Econometrica, Vol. 78, No. 4 (July, 2010), 1239–1283

INEQUALITY AND UNEMPLOYMENT IN A GLOBAL ECONOMY

BY ELHANAN HELPMAN, OLEG ITSKHOKI, AND STEPHEN REDDING[1]

This paper develops a new framework for examining the determinants of wage distributions that emphasizes within-industry reallocation, labor market frictions, and differences in workforce composition across firms. More productive firms pay higher wages and exporting increases the wage paid by a firm with a given productivity. The opening of trade enhances wage inequality and can either raise or reduce unemployment. While wage inequality is higher in a trade equilibrium than in autarky, gradual trade liberalization first increases and later decreases inequality.

KEYWORDS: Wage inequality, international trade, risk, unemployment.

1. INTRODUCTION

TWO CORE ISSUES IN INTERNATIONAL TRADE are the allocation of resources across economic activities and the distribution of incomes across factors of production. In this paper, we develop a new framework for examining the determinants of resource allocation and income distribution in which both wage inequality and unemployment respond to trade. Our framework encompasses a number of important features of product and labor markets, as a result of which it generates predictions that match features of the data. This framework is rich, flexible, and tractable, as we demonstrate by deriving a number of interesting results on trade, inequality, and unemployment. In addition, we show how this framework can be extended in various ways and how it can accommodate different general equilibrium structures. Moreover, our framework fits squarely into the new view of foreign trade that emphasizes firm heterogeneity in differentiated product sectors.

We introduce standard Diamond–Mortensen–Pissarides search and matching frictions into a Melitz (2003) model, but unlike previous work in this area, such as Helpman and Itskhoki (2010), we also introduce ex post match-specific heterogeneity in a worker's ability. Because a worker's ability is not directly observable by his employer, firms screen workers to improve the composition of their employees. Complementarities between workers' abilities and firm productivity imply that firms have an incentive to screen workers to exclude those

[1]This paper is a combined version of Helpman, Itskhoki, and Redding (2008a, 2008b). Work on these papers started when Redding was a Visiting Professor at Harvard University. We thank the National Science Foundation for financial support. Redding thanks the Centre for Economic Performance at the London School of Economics and the Yale School of Management for financial support. We are grateful to a co-editor, four anonymous referees, Pol Antràs, Matilde Bombardini, Arnaud Costinot, Gilles Duranton, Gene Grossman, James Harrigan, Larry Katz, Marc Melitz, Guy Michaels, Steve Pischke, Esteban Rossi-Hansberg, Peter Schott, Dan Trefler, and conference and seminar participants at AEA, Berkeley, CEPR, Chicago, Columbia, Harvard, LSE, NBER, NYU, Northwestern, Penn State, Princeton, Stanford, Stockholm, Tel Aviv, UCLA, and Yale for helpful comments. The usual disclaimer applies.

 DOI: 10.3982/ECTA8640

with lower abilities. As larger firms have higher returns to screening and the screening technology is the same for all firms, more productive firms screen more intensively and have workforces of higher average ability than less productive firms. Search frictions induce multilateral bargaining between a firm and its workers, and since higher-ability workforces are more costly to replace, more productive firms pay higher wages. When the economy is opened to trade, the selection of more productive firms into exporting increases their revenue relative to less productive firms, which further enhances their incentive to screen workers to exclude those of lower ability. As a result, exporters have workforces of higher average ability than nonexporters and hence pay higher wages. This mechanism generates a wage-size premium and implies that exporting increases the wage paid by a firm with a given productivity. Both features of the model have important implications for wage inequality within sectors and within groups of workers with the same ex ante characteristics.

Our first main result is that the opening of a closed economy to trade raises wage inequality. The intuition for this result is that larger firms pay higher wages and the opening of trade increases the dispersion of firm revenues, which in turn increases the dispersion of firm wages. This result is more general than our model in the sense that it holds in a wider class of models in which firm wages are increasing in firm revenue and there is selection into export markets. We provide a proof that the opening of trade raises wage inequality for any inequality measure that respects second-order stochastic dominance, and this result holds for a class of models satisfying the following three sufficient conditions: firm wages and employment are power functions of firm productivity, exporting increases the wage paid by a firm with a given productivity, and firm productivity is Pareto distributed.[2,3]

Our second main result is that once the economy is open to trade, the relationship between wage inequality and trade openness is at first increasing and later decreasing. As a result, a given change in trade frictions can either raise or reduce wage inequality, depending on the initial level of trade openness. The intuition for this result stems from the increase in firm wages that occurs

[2]While the assumption that firm productivity is Pareto distributed is strong, this assumption is standard in the literature on firm heterogeneity and trade, provides a reasonable approximation for the observed distribution of firm sizes (e.g., Axtell (2001)), and provides a reasonable approximation for the upper tail of the observed distribution of worker wages (e.g., Saez (2001)).

[3]A number of recent studies examine similar issues in models of fair wages (e.g., Egger and Kreickemeier (2009a, 2009b) and Amiti and Davis (2008)). In these models the wage distribution depends on the formulation of the fair wage hypothesis and this formulation differs across studies: wages are *assumed* to rise with either a firm's productivity, its revenue, or its profits. For example, Egger and Kreickemeier (2009a) assumed that the fair wage is a power function of a firm's productivity. In other words, more productive firms pay higher wages *by assumption*. This implies that the relative wage of two firms with different productivity levels is the same, independently of whether one or both or neither of them exports. The mechanism in our model is quite different and our inequality results hold for all standard measures of inequality.

at the productivity threshold above which firms export, which is only present when some but not all firms export. When no firm exports, a small *reduction* in trade costs increases wage inequality, because it induces some firms to export and raises the wages paid by these exporting firms relative to domestic firms. When all firms export, a small *rise* in trade costs increases wage inequality, because it induces some firms to cease exporting and reduces the wages paid by these domestic firms relative to exporting firms.

Another key prediction of our framework is that these two results hold regardless of general equilibrium effects. To demonstrate this, we derive these results from comparisons across firms that hold in sectoral equilibrium irrespective of how the sector is embedded in general equilibrium. It follows that our results for sectoral wage inequality do not depend on the impact of trade on aggregate variables and variables in other sectors. We use our framework to derive closed-form expressions for the sectoral wage distribution. This distribution depends on an extensive margin of trade openness (the fraction of exporting firms) and an intensive margin of trade openness (relative revenue in the export and domestic markets). We characterize the relationship between these extensive and intensive margins of trade openness and the exogenous parameters of the model such as fixed and variable trade costs.

Since workers are ex ante homogeneous, wage inequality in our model is *within-group* inequality. Our theoretical results are therefore consistent with empirical findings of increased within-group wage inequality following trade liberalization (see, for example, Attanasio, Goldberg, and Pavcnik (2004) and Menezes-Filho, Muendler, and Ramey (2008)). As these theoretical results hold for asymmetric countries, they are also consistent with empirical findings of increased wage inequality following trade liberalization in both developed and developing countries (see, for example, the survey by Goldberg and Pavcnik (2007)).

While our focus is within-group inequality, we also develop an extension in which there are multiple types of workers with different observable ex ante characteristics. We show that our results for the impact of trade on within-group wage inequality hold in this more general framework; in particular, trade raises wage inequality within every group of workers. While between-group inequality of wages can rise or fall, the rise in within-group inequality can dominate when between-group inequality falls, so that overall inequality rises.

In our framework, the opening of trade results in endogenous changes in workforce composition and measured firm productivity. The increase in revenue at exporters induces them to screen workers more intensively, while the decrease in revenue at nonexporters causes them to screen workers less intensively. It follows that more productive firms experience larger increases in average worker ability and wages following the opening of trade, which strengthens the correlation between firm productivity and average worker ability and echoes the empirical findings of Verhoogen (2008). To the extent that empirical measures of productivity do not completely control for differences in workforce composition, these endogenous changes in workforce composition also result in endogenous changes in measured firm productivity. As more productive firms experience the largest increases in average worker ability following the opening of trade, they also exhibit the largest increases in measured firm productivity.

Another distinctive feature of our framework is the interaction between wage inequality and unemployment. The unemployment rate depends on the fraction of workers searching for employment that are matched (the tightness of the labor market) and the fraction of these matched workers that are hired (the hiring rate). While the more intensive screening of more productive firms implies that they pay higher wages, it also implies that they hire a smaller fraction of the workers with whom they are matched. As a result, the reallocation of resources toward more productive firms that occurs following the opening of trade reduces the hiring rate and increases the unemployment rate. In contrast, the tightness of the labor market can either remain constant following the opening of trade (as in Helpman and Itskhoki (2010)) or can rise (as in Felbermayr, Prat, and Schmerer (2008) and Felbermayr, Larch, and Lechthaler (2009)) depending on what happens to expected worker income, which in turn depends on how the model is closed in general equilibrium. Therefore, the net effect on the unemployment rate of opening a closed economy to trade is ambiguous. In contrast to the unambiguous results for wage inequality, our analysis suggests that unemployment can either rise or fall following the opening of trade, which is consistent with the lack of a clear consensus on the empirical relationship between trade and unemployment (see, for example, the discussion in Davidson and Matusz (2009)).

Worker ability admits two possible interpretations within our framework. One interpretation is that ability is match-specific and independently distributed across matches. Another interpretation is that ability is a general talent of a worker that does not depend on his match, but is unobservable to both workers and firms. In the static model that we develop here, the analysis is the same irrespective of which interpretation is taken. In both cases, workers do not know their ability and have no incentive to direct their search across firms or sectors. However, our preferred interpretation is that ability is match-specific, because this interpretation makes our static model consistent with the steady state of a dynamic search and matching model, since the screening of a worker for one match reveals no information about their ability for other potential matches.[4] Under the alternative interpretation that ability is a general talent of workers, wage inequality in the model has a worker component as well as a firm component. Since more productive firms screen more intensively

[4]This type of dynamic analysis can be done along the lines of Helpman and Itskhoki (2009), whose model has, however, no worker heterogeneity and no screening.

and pay higher wages, only more able workers receive the higher wages paid by more productive firms.[5]

Our paper is related to a large literature on trade, wage inequality, and unemployment. One broad area of research has explored the relationship between trade and wage inequality in models with neoclassical labor markets. Yeaple (2005) and Bustos (2009) developed models of monopolistic competition in which firms make endogenous choices about production technology and observable skill composition. Ohnsorge and Trefler (2007) and Costinot and Vogel (2009) examined the relationship between trade and wage inequality in competitive assignment models. Burstein and Vogel (2009) developed a model in which both comparative advantage and skill-biased technology play a role in determining the relationship between trade and wage inequality.

Another broad area of research has examined the implications of labor market frictions for the impact of trade on unemployment and wage inequality. One strand of this literature has considered models of efficiency or fair wages. Davis and Harrigan (2007) developed a model of firm heterogeneity and efficiency wages in which wages vary across firms because of differences in monitoring technology, and equilibrium unemployment exists to induce workers to supply effort. Egger and Kreickemeier (2009a, 2009b) and Amiti and Davis (2008) developed models of firm heterogeneity and fair wages in which the fair wage at which workers supply effort is assumed to vary with either firm productivity, revenue, or profits.

