1995. "Exporters, Jobs, and Wages in U.S. Manufacturing: 1976-1987," *Brookings Papers on Economic Activity: Microeconomics 1995*, 67-119.

[7] Bombardini, Matilde, Gallipoli, Giovanni and Pupato, Germán, 2011. "Unobservable Skill Dispersion and Comparative Advantage." http://faculty.arts.ubc.ca/mbombardini/skilldisp-theory.pdf

[8] Bombardini, Matilde, Gallipoli, Giovanni and Pupato, Germán, 2012. "Skill Dispersion and Trade Flows," *American Economic Review 102(5)*, 2327-48.

[9] Bougheas, Spiros and Riezman, Raymond, 2007. "Trade and the Distribution of Human Capital," *Journal of International Economics 73(2)*, 421-33.

[10] Bustos, Paula, 2011. "Trade Liberalization, Exports, and Technology Upgrading: Evidence on the Impact of MERCOSUR on Argentinian Firms," *American Economic Review 101(1)*, 304-40.

[11] Cahuc, Pierre, Postel-Vinay, Fabien and Robin, Jean-Marc, 2006. "Wage Bargaining with On-the-Job Search: Theory and Evidence, *Econometrica 74(2)*, 323–64.

[12] Costinot, Arnaud, 2009. "An Elementary Theory of Comparative Advantage," *Econometrica 77(4)*, 1165-92.

[13] Costinot, Arnaud and Vogel, Jonathan, 2010. "Matching and Inequality in the World Economy," *Journal of Political Economy 118(4)*, 747-86.

[14] Davidson, Carl and Matusz, Stephen J., 2012. "A Model of Globalization and Firm-Worker Matching: How Good is Good Enough?" *Review of Economics and Finance 23*, 5-15.

[15] Davidson, Carl, Matusz, Stephen J. and Shevchenko, Andrei, 2008. "Globalization and Firm Level Adjustment with Imperfect Labor Markets," *Journal of International Economics 75(2)*, 295-309.

[16] Davidson, Carl, Heyman, Fredrik, Matusz, Stephen J., Sjoholm, Fredrik, and Zhu, Susan, 2010. "Globalization and Imperfect Labor Market Sorting," Leverhulme Centre for Research on Globalization and Economic Policy Research Paper 2010/30.

[17] Eeckhout, Jan and Kircher, Philipp, 2011. "Identifying Sorting – in Theory," *Review of Economic Studies 78(3)*, 872-906.

[18] Eeckhout, Jan and Kircher, Philipp, 2012. "Assortative Matching with Large Firms: Span of Control over More versus Better Workers." http://personal.lse.ac.uk/kircher/Papers/sorting-and-factor-intensity.pdf

[19] Garicano, Luis, and Rossi-Hansberg, Esteban, 2006. "Organization and Inequality in a Knowledge Economy," *Quarterly Journal of Economics 121(4)*, 1383-1435.

[20] Grossman, Gene M., 1983. "Partially-Mobile Capital: A General Approach to Two-Sector Trade Theory," *Journal of International Economics 15(1-2)*, 1-17.

[21] Grossman, Gene M., 2004. "The Distribution of Talent and the Pattern and Consequences of International Trade," *Journal of Political Economy 112(1)*, 209-39.

[22] Grossman, Gene M., Helpman, Elhanan, and Kircher, Philipp, 2013. "Matching and Sorting in a Global Economy." Downloadable at http://www.princeton.edu/~grossman/grossman_working_papers.htm.

[23] Grossman, Gene M. and Maggi, Giovanni, 2000. "Diversity and Trade," *American Economic Review 90(5)*, 1255-75.

[24] Helpman, Elhanan, Itskhoki, Oleg and Redding, Stephen, 2010. "Inequality and Unemployment in a Global Economy," *Econometrica 78(4)*, 1239-83.

[25] Helpman, Elhanan, Itskhoki, Oleg, Redding, Stephen, and Muendler, Marc-Andreas, 2012. "Trade and Inequality: From Theory to Estimation," NBER Working Paper No. 17991, National Bureau of Economic Research.

[26] Lopes de Melo, Rafael, 2009. "Sorting in the Labor Market: Theory and Measurement." Downloadable at https://sites.google.com/site/rlopesdemelo/research.

[27] Matsuyama, Kiminori, 1992. "A Simple Model of Sectoral Adjustment," *The Review of Economic Studies 59(2)*, 375-87.

[28] Melitz, Marc, 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregrate Industry Performance, *Econometrica 71(6)*, 1695-1726.

[29] Mussa, Michael, 1982. "Imperfect Factor Mobility and the Distribution of Income," *Journal of International Economics 12(1-2)*, 125-141.

[30] Ohnsorge, Franziska and Trefler, Daniel, 2007. "Sorting it Out: International Trade with Heterogeneous Workers," *Journal of Political Economy 115(5)*, 868-92.

[31] Postel-Vinay, Fabien and Robin, Jean-Marc, 2006. "Microeconometric Search-Matching Models and Matched Employer-Employee Data", in R. Blundell, W. Newey and T. Persson, eds, *Advances in Economics and Econometrics: Theory and Applications*, Cambridge: Cambridge University Press.

[32] Ruffin, Roy, 1988. "The Missing Link: The Ricardian Approach to the Factor Endowment Theory of International Trade," *American Economic Review 78(4)*, 759-72.

[33] Sampson, Thomas, 2012. "Selection into Trade and Wage Inequality," CEP Discussion Papers No. 1152, Centre for Economic Policy Performance, London School of Economics.

[34] Schank, Thorsten, Schnabel, Klaus and Wagner, Joachim, 2007. "Do Exporters Really Pay Higher Wages? First Evidence from German Linked Employer-Employee Data," *Journal of International Economics 72(1)*, 52-74.

[35] Stole, Lars A. and Zwiebel, Jeffrey, 1996a. "Organizational Design and Technology Choice under Intrafirm Bargaining," *American Economic Review 86(1)*, 195-222.

[36] Stole, Lars A. and Zwiebel, Jeffrey, 1996b. "Intra-Firm Bargaining under Non-Binding Contracts," *Review of Economic Studies 63(3)*, 375-410.

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