Yet another strand of this literature has considered models of search and matching, which provide natural microfoundations for labor market frictions.[6] In important research, Davidson, Martin, and Matusz (1988, 1999) showed that the introduction of search and matching frictions into competitive models of international trade has predictable implications for the relationship between relative goods and factor prices. Using models of firm heterogeneity with search and matching frictions, Felbermayr, Prat, and Schmerer (2008), Felbermayr, Larch, and Lechthaler (2009), and Helpman and Itskhoki (2010) examined the relationship between trade and unemployment. None of these

[5] Embedding this alternative interpretation in a dynamic framework would be more complicated, because screening for one match reveals information about a worker's productivity for other potential matches. As a result, the ex post distribution of worker ability among the unemployed would no longer be equal to the ex ante distribution of worker ability. Additionally, as workers gradually learn about their ability, more able workers would have an incentive to direct their search toward more productive firms.

[6] Seminal models of search and matching include Mortensen (1970), Pissarides (1974), Diamond (1982a, 1982b), and Mortensen and Pissarides (1994). One line of research follows Burdett and Mortensen (1998) in analyzing wage dispersion in models of wage posting and random search. Another line of research examines wage dispersion when both firms and workers are heterogeneous, including models of pure random search, such as Acemoglu (1999) and Shimer and Smith (2000), and models incorporating on-the-job search, such as Postel-Vinay and Robin (2002), Cahuc, Postel-Vinay, and Robin (2006), and Lentz (2010).

papers, however, features wage dispersion across firms, because more productive firms expand on the extensive margin of matched workers until the bargained wage rate equals the replacement cost of a worker. Our main point of departure from these studies is the introduction of ex post heterogeneity in worker ability and screening of workers by firms, which generates wage dispersion across firms that is influenced by both search frictions and trade liberalization.

Our modelling of labor market frictions is also related to the one-period search models of Acemoglu (1999) and Acemoglu and Shimer (1999), in which firms make irreversible investments in capacity or technology before being matched one-to-one with workers. Davidson, Matusz, and Shevchenko (2008) examined the impact of international trade in a model of this form, where firms choose either a high or low technology and can be matched with either a high- or low-skill manager. In equilibria where high-skill managers are willing to accept jobs at low-technology firms and only high-technology firms export, the model features an exporter wage premium and trade liberalization increases the wage gap between high- and low-skill managers. One key difference between our approach and these models is that we allow for an endogenous measure of matched workers for each firm rather than assuming one-to-one matching between firms and workers. Modelling endogenous variation in firm size enables our framework to speak in a meaningful way to empirically observed correlations between productivity, employment, and wages, which is important for the within-industry reallocations induced by trade liberalization. Once this endogenous variation in firm size is introduced, differences in workforce composition play a central role in generating differences in wages across firms, as discussed above.

The remainder of the paper is structured as follows. Section 2 outlines the model and its sectoral equilibrium. Section 3 presents our results on sectoral wage inequality. Section 4 presents our results on sectoral unemployment, and Section 5 extends our analysis to incorporate observable ex ante heterogeneity between multiple types of workers. Section 6 examines alternative ways of closing the model in general equilibrium. Section 7 concludes. The online Technical Appendix (Helpman, Itskhoki, and Redding (2010)) contains technical details, including proofs of propositions and other results.

2. SECTORAL EQUILIBRIUM

The key predictions of our model relate to the distribution of wages and employment across firms and workers within a sector. In this section, we derive these distributions from comparisons across firms that hold in sectoral equilibrium for *any* value of a worker's expected income outside the sector, that is, his outside option. An important implication of this result is that the model's predictions for sectoral equilibrium hold regardless of general equilibrium effects. Throughout this section, all prices, revenues, and costs are measured in terms of a numeraire, where the choice of this numeraire is specified when we embed the sector in general equilibrium.

2.1. *Model Setup*

We consider a world of two countries, home and foreign, where foreign variables are denoted by an asterisk. In each country there is a continuum of workers who are ex ante identical. Initially, we assume workers are risk neutral, but we consider risk aversion in Section 6. Demand within the sector is defined over the consumption of a continuum of horizontally differentiated varieties and takes the constant elasticity of substitution (CES) form. The real consumption index for the sector (Q) is therefore defined as

$$Q = \left[\int_{j\in J} q(j)^{\beta}\,dj\right]^{1/\beta}, \quad 0 < \beta < 1,$$

where j indexes varieties, J is the set of varieties within the sector, $q(j)$ denotes consumption of variety j, and β controls the elasticity of substitution between varieties. To simplify notation, we suppress the sector subscript except where important, and while we display expressions for home, analogous relationships hold for foreign. The price index dual to Q is denoted by P and depends on the prices $p(j)$ of individual varieties j. Given this specification of sectoral demand, the equilibrium revenue of a firm is

$$r(j) = p(j)q(j) = Aq(j)^{\beta}, \tag{1}$$

where A is a demand shifter for the sector.[7] Each firm takes the demand shifter as given when making its decisions, because it supplies one of a continuum of varieties within the sector and is, therefore, of measure zero relative to the sector as a whole. In this section we show how A is determined in sectoral equilibrium, and we show in Section 6 how it is related to prices and expenditures in general equilibrium.

The product market is modelled in the same way as in Melitz (2003). There is a competitive fringe of potential firms that can choose to enter the differentiated sector by paying an entry cost of $f_e > 0$. Once a firm incurs the sunk entry cost, it observes its productivity θ, which is independently distributed and drawn from a Pareto distribution $G_\theta(\theta) = 1 - (\theta_{\min}/\theta)^z$ for $\theta \geq \theta_{\min} > 0$ and $z > 1$. The Pareto distribution is not only tractable, but together with our other assumptions implies a Pareto firm-size distribution, which provides a reasonable approximation to observed data (see Axtell (2001)). Since in equilibrium

[7]As is well known, the demand function for a variety j can be expressed as

$$q(j) = A^{1/(1-\beta)} p(j)^{-1/(1-\beta)}, \quad \text{where} \quad A = E^{1-\beta}P^{\beta}$$

and E is total expenditure on varieties within the sector, while P is the sector's ideal price index.

all firms with the same productivity behave symmetrically, we index firms by θ from now onward.

Once firms observe their productivity, they decide whether to exit, produce solely for the domestic market, or produce for both the domestic and export markets. Production involves a fixed cost of $f_d > 0$ units of the numeraire. Similarly, exporting involves a fixed cost of $f_x > 0$ units of the numeraire and an iceberg variable trade cost, such that $\tau > 1$ units of a variety must be exported for one unit to arrive in the foreign market.

Output of each variety (y) depends on the productivity of the firm (θ), the measure of workers hired (h), and the average ability of these workers ($\bar{a}$):

$$(2) \qquad y = \theta h^{\gamma} \bar{a}, \quad 0 < \gamma < 1.$$

This production technology can be interpreted as capturing either human capital complementarities (e.g., production in teams where the productivity of a worker depends on the average productivity of her team) or a managerial time constraint (e.g., a manager with a fixed amount of time who needs to allocate some time to each worker). In the Technical Appendix, we derive the production technology under each of these interpretations. A key feature of the production technology is complementarities in worker ability, where the productivity of a worker is increasing in the abilities of other workers employed by the firm.[8]

Worker ability is assumed to be independently distributed and drawn from a Pareto distribution, $G_a(a) = 1 - (a_{\min}/a)^k$ for $a \geq a_{\min} > 0$ and $k > 1$. Under our preferred interpretation, worker ability is match-specific and hence a worker's ability draw for a given match conveys no information about ability draws for other potential matches. The labor market is characterized by search and matching frictions which are modelled following the standard Diamond–Mortensen–Pissarides approach. A firm that pays a *search cost* of bn units of the numeraire can randomly match with a measure n of workers, where the search cost b is endogenously determined by the tightness of the labor market as discussed below.

Consistent with a large empirical literature in labor economics, we assume that worker ability cannot be costlessly observed when firms and workers are matched.[9] Instead, we assume that firms can undertake costly investments in worker screening to obtain an imprecise signal of worker ability, which is in line with a recent empirical literature on firm screening and other recruitment

[8]The existence of these production complementarities is the subject of a long line of research in economics, including Lucas (1978), Rosen (1982), and Garicano (2000). For empirical evidence, see, for example, Moretti (2004).

[9]For example, Altonji and Pierret (2001) found that as employers learn about worker productivity, the wage equation coefficients on easily observed characteristics, such as education, fall relative to the coefficients on hard-to-observe correlates of worker productivity.

policies.[10] To capture the idea of an imprecise signal in as tractable a way as possible, we assume that by paying a *screening cost* of ca_c^δ/δ units of the numeraire, where $c > 0$ and $\delta > 0$, a firm can identify workers with an ability below a_c.[11] Screening costs are increasing in the ability threshold a_c chosen by the firm, because more complex and costlier tests are required for higher ability cutoffs.[12]

This specification of worker screening is influenced by empirical evidence that more productive firms not only employ more workers, but also screen more intensively, have workforces of higher average ability, and pay higher wages. Each of these features emerges naturally from our specification of production and screening, as demonstrated below, because production complementarities imply a greater return to screening for more productive firms and the screening technology is the same for all firms. Our formulation also ensures that the multilateral bargaining game between firms and workers over the surplus from production remains tractable. As the only information revealed by screening is which workers have match-specific abilities above and below a_c, neither the firm nor the workers know the match-specific abilities of individual workers, and hence bargaining occurs under conditions of symmetric information.

The key feature of our analysis is not the precise formulation of screening, which is chosen partly for tractability, but the variation in workforce composition across firms after screening. Since more productive firms have workforces of higher average ability after screening, they pay higher wages as the outcome of the bargaining game. Other formulations in which more productive firms have workforces of higher average ability after screening would also generate the prediction that more productive firms pay higher bargained wages.

2.2. *Firm's Problem*

The complementarities between workers' abilities in the production technology provide the incentive for firms to screen workers. By screening and not employing workers with abilities less than a_c, a firm reduces output (and hence revenue and profits) by decreasing the measure of workers hired (h), but raises output by increasing average worker ability ($\bar{a}$). Since there are diminishing returns to the measure of workers hired ($0 < \gamma < 1$), output can be increased by

[10]For empirical evidence on the resources devoted by firms to the screening of job applicants, see, for example, Barron, Bishop, and Dunkelberg (1985), Barron, Black, and Loewenstein (1987), Pellizzari (2005), and Autor and Scarborough (2008).

[11]In this formulation, there is a fixed cost of screening, even when the screening is not informative, that is, when $a_c = a_{\min}$. We focus on interior equilibria in which firms of all productivities choose screening tests that are informative, $a_c > a_{\min}$, and so the fixed cost of screening is always incurred. As we show below, this is the case when the screening cost, c, is sufficiently small.

[12]All results generalize immediately to the case where the screening costs are separable in a_c and n, and linear in n, so that we can allow screening costs to rise with the measure of matched workers.

screening as long as there is sufficient dispersion in worker ability (sufficiently low k).[13] With a Pareto distribution of worker ability, a firm that chooses a screening threshold a_c hires a measure $h = n(a_{\min}/a_c)^k$ of workers with average ability $\bar{a} = ka_c/(k-1)$. Therefore, the production technology can be rewritten as

$$y = \kappa_y \theta n^\gamma a_c^{1-\gamma k}, \qquad \kappa_y \equiv \frac{k}{k-1} a_{\min}^{\gamma k}, \tag{3}$$

where we require $0 < \gamma k < 1$ for a firm to have an incentive to screen.[14]

Given consumer love of variety and a fixed production cost, no firm will ever serve the export market without also serving the domestic market. If a firm exports, it allocates its output ($y(\theta)$) between the domestic and export markets ($y_d(\theta)$ and $y_x(\theta)$, respectively) to equate its marginal revenues in the two markets, which from (1) implies $[y_x(\theta)/y_d(\theta)]^{1-\beta} = \tau^{-\beta}(A^*/A)$. Therefore, a firm's total revenue can be expressed as

$$r(\theta) \equiv r_d(\theta) + r_x(\theta) = \Upsilon(\theta)^{1-\beta} A y(\theta)^\beta, \tag{4}$$

where $r_d(\theta) \equiv A y_d(\theta)^\beta$ is revenue from domestic sales and $r_x(\theta) \equiv A^*[y_x(\theta)/\tau]^\beta$ is revenue from exporting. The variable $\Upsilon(\theta)$ captures a firm's "market access," which depends on whether it chooses to serve both the domestic and foreign markets or only the domestic market:

$$\Upsilon(\theta) \equiv 1 + I_x(\theta)\tau^{-\beta/(1-\beta)}\left(\frac{A^*}{A}\right)^{1/(1-\beta)}, \tag{5}$$

where $I_x(\theta)$ is an indicator variable that equals 1 if the firm exports and 0 otherwise.[15]

[13]Since production complementarities provide the incentive for firms to screen, the marginal product of workers with abilities below a_c is negative for a firm with screening threshold a_c, as shown in the Technical Appendix. Note that in this production technology, the marginal product of a worker depends not only on his ability, but also on the ability of his co-workers. Therefore, a worker with a given ability can have a positive or negative marginal product, depending on the ability of his co-workers. While worker screening is a key feature of firms' recruitment policies, and production complementarities provide a tractable explanation for it, other explanations are also possible, such as fixed costs of maintaining an employment relationship (e.g., in terms of office space or other scarce resources).

[14]In contrast, when $\gamma > 1/k$, no firm screens and the model reduces to the model of Helpman and Itskhoki (2010), which has no screening or ex post worker heterogeneity. We do not discuss this case here. While for simplicity, we assume a unit exponent on average ability in the production technology (2), a more general specification is $y = \theta h^\gamma \bar{a}^\xi$, in which case the condition for firms to screen is $0 < \gamma < \xi/k$.

[15]Note that $[y_x(\theta)/y_d(\theta)]^{1-\beta} = \tau^{-\beta}(A^*/A)$ and $y_d(\theta) + y_x(\theta) = y(\theta)$ imply $y_d(\theta) = y(\theta)/\Upsilon(\theta)$ and $y_x(\theta) = y(\theta)[\Upsilon(\theta) - 1]/\Upsilon(\theta)$, and hence $r_d(\theta) = r(\theta)/\Upsilon(\theta)$ and $r_x(\theta) = r(\theta)[\Upsilon(\theta) - 1]/\Upsilon(\theta)$.

After having observed its productivity, a firm chooses whether or not to produce, whether or not to export, the measure of workers to sample, and the screening ability threshold (and hence the measure of workers to hire). Once these decisions have been made, the firm and its hired workers engage in strategic bargaining with equal weights over the division of revenue from production in the manner proposed by Stole and Zwiebel (1996a, 1996b). The only information known by firms and workers at the bargaining stage is that each hired worker has an ability greater than a_c. Therefore, the expected ability of each worker is $\bar{a} = k/(k-1)a_c$ and each worker is treated as if they have an ability of $\bar{a}$. Combining (2) and (4), firm revenue can be written as $r = Y(\theta)^{1-\beta}A(\theta\bar{a})^{\beta}h^{\beta\gamma}$, which is continuous, increasing, and concave in h. As the fixed production, and fixed exporting, search, and screening costs have all been sunk before the bargaining stage, all other arguments of firm revenue are fixed. Furthermore, the outside option of hired workers is unemployment, whose value we normalize to 0. As a result, the solution to the bargaining game is that the firm receives the fraction $1/(1+\beta\gamma)$ of revenue (4), while each worker receives the fraction $\beta\gamma/(1+\beta\gamma)$ of average revenue per worker.[16]

Anticipating the outcome of the bargaining game, the firm maximizes its profits. Combining (3), (4), and (5), this profit maximization problem can be written as

$$\pi(\theta) \equiv \max_{\substack{n\geq 0,\\ a_c \geq a_{\min},\\ I_x\in\{0,1\}}} \left\{ \frac{1}{1+\beta\gamma}\left[1 + I_x \tau^{-\beta/(1-\beta)}\left(\frac{A^*}{A}\right)^{1/(1-\beta)}\right]^{1-\beta} \times A(\kappa_y \theta n^{\gamma} a_c^{1-\gamma k})^{\beta} - bn - \frac{c}{\delta}a_c^{\delta} - f_d - I_x f_x \right\}. \tag{6}$$

The firm's decision whether or not to produce and whether or not to export takes a standard form. The presence of a fixed production cost implies that there is a zero-profit cutoff for productivity, θ_d, such that a firm drawing a productivity below θ_d exits without producing. Similarly, the presence of a fixed exporting cost implies that there is an exporting cutoff for productivity, θ_x, such that a firm drawing a productivity below θ_x does not find it profitable to serve the export market. Given that a large empirical literature finds evidence of

[16]See Acemoglu, Antràs, and Helpman (2007) and the Technical Appendix for the derivation of the solution to the bargaining game. Stole–Zwiebel bargaining is a natural generalization of Nash bargaining to the multiple workers case: the firm bargains bilaterally with every worker, but unlike in Nash bargaining, it internalizes the effect of a worker's departure on the wages of the remaining workers. As a result, the equilibrium wage as a function of employment is the solution to the differential equation $\partial(r - wh)/\partial h = w$, which equalizes the marginal surplus of the firm and the surplus of the worker from employment.

selection into export markets, where only the most productive firms export, we focus on values of trade costs for which $\theta_x > \theta_d > \theta_{\min}$.[17] The firm market access variable is, therefore, determined as

$$\Upsilon(\theta) = \begin{cases} 1, & \theta < \theta_x, \\ \Upsilon_x, & \theta \geq \theta_x, \end{cases} \tag{7}$$

$$\Upsilon_x \equiv 1 + \tau^{-\beta/(1-\beta)} \left(\frac{A^*}{A} \right)^{1/(1-\beta)} > 1.$$

The firm's first-order conditions for the measure of workers sampled (n) and the screening ability threshold (a_c) are

$$\frac{\beta\gamma}{1+\beta\gamma} r(\theta) = bn(\theta),$$

$$\frac{\beta(1-\gamma k)}{1+\beta\gamma} r(\theta) = ca_c(\theta)^\delta.$$

These conditions imply that firms with larger revenue sample more workers and screen to a higher ability threshold. While the measure of workers hired, $h = n(a_{\min}/a_c)^k$, is increasing in the measure of workers sampled, n, it is decreasing in the screening ability threshold, a_c. Under the assumption $\delta > k$, firms with larger revenue not only sample more workers, but also hire more workers. Finally, from the division of revenue in the bargaining game, the total wage bill is a constant share of revenue, which implies that firm wages are monotonically increasing in the screening ability cutoff:

$$w(\theta) = \frac{\beta\gamma}{1+\beta\gamma} \frac{r(\theta)}{h(\theta)} = b \frac{n(\theta)}{h(\theta)} = b \left[\frac{a_c(\theta)}{a_{\min}} \right]^k.$$

By adjusting employment, firms are able to push their bargained wage down to the replacement cost of a worker. As larger firms have workers of higher average ability, which are more costly to replace, they pay higher wages. Thus, firms with larger revenue have higher screening ability cutoffs and pay higher wages, but the expected wage conditional on being sampled is the same across all firms,

$$\frac{w(\theta)h(\theta)}{n(\theta)} = b,$$

[17] For empirical evidence of selection into export markets, see, for example, Bernard and Jensen (1995) and Roberts and Tybout (1997).

which implies that workers have no incentive to direct their search.[18] Combining the measure of workers hired, $h = n(a_{\min}/a_c)^k$, with the first-order conditions above yields the following relationship between firm wages and the measure of workers hired:

$$\ln w(\theta) = \text{constant} + \frac{k}{\delta - k} \ln h(\theta).$$

Therefore, under the assumption $\delta > k$, the model exhibits an employer-size wage premium, where firms that employ more workers pay higher wages. To match empirical findings of such an employer-size wage premium, we focus on parameter values satisfying this inequality.

Using the firms' first-order conditions, firm revenue (4), and the production technology (3), we can solve explicitly for firm revenue as a function of firm productivity (θ), the demand shifter (A), the search cost (b), and parameters:

$$r(\theta) = \kappa_r \big[c^{-\beta(1-\gamma k)/\delta} b^{-\beta\gamma} \Upsilon(\theta)^{(1-\beta)} A \theta^{\beta}\big]^{1/\Gamma}, \tag{8}$$

where $\Gamma \equiv 1 - \beta\gamma - \beta(1-\gamma k)/\delta > 0$ and the constant κ_r is defined in the Technical Appendix. An implication of this expression is that the relative revenues of any two firms depend solely on their relative productivities and relative market access: $r(\theta')/r(\theta'') = (\theta'/\theta'')^{\beta/\Gamma} (\Upsilon(\theta')/\Upsilon(\theta''))^{\beta/\Gamma}$.

Finally, using the two first-order conditions in the firm's problem (6), firm profits can be expressed in terms of firm revenue and the fixed costs of production and exporting:

$$\pi(\theta) = \frac{\Gamma}{1+\beta\gamma} r(\theta) - f_d - I_x(\theta) f_x. \tag{9}$$

2.3. *Sectoral Variables*

To determine sectoral equilibrium, we use the recursive structure of the model. In a first block of equations, we solve for the tightness of the labor market (x, x^*) and search costs (b, b^*) in each country. In a second block of equations, we solve for the zero-profit productivity cutoffs (θ_d, θ_d^*), the exporting productivity cutoffs (θ_x, θ_x^*), and sectoral demand shifters (A, A^*). A third and final block of equations, to be described in Section 6, determines the remaining components of the sector's variables: the dual price index (P, P^*), the real consumption index (Q, Q^*), the mass of firms (M, M^*), and the size of the labor force (L, L^*).

[18]We note that search frictions and wage bargaining alone are not enough to generate wage variation across firms in our model. From the firm's first-order condition for the measure of workers sampled, each firm equates workers' share of revenue per sampled worker to the common search cost. In the special case of our model without worker heterogeneity and screening, all sampled workers are hired, which implies that each firm's wage is equal to the common search cost.

2.3.1. *Labor Market Tightness and Search Costs*

Following the standard Diamond–Mortensen–Pissarides approach, the search cost (b) is assumed to be increasing in labor market tightness (x):

$$b = \alpha_0 x^{\alpha_1}, \quad \alpha_0 > 1, \alpha_1 > 0, \tag{10}$$

where labor market tightness equals the ratio of workers sampled (N) to workers searching for employment in the sector (L): $x = N/L$.[19] Under the assumption of risk neutrality, the supply of workers searching for employment in the sector depends on their expected income outside the sector, that is, their outside option, ω. In particular, workers are indifferent between searching for employment inside and outside the sector if their expected income in the sector, which equals the probability of being sampled (x) times the expected wage conditional on being sampled ($w(\theta)h(\theta)/n(\theta) = b$ from the analysis above), is equal to ω:

$$\omega = xb. \tag{11}$$

We discuss in Section 6 how this condition is modified when workers are risk averse. Together (10) and (11) determine the search cost and the labor market tightness (b, x) for a given value of expected income (ω):

$$b = \alpha_0^{1/(1+\alpha_1)} \omega^{\alpha_1/(1+\alpha_1)} \quad \text{and} \quad x = \left(\frac{\omega}{\alpha_0}\right)^{1/(1+\alpha_1)}, \tag{12}$$

where we assume $\alpha_0 > \omega$ so that $0 < x < 1$, as discussed in Section 6. Analogous relationships determine search costs and labor market tightness (b^*, x^*) for a given value of expected income (ω^*) in foreign. The search cost in (12) depends solely on parameters of the search technology (α_0, α_1) and expected income (ω). In particular, we can make the following statement.

LEMMA 1: *The search cost b and the measure of labor market tightness x are both increasing in expected worker income ω.*

When we subsequently embed the sector in general equilibrium, we specify conditions under which expected worker income (ω) is constant and conditions under which it changes with the other endogenous variables of the model.

[19]As shown by Blanchard and Galí (2010) and in the Technical Appendix, this relationship can be derived from a constant returns to scale Cobb–Douglas matching function and a cost of posting vacancies. The parameter α_0 is increasing in the cost of posting vacancies and decreasing in the productivity of the matching technology, while α_1 depends on the weight of vacancies in the Cobb–Douglas matching function. Other static models of search and matching include Acemoglu (1999) and Acemoglu and Shimer (1999).

2.3.2. *Productivity Cutoffs and Demand*

The two productivity cutoffs can be determined using firm revenue (8) and profits (9). The productivity cutoff below which firms exit (θ_d) is determined by the requirement that a firm with this productivity makes zero profits:

$$\frac{\Gamma}{1+\beta\gamma}\kappa_r\big[c^{-\beta(1-\gamma k)/\delta}b^{-\beta\gamma}A\theta_d^{\beta}\big]^{1/\Gamma}=f_d. \tag{13}$$

Similarly, the exporting productivity cutoff above which firms export (θ_x) is determined by the requirement that at this productivity a firm is indifferent between serving only the domestic market and serving both the domestic and foreign markets:

$$\frac{\Gamma}{1+\beta\gamma}\kappa_r\big[c^{-\beta(1-\gamma k)/\delta}b^{-\beta\gamma}A\theta_x^{\beta}\big]^{1/\Gamma}\big[\Upsilon_x^{(1-\beta)/\Gamma}-1\big]=f_x. \tag{14}$$

These two conditions imply the following relationship between the productivity cutoffs:

$$\big[\Upsilon_x^{(1-\beta)/\Gamma}-1\big]\left(\frac{\theta_x}{\theta_d}\right)^{\beta/\Gamma}=\frac{f_x}{f_d}. \tag{15}$$

In equilibrium, we also require the free entry condition to hold, which equates the expected value of entry to the sunk entry cost. Using the zero-profit and exporting cutoff conditions, (13) and (14), respectively, and the relationship between revenues for firms with different productivities from (8), the free entry condition can be written as

$$f_d\int_{\theta_d}^{\infty}\left[\left(\frac{\theta}{\theta_d}\right)^{\beta/\Gamma}-1\right]dG_\theta(\theta)+f_x\int_{\theta_x}^{\infty}\left[\left(\frac{\theta}{\theta_x}\right)^{\beta/\Gamma}-1\right]dG_\theta(\theta)=f_e. \tag{16}$$

Equations (13), (14), and (16) can be used to solve for home's productivity cutoffs and the demand shifter (θ_d, θ_x, A) for a given value of the foreign demand shifter (A^*), which only influences home sectoral equilibrium through exporter market access ($\Upsilon_x > 1$).[20] Three analogous equations can be used to solve for foreign variables (θ_d^*, θ_x^*, A^*) for a given value of A. Together these six equations allow us to solve for the productivity cutoffs and demand shifters in the two countries (θ_d, θ_x, A, θ_d^*, θ_x^*, A^*) for given values of search costs (b, b^*), which were determined in the previous block of equations. Having solved for the productivity cutoffs and demand shifters, firm market access in each country ($\Upsilon(\theta)$, $\Upsilon^*(\theta)$) follows immediately from (5).

[20]In a symmetric equilibrium $A = A^*$ and $\Upsilon_x = 1+\tau^{-\beta/(1-\beta)}$, which implies that the ratio of the two productivity cutoffs is pinned down by (15) alone.

The productivity cutoffs and demand shifter depend on two dimensions of trade openness in (13), (14), and (16). First, they depend on an extensive margin of trade openness, as captured by the ratio of the productivity cutoffs $\rho \equiv \theta_d/\theta_x \in [0,1]$, which determines the fraction of exporting firms $[1-G_\theta(\theta_x)]/[1-G_\theta(\theta_d)] = \rho^z$. Second, they depend on an intensive margin of trade openness, as captured by the market access variable, $\Upsilon_x > 1$, which determines the ratio of revenues from domestic sales and exporting, as discussed in footnote 15. These two dimensions of trade openness are linked through the relationship between the productivity cutoffs (15).

2.4. *Firm-Specific Variables*

In this section, we use the solutions for sectoral equilibrium to solve for firm-specific variables. We show that the model's predictions for firm-specific variables are consistent with empirically observed relationships between wages, employment, workforce composition, productivity, and export participation. To solve for firm-specific variables, we use two properties of the model. First, from firm revenue (8), the relative revenue of any two firms depends solely on their relative productivities and relative market access. Second, from firm profits (9), the lowest productivity firm with productivity θ_d makes zero profits. Combining these two properties with the firm's first-order conditions above allows all firm-specific variables to be written as functions of firm productivity (θ), firm market access ($\Upsilon(\theta)$), the zero-profit productivity cutoff (θ_d), search costs (b), and parameters,

$$
\begin{aligned}
(17)\qquad & r(\theta) = \Upsilon(\theta)^{(1-\beta)/\Gamma}\cdot r_d\cdot\left(\frac{\theta}{\theta_d}\right)^{\beta/\Gamma}, \quad r_d \equiv \frac{1+\beta\gamma}{\Gamma} f_d,\\
& h(\theta) = \Upsilon(\theta)^{(1-\beta)(1-k/\delta)/\Gamma}\cdot h_d\cdot\left(\frac{\theta}{\theta_d}\right)^{\beta(1-k/\delta)/\Gamma},\\
& h_d \equiv \frac{\beta\gamma}{\Gamma}\frac{f_d}{b}\left[\frac{\beta(1-\gamma k)}{\Gamma}\frac{f_d}{c a_{\min}^{\delta}}\right]^{-k/\delta},\\
& w(\theta) = \Upsilon(\theta)^{k(1-\beta)/(\delta\Gamma)}\cdot w_d\cdot\left(\frac{\theta}{\theta_d}\right)^{\beta k/(\delta\Gamma)},\\
& w_d \equiv b\left[\frac{\beta(1-\gamma k)}{\Gamma}\frac{f_d}{c a_{\min}^{\delta}}\right]^{k/\delta},
\end{aligned}
$$

where market access ($\Upsilon(\theta)$) is determined as a function of firm productivity in (7). Combining these expressions with the Pareto productivity distribution, firm revenue and employment are also Pareto distributed, with shape parameters that depend on the dispersion of firm productivity, the dispersion of

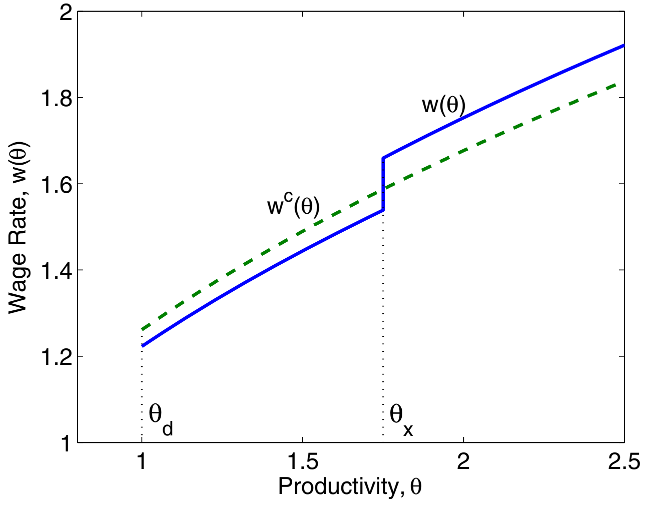

FIGURE 1.—Wages as a function of firm productivity.

worker ability, and product and labor market parameters that influence workforce composition.[21]

From the firm-specific solutions (17), more productive firms not only have higher revenue, profits, and employment, as in the benchmark model of firm heterogeneity of Melitz (2003), but also pay higher wages as shown in Figure 1. These results are consistent with empirical evidence on rent sharing, whereby higher firm revenue and profits are shared with workers through higher wages (e.g., Van Reenen (1996)), and with the large empirical literature that finds an employer-size wage premium (see the survey by Oi and Idson (1999)).

Additionally, the differences in wages across firms are driven by differences in workforce composition. More productive firms have workforces of higher average ability, which are more costly to replace in the bargaining game, and, therefore, they pay higher wages. These features of the model are consistent with empirical findings that the employer-size wage premium is in part explained by differences in the unobserved characteristics of workers across firms.[22] The reason more productive firms have workforces of higher average ability in the model is that they screen more intensively, which also receives empirical support. An emerging literature on firm recruitment policies provides

[21]See Helpman, Itskhoki, and Redding (2008a) for further discussion. As the shape parameters of these distributions depend on ratios of parameters in the model, the observed distributions can be used to discipline the model's parameters.

[22]See, for example, Abowd, Kramarz, and Margolis (1999), Abowd, Creecy, and Kramarz (2002), and De Melo (2008).

evidence of more intensive screening policies for larger firms and higher-wage matches.[23]

Finally, firm characteristics are systematically related to export participation in the model. As a result of fixed costs of exporting, there is a discrete increase in firm revenue at the productivity threshold for entry into exporting (θ_x), where $\Upsilon(\theta)$ increases from 1 to $\Upsilon_x > 1$, which implies a discrete increase in all other firm variables except for profits. Therefore, exporters not only have higher revenue and employment than nonexporters, as in the benchmark model of firm heterogeneity of Melitz (2003), but also pay higher wages, as found empirically by Bernard and Jensen (1995, 1997). While exporting increases the wage paid by a firm with a given productivity, so that the model features an exporter wage premium conditional on firm productivity, it does not feature an exporter wage premium conditional on firm size, because both firm wages and firm size increase discretely at the productivity threshold for entry into export markets.[24] The wage differences between exporters and nonexporters in the model are accompanied by differences in workforce composition, as found empirically by Schank, Schanbel, and Wagner (2007), Munch and Skaksen (2008), and Frías, Kaplan, and Verhoogen (2009).

3. SECTORAL WAGE INEQUALITY

While workers are ex ante identical and have the same expected income, there is ex post wage inequality because workers receive different wages depending on the employer with whom they are matched. In this section, we consider the within-sector distribution of wages across employed workers. This sectoral wage distribution is a weighted average of the distributions of wages for workers employed by domestic firms, $G_{w,d}(w)$, and for workers employed by exporters, $G_{w,x}(w)$, with weights equal to the shares of employment in the two groups of firms:

$$G_w(w) = \begin{cases} S_{h,d} G_{w,d}(w), \\ \qquad \text{for } w_d \leq w \leq w_d/\rho^{\beta k/(\delta \Gamma)}, \\ S_{h,d}, \quad \text{for } w_d/\rho^{\beta k/(\delta \Gamma)} \leq w \leq w_d \Upsilon_x^{k(1-\beta)/(\delta \Gamma)}/\rho^{\beta k/(\delta \Gamma)}, \\ S_{h,d} + (1 - S_{h,d}) G_{w,x}(w), \\ \qquad \text{for } w \geq w_d \Upsilon_x^{k(1-\beta)/(\delta \Gamma)}/\rho^{\beta k/(\delta \Gamma)}, \end{cases} \tag{18}$$

[23]For example, Barron, Black, and Loewenstein (1987) found that expenditures on screening workers are positively and significantly related to employer size, while Pellizzari (2005) found that matches created through more intensive screening pay higher wages.

[24]One potential explanation for an exporter wage premium conditional on firm size emerges from the extension of our model (discussed below) to incorporate multiple types of workers with different observed characteristics. Differences in workforce composition across these types between exporters and nonexporters that are imperfectly controlled for in empirical studies can give rise to such an exporter wage premium conditional on firm size.

where ρ and Y_x are the extensive and intensive margins of trade openness defined above, $w_d = w(\theta_d)$ is the wage paid by the least productive firm in (17), $w_d/\rho^{k\beta/(\delta\Gamma)} = w(\theta_x^-)$ is the wage paid by the most productive nonexporter, and $w_d Y_x^{k(1-\beta)/(\delta\Gamma)}/\rho^{k\beta/(\delta\Gamma)} = w(\theta_x^+)$ is the wage paid by the least productive exporter. Note that w_d depends solely on parameters and search costs (b), which in turn depend on expected worker income (ω). The share of workers employed by domestic firms, $S_{h,d}$, can be evaluated using the Pareto productivity distribution and the solution for firm-specific variables (17) as

$$S_{h,d} = \frac{1-\rho^{z-\beta(1-k/\delta)/\Gamma}}{1+\rho^{z-\beta(1-k/\delta)/\Gamma}[Y_x^{(1-\beta)(1-k/\delta)/\Gamma}-1]},$$

which depends on the extensive and intensive margins of trade openness.

The distributions of wages across workers employed by domestic and exporting firms can also be derived from the solutions for firm-specific variables (17). Given that productivity is Pareto distributed, and both wages and employment are power functions of productivity, the distribution of wages across workers employed by domestic firms is a truncated Pareto distribution

$$G_{w,d}(w) = \frac{1-\left(\dfrac{w_d}{w}\right)^{1+1/\mu}}{1-\rho^{z-\beta(1-k/\delta)/\Gamma}} \quad \text{for} \quad w_d \leq w \leq w_d/\rho^{k\beta/(\delta\Gamma)}.$$

Similarly, the distribution of wages across workers employed by exporters, $G_{w,x}(w)$, is an untruncated Pareto distribution

$$G_{w,x}(w) = 1-\left[\frac{w_d}{w} Y_x^{k(1-\beta)/(\delta\Gamma)} \rho^{-k\beta/(\delta\Gamma)}\right]^{1+1/\mu}$$
$$\text{for} \quad w \geq w_d Y_x^{k(1-\beta)/(\delta\Gamma)}/\rho^{k\beta/(\delta\Gamma)}.$$

The wage distributions for workers employed by domestic firms and by exporters have the same shape parameter, $1+1/\mu$, where μ is defined as

$$\mu \equiv \frac{\beta k/\delta}{z\Gamma-\beta}, \quad \text{where} \quad \Gamma \equiv 1-\beta\gamma-\frac{\beta}{\delta}(1-\gamma k).$$

For the mean and variance of the sectoral wage distribution to be finite, we require $0 < \mu < 1$ and hence $z\Gamma > 2\beta$, which is satisfied for sufficiently large z (a not too dispersed productivity distribution).[25] The dispersion of firm wages

[25]While we concentrate on the wage distribution, as this is typically the subject of the economic debate over the impact of trade liberalization, the income distribution could also be influenced by profits. The model can also be used to determine the distribution of revenue (and hence profits) across firms as discussed above.

is systematically related to the dispersion of firm revenue in the model, because larger firms have workforces of higher average ability and hence pay higher wages. While this mechanism is more general than our distributional assumptions, the assumption of Pareto distributions of firm productivity and worker ability enables closed-form solutions for the wage distribution to be derived. While the log normal distribution is generally believed to provide a closer approximation to the empirical wage distribution, the Pareto distribution provides a close approximation for the upper tail.

3.1. *Sectoral Wage Inequality in the Closed Economy*

The closed economy wage distribution can be obtained by considering the case of arbitrarily large values of trade costs, which imply $\rho \to 0$ in (18). In the closed economy, the share of employment in domestic firms is equal to 1, and the sectoral wage distribution across workers employed by domestic firms is an untruncated Pareto distribution with lower limit w_d and shape parameter $1 + 1/\mu$. Given an untruncated Pareto distribution, all scale-invariant measures of inequality, such as the coefficient of variation, the Gini coefficient, and the Theil index, depend solely on the distribution's shape parameter. None of these measures depends on the lower limit of the support of the wage distribution (w_d), and they therefore do not depend on search costs (b) and expected worker income (ω). While these variables affect the mean of the wage distribution, they do not affect its dispersion. An important implication of this result is that the model's predictions for wage inequality do not depend on the equilibrium value of expected worker income (ω) and hence are robust to alternative ways of closing the model in general equilibrium.

PROPOSITION 1: *In the closed economy*, μ *is a sufficient statistic for sectoral wage inequality*. *In particular*, (i) *the coefficient of variation of wages is* $\mu/\sqrt{1-\mu^2}$; (ii) *the Lorenz curve is represented by* $s_w = 1 - (1 - s_h)^{1/(1+\mu)}$, *where* s_h *is the fraction of workers and* s_w *is the fraction of their wages when workers are ordered from low- to high-wage earners*; (iii) *the Gini coefficient is* $\mu/(2+\mu)$; *and* (iv) *the Theil index is* $\mu - \ln(1+\mu)$.

The existence of a sufficient statistic for wage inequality in the closed economy is more general than our model in the sense that it holds for a wider class of models in which firm wages and employment are power functions of productivity, and productivity is Pareto distributed. Together these features imply that the wage distribution in the closed economy is an untruncated Pareto distribution and, hence, the shape parameter of this distribution is a sufficient statistic for wage inequality. In our model, this shape parameter is linked to the underlying structural parameters of the model that influence workforce composition and, hence, enter the derived parameter μ. Evidently, sectoral wage inequality is monotonically increasing in μ (the lower the shape parameter of the wage distribution $1 + 1/\mu$, the greater the wage inequality).

PROPOSITION 2: *In the closed economy, inequality in the sectoral distribution of wages is increasing in firm productivity dispersion (lower z) and increasing in worker ability dispersion (lower k) if and only if $z^{-1} + \delta^{-1} + \gamma > \beta^{-1}$.*

Since more productive firms pay higher wages, greater dispersion in firm productivity (lower z) implies greater sectoral wage inequality. In contrast, greater dispersion in worker ability (lower k) has an ambiguous effect on sectoral wage inequality because of two counteracting forces. On the one hand, a reduction in k increases relative employment in more productive firms (from (17)) that pay higher wages, which increases wage inequality. On the other hand, a reduction in k decreases relative wages paid by more productive firms (from (17)), which reduces wage inequality. When the parameter inequality in the proposition is satisfied, the change in relative employment dominates the change in relative wages, and greater dispersion in worker ability implies greater sectoral wage inequality.

The model's prediction that sectoral wage inequality is closely linked to the dispersion of firm productivity receives empirical support. In particular, Davis and Haltiwanger (1991) showed that wage dispersion across plants within sectors accounts for a large share of overall wage dispersion and is responsible for more than one-third of the growth in overall wage dispersion in U.S. manufacturing between 1975 and 1986. Additionally, they found that between-plant wage dispersion is strongly related to between-plant size dispersion, which in our model is driven by productivity dispersion. Similarly, Faggio, Salvanes, and Van Reenen (2007) showed that a substantial component of the increase in individual wage inequality in the United Kingdom in recent decades has occurred between firms within sectors and is linked to increased productivity dispersion between firms within sectors.

While greater firm productivity dispersion (associated, for example, with innovations such as information and communication technologies (ICTs)) is one potential source of increased wage inequality in the model, another potential source is international trade as considered in the next section. Indeed, both greater firm productivity dispersion and international trade raise wage inequality through the same mechanism of greater dispersion in firm revenue and wages within industries, and both raise measured productivity at the industry level through reallocations of resources across firms.

3.2. *Open versus Closed Economy*

The sectoral wage distribution in the open economy depends on the sufficient statistic for wage inequality in the closed economy (μ) and the extensive and intensive measures of trade openness (ρ and Υ_x, respectively). In the two limiting cases where trade costs are sufficiently high that no firm exports ($\rho = 0$) and trade costs are sufficiently low that all firms export ($\rho = 1$), the open economy wage distribution is an untruncated Pareto distribution with shape parameter $1 + 1/\mu$. From Proposition 1, all scale-invariant measures of inequality for an untruncated Pareto distribution depend solely on the distribution's shape parameter. Therefore, the same level of wage inequality exists in the open economy when all firms export as in the closed economy.

To characterize sectoral wage inequality in the open economy when $0 < \rho < 1$ (only some firms export), we compare the actual open economy wage distribution ($G_w(w)$) to a counterfactual wage distribution ($G_w^c(w)$). For the counterfactual wage distribution, we choose an untruncated Pareto distribution with the same shape parameter as the wage distribution in the closed economy ($1 + 1/\mu$), but the same mean as the wage distribution in the open economy. An important feature of this counterfactual wage distribution is that it has the same level of inequality as the closed economy wage distribution. Therefore, if we show that there is more inequality with the open economy wage distribution than with the counterfactual wage distribution, this will imply that there is more wage inequality in the open economy than in the closed economy.

The counterfactual wage distribution has two other important properties, as shown formally in the Technical Appendix. First, the lowest wage in the counterfactual wage distribution (w_d^c) lies strictly in between the lowest wage paid by domestic firms (w_d) and the lowest wage paid by exporters ($w(\theta_x^+)$) in the actual open economy wage distribution. Otherwise, the counterfactual wage distribution would have a mean either lower or higher than the actual open economy wage distribution, which contradicts the requirement that the two distributions have the same mean. Second, the counterfactual wage distribution has a smaller slope than the actual wage distribution at $w(\theta_x^+)$. Otherwise, the counterfactual wage distribution would have a greater density than the actual wage distribution for $w \geq w(\theta_x^+)$, and would, therefore, have a higher mean than the actual wage distribution.

Together, these two properties imply that the relative location of the cumulative distribution functions for actual and counterfactual wages is as shown in Figure 2.[26] The actual and counterfactual cumulative distributions intersect only once, and the actual distribution lies above the counterfactual distribution for low wages and below it for high wages.[27] This pattern provides a sufficient condition for the counterfactual wage distribution to second-order stochastically dominate the wage distribution in the open economy. Therefore, for all measures of inequality that respect second-order stochastic dominance, the open economy wage distribution exhibits greater inequality than the counterfactual wage distribution. It follows that the wage distribution in the open economy exhibits more inequality than the wage distribution in the

[26]To generate Figures 1–3, we set the parameters of the model to match some of the salient features of the data. For details see Helpman, Itskhoki, and Redding (2008b).

[27]Note that the actual and counterfactual distributions can intersect either above the wage at the most productive nonexporter, $w(\theta_x^-)$ (as shown in Figure 2), or below it. In both cases, the actual and counterfactual distributions have the properties discussed in the text.

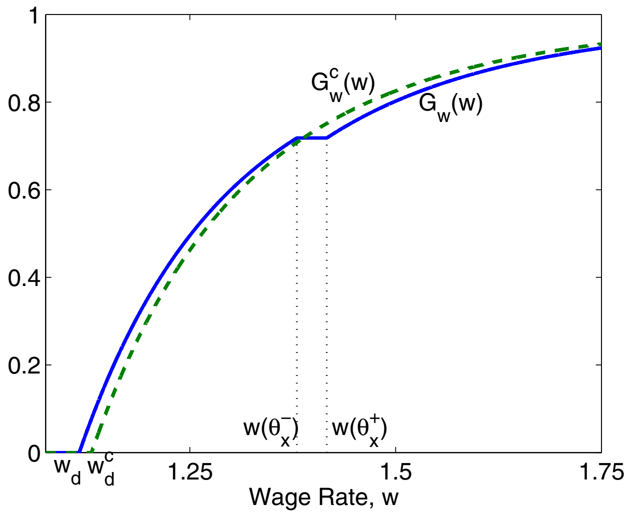

FIGURE 2.—Cumulative distribution function of wages.

closed economy. This result holds independently of whether the opening of trade affects expected worker income (ω), because ω affects the lower limit of the actual open economy wage distribution (and hence the lower limit of the counterfactual wage distribution), but does not affect the comparison of levels of inequality between the two distributions.

PROPOSITION 3: (i) *Sectoral wage inequality in the open economy when some but not all firms export is strictly greater than in the closed economy and* (ii) *sectoral wage inequality in the open economy when all firms export is the same as in the closed economy*.

Proposition 3 highlights a new mechanism for international trade to affect wage inequality that is absent from neoclassical trade theories such as the Heckscher–Ohlin model; namely, the participation of some but not all firms in exporting. This generic mechanism applies in any heterogeneous firm model in which firm wages are related to firm revenue and there is selection into export markets. As a result of this mechanism, Proposition 3 holds in models in which the following three conditions are satisfied: (i) firm wages and employment are power functions of firm productivity, (ii) there is firm selection into export markets and exporting increases wages for a firm with a given productivity, and (iii) firm productivity is Pareto distributed.

An important implication of Proposition 3, which applies for symmetric and asymmetric countries alike, is that the opening of trade can increase wage inequality in all countries. In contrast, the Stolper–Samuelson theorem of the Heckscher–Ohlin model predicts rising wage inequality in developed countries but falling wage inequality in developing countries. Proposition 3 is, therefore, consistent with empirical findings of increased wage inequality in developing countries following trade liberalization, as reviewed by Goldberg and Pavcnik (2007). Similarly, Proposition 3 is consistent with empirical evidence that much of the observed reallocation in the aftermath of trade liberalization occurs across firms within sectors and is accompanied by increases in within-group wage inequality.[28]

Since sectoral wage inequality when all firms export is the same as in the closed economy, but sectoral wage inequality when only some firms export is higher than in the closed economy, the relationship between sectoral wage inequality and the fraction of exporters is at first increasing and later decreasing.

COROLLARY TO PROPOSITION 3: *An increase in the fraction of exporting firms raises sectoral wage inequality when the fraction of exporting firms is sufficiently small and reduces sectoral wage inequality when the fraction of exporting firms is sufficiently large*.

The intuition for this result is that the increase in firm wages that occurs at the productivity threshold above which firms export is present only when some, but not all, firms export. When no firm exports ($\rho = 0$), a small reduction in trade costs that induces some firms to start exporting raises sectoral wage inequality because of the higher wages paid by exporters. When all firms export ($\rho = 1$), a small increase in trade costs that induces some firms to stop exporting raises sectoral wage inequality because of the lower wages paid by domestic firms. Furthermore, as the proof follows from that for Proposition 3 above, this corollary holds for any measure of wage inequality that respects second-order stochastic dominance and for asymmetric countries. One important implication of these results is that the initial level of trade openness is a relevant control for empirical studies examining the relationship between wage inequality and trade.

Our closed-form expression for the wage distribution (18) in terms of the extensive and intensive margins of trade openness (ρ and Υ_x, respectively) holds both in the extreme cases of autarky and frictionless trade, as well as in a trade equilibrium where only a fraction of firms export. While both fixed and variable trade costs influence sectoral wage inequality, they do so through slightly different mechanisms, because they have different effects on ρ and Υ_x. This can be seen most clearly for symmetric countries, where the intensive margin depends on variable trade costs alone ($\Upsilon_x = 1 + \tau^{-\beta/(1-\beta)}$), and changes in the fixed costs of exporting affect only the extensive margin (ρ). To illustrate the relationship between sectoral wage inequality and trade openness in the

[28]To account for reallocation within industries, however, the Stolper–Samuelson theorem can be reinterpreted as applying at a more disaggregated level within industries as, for example, in Feenstra and Hanson (1996).

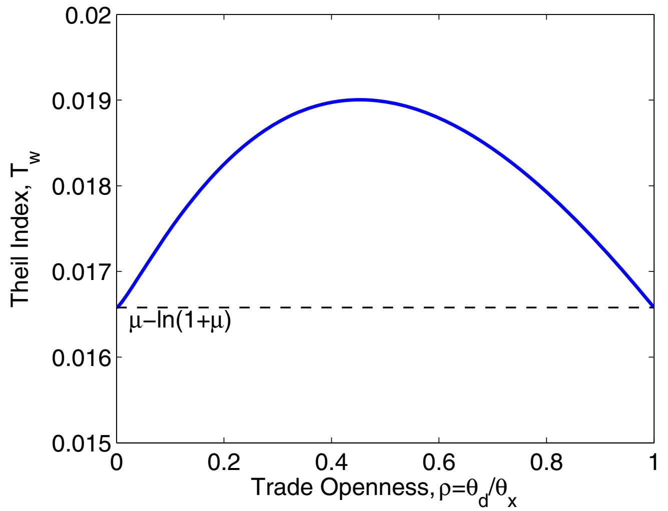

FIGURE 3.—Trade openness and sectoral wage inequality.

model, Figure 3 graphs the variation in the Theil index of wage inequality with symmetric countries as we vary the fixed cost of exporting (f_x) and hence the extensive margin of trade openness (ρ). A similar pattern of at first increasing and then later decreasing wage inequality can emerge as we vary variable trade costs (τ). While we have not been able to show analytically that the relationship between wage inequality and the fraction of exporting firms is single peaked, as shown in Figure 3, this pattern emerged from all of our simulations of the model for a wide range of different parameter configurations.

In our framework, the relationship between firm wages and revenue arises from differences in average worker ability across firms. As the opening of trade changes the dispersion of firm revenue, this in turn changes the dispersion of average worker ability across firms and, hence, changes the dispersion of wages.

PROPOSITION 4: *The opening of the closed economy to trade amplifies differences in workforce composition across firms*.

As a result of the opening of trade, the revenue of exporters increases, which induces them to screen more intensively, while the revenue of domestic firms decreases, which induces them to screen less intensively. Hence a worker with a given ability who would be hired by a high productivity firm in the closed economy may not be hired by this firm in the open economy if it becomes an exporter. The opening of trade, therefore, strengthens the correlation between firm productivity and average worker ability, which echoes the empirical findings of greater wage and skill upgrading at more productive exporting firms in Verhoogen (2008). To the extent that empirical measures of productivity do not adequately control for worker ability, changes in average worker ability are reflected in changes in measured firm productivity. As the opening of trade amplifies differences in workforce composition across firms, it, therefore, also magnifies differences in measured firm productivity.

4. SECTORAL UNEMPLOYMENT

In the model workers can be unemployed either because they are not matched with a firm or because their match-specific ability draw is below the screening threshold of the firm with which they are matched. Both components of unemployment are frictional in the sense that workers cannot immediately achieve another match. The sectoral unemployment rate u includes both of these components and can be written as 1 minus the product of the hiring rate σ and the tightness of the labor market x,

$$u = \frac{L - H}{L} = 1 - \frac{H}{N}\frac{N}{L} = 1 - \sigma x, \tag{19}$$

where $\sigma \equiv H/N$, H is the measure of hired workers, N is the measure of matched workers, and L is the measure of workers seeking employment in the sector.

The sectoral tightness of the labor market (x) in (12) depends on the search friction parameter (α_0) and expected worker income (ω). Therefore, the tightness of the labor market is not directly affected by trade openness and is only indirectly affected insofar as trade openness influences ω. In contrast, the sectoral hiring rate (σ) depends directly on trade openness, which influences firm revenues and hence screening ability thresholds. Using the Pareto productivity distribution, the sectoral hiring rate can be expressed as a function of the extensive and intensive margins of trade openness (ρ and Y_x, respectively), the sufficient statistic for wage inequality (μ), and other parameters, as shown in the Technical Appendix:

$$\sigma = \varphi(\rho, Y_x)\sigma^A, \quad \sigma^A = \frac{1}{1+\mu}\left[\frac{\Gamma}{\beta(1-\gamma k)}\frac{c a_{\min}^{\delta}}{f_d}\right]^{k/\delta}, \tag{20}$$

where σ^A is the hiring rate in autarky, the term in square brackets is the hiring rate of the least productive firm (h_d/n_d), and

$$\varphi(\rho, Y_x) \equiv \frac{1 + [Y_x^{(1-\beta)(1-k/\delta)/\Gamma} - 1]\rho^{z-\beta(1-k/\delta)/\Gamma}}{1 + [Y_x^{(1-\beta)/\Gamma} - 1]\rho^{z-\beta/\Gamma}}.$$

Evidently, we have $\varphi(0, Y_x) = 1$ and $0 < \varphi(\rho, Y_x) < 1$ for $0 < \rho \leq 1$, since $Y_x > 1$ and $\delta > k$.

Search and screening costs have quite different effects on the closed economy unemployment rate. For a given expected worker income, a rise in the search friction α_0 raises search costs (b), which reduces the sectoral tightness of the labor market (x) and increases the sectoral unemployment rate. In contrast, as the screening cost (c) increases, firms screen less intensively, which increases the sectoral hiring rate (σ) and thereby reduces the sectoral unemployment rate.

The opening of the closed economy to trade affects the sectoral unemployment rate through two channels. The first channel is through expected worker income (ω) and the tightness of the labor market (x). When we embed the sector in general equilibrium, we show that expected worker income can either remain constant or rise following the opening of trade. As a result, the tightness of the labor market is either unaffected by the opening of trade (as in Helpman and Itskhoki (2010)) or rises (as in Felbermayr, Prat, and Schmerer (2008), Felbermayr, Larch, and Lechthaler (2009)).

The second channel is through the hiring rate (σ), which depends on firms' screening decisions and is distinctive to our approach. As firms' screening decisions determine both firm wages and hiring rates, this second channel introduces a two-way dependence between wage inequality and unemployment. The opening of trade results in an expansion in the revenue of exporters and a contraction in the revenue of nonexporters, which changes industry composition toward more productive firms that screen more intensively. Therefore, the opening of trade reduces the hiring rate, which increases sectoral unemployment.

PROPOSITION 5: *The opening of the closed economy to trade has an ambiguous overall effect on the sectoral unemployment rate*: (a) *The tightness of the labor market can either remain constant or rise following the opening of trade*, *which leaves unchanged or reduces the rate of unemployment*. (b) *The hiring rate is strictly lower in the open economy than in the closed economy*, *which raises the rate of unemployment*.

While the model's predictions for the impact on sectoral wage inequality of the opening of the closed economy to trade are unambiguous irrespective of how expected worker income is determined in general equilibrium, its predictions for sectoral unemployment are ambiguous and depend on general equilibrium effects. This ambiguity of the results for unemployment is consistent with the absence of a clear empirical consensus on the relationship between trade and unemployment, as discussed, for example, in Davidson and Matusz (2009).

The sectoral distribution of income depends on both the sectoral distribution of wages and the unemployment rate, where unemployed workers all receive the same income of zero. As the opening of trade raises wage inequality and has an ambiguous effect on unemployment, income and wage inequality can move in opposite directions in the model. Therefore, our framework highlights that conclusions based on wage inequality can be misleading if the ultimate concern is income inequality.

5. OBSERVABLE WORKER HETEROGENEITY

In this section, we introduce ex ante heterogeneity across workers. We consider a setting in which there are multiple occupations and occupation-specific supplies of workers, where workers from one occupation cannot perform the tasks of workers from another occupation. There are observable differences in ex ante worker characteristics across the occupations, which introduces a distinction between within-group wage inequality (among workers with the same ex ante characteristics) and between-group wage inequality (across workers with different ex ante characteristics). Although, for expositional simplicity, we confine the discussion to two occupations only, it will become clear how the main specification can be generalized to any number of occupations.

To demonstrate in a simple way the robustness of our results to the introduction of ex ante heterogeneity, we concentrate on a Cobb–Douglas production technology. We show that the opening of trade raises within-group wage inequality for each group of workers, whereas between-group wage inequality can rise or decline with trade. When between-group wage inequality declines, the rise of within-group wage inequality can dominate, so that overall wage inequality still rises. To illustrate the flexibility of our framework, we also briefly discuss at the end of this section some implications of technology–skill complementarity.

5.1. *Main Specification*

There are two types of labor, $\ell = 1, 2$, with h_ℓ denoting a firm's employment of labor of type ℓ and n_ℓ denoting a firm's measure of matches with labor of this type. Labor markets are occupation-specific and each one of them is similar to the labor market specified above. In particular, search and matching occurs separately for every occupation. We allow the expected income of a type-ℓ worker ω_ℓ, the resulting hiring costs b_ℓ, and tightness in the labor market x_ℓ to vary across occupations. The ability of every group is Pareto distributed with shape parameter k_ℓ and lower bound $a_{\min \ell}$.

The generalized production function is

$$(21) \qquad y = \theta(\bar{a}_1 h_1^{\gamma_1})^{\lambda_1}(\bar{a}_2 h_2^{\gamma_2})^{\lambda_2}, \quad \lambda_1 + \lambda_2 = 1.$$

This is a Cobb–Douglas extension of the production function (2) that allows for two occupation-specific tasks (e.g., engineers and managers). As before, the revenue function is $r = Ay^\beta$ for nonexporting firms and $r = Y_x^{1-\beta}Ay^\beta$ for exporting firms. The wage rates are determined in a Stole–Zwiebel bargaining game. This results in a wage bill for every type of worker which is a constant fraction of revenue,[29]

$$w_\ell h_\ell = \frac{\beta\gamma_\ell}{1+\beta\bar{\gamma}}\lambda_\ell r \quad \text{for} \quad \ell = 1, 2,$$

where $\bar{\gamma} \equiv \lambda_1\gamma_1 + \lambda_2\gamma_2$. Therefore, the problem of the firm, which is a generalization of (6), yields the following wage and employment schedules for the two occupational groups (see the Technical Appendix for details)[30]:

$$h_\ell(\theta) = h_{d\ell} \Upsilon(\theta)^{(1-\beta)(1-k_\ell/\delta)/\Gamma} \left(\frac{\theta}{\theta_d}\right)^{\beta(1-k_\ell/\delta)/\Gamma},$$

$$w_\ell(\theta) = w_{d\ell} \Upsilon(\theta)^{k_\ell(1-\beta)/(\delta\Gamma)} \left(\frac{\theta}{\theta_d}\right)^{\beta k_\ell/(\delta\Gamma)},$$

where now

$$\Gamma \equiv 1 - \beta\bar{\gamma} - \frac{\beta}{\delta}[1 - (\lambda_1\gamma_1 k_1 + \lambda_2\gamma_2 k_2)].$$

We next use this generalized solution to discuss wage dispersion within firms, and wage inequality within and between groups.

It is evident from the equations for the firm-specific variables above that employment and wages are rising with firm productivity in every occupation (provided that $k_\ell < \delta$ for both groups). However, the relative wage bills of the two types of workers are the same in every firm. Under these circumstances, relative wages are inversely proportional to relative employment across firms. In particular, we have

$$\frac{h_1(\theta)}{h_2(\theta)} = \frac{h_{d1}}{h_{d2}} \Upsilon(\theta)^{(1-\beta)(k_2-k_1)/(\delta\Gamma)} \left(\frac{\theta}{\theta_d}\right)^{\beta(k_2-k_1)/(\delta\Gamma)}$$

[29]This results from a solution to the system of differential equations $\partial(r - w_1h_1 - w_2h_2)/\partial h_\ell = w_\ell$ for $\ell = 1, 2$. See the Technical Appendix for details.

[30]The constants $h_{d\ell}$ and $w_{d\ell}$ ($\ell = 1, 2$) are generalizations of those provided in equation (17) for the baseline model without ex ante heterogeneity. They are

$$w_{d\ell} \equiv b_\ell \left[\frac{\lambda_\ell\beta(1-\gamma_\ell k_\ell)}{\Gamma} \frac{f_d}{c a_{\min\ell}^\delta}\right]^{k_\ell/\delta},$$

$$h_{d\ell} \equiv \frac{\lambda_\ell\beta\gamma_\ell}{\Gamma}\frac{f_d}{b_\ell}\left[\frac{\lambda_\ell\beta(1-\gamma_\ell k_\ell)}{\Gamma} \frac{f_d}{c a_{\min\ell}^\delta}\right]^{-k_\ell/\delta}.$$

and

$$\frac{w_1(\theta)}{w_2(\theta)} = \frac{w_{d1}}{w_{d2}} \Upsilon(\theta)^{(1-\beta)(k_1-k_2)/(\delta\Gamma)} \left(\frac{\theta}{\theta_d}\right)^{\beta(k_1-k_2)/(\delta\Gamma)}.$$

It follows that more productive firms employ relatively more of type-1 workers if and only if $k_1 < k_2$, that is, if and only if the ability of type-1 workers is more dispersed than the ability of type-2 workers. In what follows, we assume this to be the case. Under these circumstances, more productive firms pay relatively higher wages to type-2 workers.[31] The intuition for this relationship between relative wages and employment across firms of different productivities is as follows. Since more productive firms employ relatively more of type-1 workers, this weakens their bargaining position relative to type-2 workers, and hence more productive firms pay relatively higher wages to type-2 workers.

Note that the occupation-specific degree of decreasing returns γ_ℓ only affects the relative wages and employment of firms through the composite derived parameter Γ. Note also that the relationship between relative wages and employment across firms of different productivities does not depend on the levels of human capital of workers in each group. High human capital of workers in group ℓ (high $a_{\min \ell}$) only affects relative employment and wages of the two groups through the cutoffs $h_{d\ell}$ and $w_{d\ell}$, which are common to all firms.

5.1.1. *Within-Group Inequality*

We can use the above solutions to calculate the distribution of wages within every occupation, as described in Section 3. As before, the distribution of wages within an occupation is an untruncated Pareto in the closed economy. This distribution now has an occupation-specific shape parameter $1 + 1/\mu_\ell$, where $\mu_\ell = \beta k_\ell / [\delta(z\Gamma - \beta)]$. It follows that $k_1 < k_2$ implies $\mu_1 < \mu_2$, so that there is more wage dispersion in group 2. Group 1 has a steeper employment schedule, but a flatter wage schedule across firms. The second effect dominates, leading to less wage inequality within group 1.

As in Section 3, the open economy wage distribution within an occupation has two components: a truncated Pareto among workers employed by firms that serve only the domestic market and an untruncated Pareto among workers employed by exporters, with $1 + 1/\mu_\ell$ being the common shape parameter of these distributions. As a result, we can use the same method as in Section 3 to prove that within every occupation, there is more wage inequality in the trade regime than in autarky, as long as some, but not all, firms export (see Proposition 3). Similarly, within every occupation, wage inequality rises with trade

[31] If the screening technology were also to differ across occupations, then the relative comparison would be between k_1/δ_1 and k_2/δ_2 instead of k_1 and k_2, while differences in c have no effect on these variations across firms.

openness initially and later declines (see Corollary to Proposition 3). Therefore our results for within-group inequality naturally generalize to the case of multiple occupations with ex ante differences in worker characteristics across occupations.

5.1.2. *Between-Group Inequality*

Next consider the impact of trade on wage inequality across occupations. Average wages for an occupation are given by[32]

$$\bar{w}_\ell = (1+\mu_\ell)w_{d\ell} \cdot \frac{1+\rho^{z-\beta/\Gamma}[\Upsilon_x^{(1-\beta)/\Gamma}-1]}{1+\rho^{z-\beta(1-k_\ell/\delta)/\Gamma}[\Upsilon_x^{(1-\beta)(1-k_\ell/\delta)/\Gamma}-1]}, \quad \ell = 1, 2.$$

In the case where $k_1 < k_2$ and expected worker income is unchanged as a result of the opening of trade, it is straightforward to show that trade causes the average wage in occupation 2 to rise relative to occupation 1 (i.e., $\bar{w}_2/\bar{w}_1$ increases).[33] As discussed earlier, there are two effects of trade openness on the average wage for an occupation: the higher wages paid by exporting firms and the reallocation of employment toward high-wage exporting firms. The first effect is stronger for the high-k_ℓ group, while the second effect is stronger for the low-k_ℓ group. The first effect dominates and the relative average wage of the high-k_ℓ occupation rises.

How do these results affect between-group inequality? With two occupations, different measures of between-group inequality (including the Theil index and the Gini coefficient) achieve their minimum when $\bar{w}_2/\bar{w}_1 = 1$, and inequality rises in $\bar{w}_2/\bar{w}_1$ if and only if $\bar{w}_2 > \bar{w}_1$. It follows that trade raises between-group inequality when $\bar{w}_2 > \bar{w}_1$ in autarky. This occurs, for example, when $w_{d1} = w_{d2}$, because $k_1 < k_2$ implies $\mu_1 < \mu_2$. In contrast, trade reduces between-group inequality when $\bar{w}_1 > \bar{w}_2$ in both autarky and the trade equilibrium. This happens when labor market tightness for group 1 is sufficiently larger than for group 2, so that b_1/b_2 and, hence, w_{d1}/w_{d2} are sufficiently large, where b_1/b_2 depends on relative expected incomes for the two groups of workers (ω_1/ω_2). From this discussion, it is clear that between-group inequality can either rise or fall as a result of the opening to trade. But even if between-group inequality falls, its decline can be dominated by the rise in within-group inequality, so that the opening of trade raises overall wage inequality.

This analysis of between-group inequality has so far abstracted from effects of trade on the relative expected incomes of the two groups of workers

[32]Note that $\bar{w}_\ell^A = (1+\mu_\ell)w_{d\ell}$ is the autarky average wage rate. The derivation of the average wage can be found in the Technical Appendix in the proof of Proposition 3.

[33]The formal argument is the following. Given $k_1 < k_2$, an increase of ρ from zero to a positive value increases $\bar{w}_2/\bar{w}_1$. Moreover, the partial derivative of $\bar{w}_2/\bar{w}_1$ with respect to ρ is positive. Both arguments assume b_2/b_1 and, hence, w_{d2}/w_{d1} are constant, which is the case when expected worker income is unchanged as a result of the opening of trade.

(ω_1/ω_2), as analyzed in the section on general equilibrium below. A movement in ω_1/ω_2 can, however, be a dominant force behind the change in between-group inequality by affecting w_{d1}/w_{d2} through relative search costs b_1/b_2. Our framework can be nested in a two-sector Heckscher–Ohlin model in which trade and the relative supply of the two types of workers determine the relative expected incomes of the two groups (see the Technical Appendix). In this specification, within-group inequality responds to trade according to Proposition 3 while the response of between-group inequality is shaped by the standard Stolper–Samuelson forces. Our predictions for within-group inequality are, therefore, robust to these extensions of our model to multifactor multisector environments.

5.2. *Technology–Skill Complementarity*

While the Cobb–Douglas production technology provides a tractable framework within which to demonstrate the robustness of our results to the introduction of ex ante heterogeneity, it implies that the relative wage bills of the two types of workers are constant across firms. This feature imposes a tight link between differences in ability dispersion between the two groups of workers, $k_2 - k_1$, and variation in the relative wages and employment of the two groups of workers across firms of different productivities. It, therefore, also imposes a tight link between $k_2 - k_1$ and differences in wage dispersion between the two groups of workers. To break this tight link, the model can be generalized to a more flexible CES production technology[34]:

$$y = [\lambda_1(\theta_1\bar{a}_1 h_1^{\gamma})^{\nu} + \lambda_2(\theta_2\bar{a}_2 h_2^{\gamma})^{\nu}]^{1/\nu}, \quad 0 < \nu \leq \beta,\ \lambda_1 + \lambda_2 = 1.$$

Interpreting $\ell = 1$ as skilled labor and $\ell = 2$ as unskilled labor, we can treat $\theta_1 = \theta$ as the productivity level and set $\theta_2 \equiv 1$. This specification exhibits technology–skill complementarity, which is a known feature of the data. To focus on this feature, we also assume $k_1 = k_2 = k$. The limit $\nu \to 0$ results in the Cobb–Douglas case studied above, while we have imposed $\nu \leq \beta$ to insure that employment of both types of labor increases with firm productivity.

In the Technical Appendix, we show that in this case, the Stole–Zwiebel bargaining game yields the equilibrium wages

$$w_\ell h_\ell = \frac{\beta\gamma}{1+\beta\gamma}\chi_\ell r, \quad \chi_\ell \equiv \frac{\lambda_\ell(\theta_\ell\bar{a}_\ell h_\ell^{\gamma})^{\nu}}{\sum_{j=1}^{2}\lambda_j(\theta_j\bar{a}_j h_j^{\gamma})^{\nu}},\ \chi_1 + \chi_2 \equiv 1.$$

[34]Alternatively, to address this issue, one can study a model where firms of different productivity choose different technologies which use the two types of labor with different intensities. We discuss this extension in the Technical Appendix.

As before, the aggregate wage bill is a fraction $\beta\gamma/(1+\beta\gamma)$ of revenue r, and now the wage bill of input ℓ is a fraction χ_ℓ of the total wage bill. Using this wage structure, we show in the Technical Appendix that in this case, the solution to the firm's problem, which is a generalization of (6), yields

$$\frac{h_1(\theta)}{h_2(\theta)} = \kappa_h \theta^{(\nu/\Lambda)(1-k/\delta)},$$
$$\frac{w_1(\theta)}{w_2(\theta)} = \kappa_w \theta^{k\nu/(\delta\Lambda)},$$

where κ_h and κ_w are constants that depend on parameters and the relative equilibrium search costs for the two groups of workers (b_1/b_2), and $\Lambda \equiv 1 - \nu\gamma - \nu(1-\gamma k)/\delta > 0$. In words, more productive firms employ relatively more skilled workers and skilled workers are paid relatively more by more productive firms. That is, the share of group-1 workers in the total wage bill increases with firm productivity, in contrast to the constant share imposed by the Cobb–Douglas specification.

As in the single input case, we show in the Technical Appendix that more productive firms pay higher wages to both types of workers and they employ more workers of each type. More productive firms also select into exporting, and there is a discontinuous upward jump in revenue and, hence, wages and employment for each group of workers at the productivity cutoff θ_x above which firms export. This jump in wages contributes to wage inequality in the trade equilibrium when not all firms export, although we cannot directly apply the arguments from Section 3 so as to extend Proposition 3 to this case.[35]

An interesting special case arises when $\nu = \beta$. In this case, all nonexporters pay the same wages to the unskilled workers and employ the same number of these workers, and, similarly, all exporters pay the same wages to the unskilled and employ the same number of these workers, except that exporters pay higher unskilled wages than nonexporters. Additionally, in this case, wages and employment of skilled workers are power functions of firm productivity θ. Therefore, Proposition 3 applies. It follows that in the closed economy, wage dispersion is greater for skilled workers than unskilled workers, and the opening of the closed economy to trade increases within-group wage inequality for both skilled and unskilled workers.

6. GENERAL EQUILIBRIUM

In this section, we return to our benchmark model with a single type of worker and embed the sector in general equilibrium to determine expected

[35]The reason that the previous proof does not directly apply is that wages are no longer power functions of θ. The Technical Appendix provides closed-form solutions for employment and wages of the two skill groups as a function of firm productivity.

worker income (ω), prices, and aggregate income. We use this characterization of general equilibrium to examine the impact on worker welfare of opening the closed economy to trade.

Individual workers in the differentiated sector experience idiosyncratic income risk as a result of the positive probability of unemployment and wage dispersion. We assume that preferences are defined over an aggregate consumption index ($\mathcal{C}$) and exhibit constant relative risk aversion (CRRA):

$$\mathbb{U} = \frac{\mathbb{E}\mathcal{C}^{1-\eta}}{1-\eta}, \quad 0 \leq \eta < 1, \tag{22}$$

where $\mathbb{E}$ is the expectation operator. Expected indirect utility is therefore

$$\mathbb{V} = \frac{1}{1-\eta}\mathbb{E}\left(\frac{w}{\mathcal{P}}\right)^{1-\eta},$$

where $\mathcal{P}$ is the price index of the aggregate consumption index $\mathcal{C}$. While we initially assume that workers are risk neutral ($\eta = 0$), we discuss the implications of introducing risk aversion below ($0 < \eta < 1$).

To demonstrate the robustness of our sectoral equilibrium results, we consider several approaches to embedding the sector in general equilibrium. First, we adopt a standard approach from international trade of introducing an outside good, which is homogeneous and produced without search frictions.[36] This approach allows for the possibility that some sectors of the economy are best characterized by neoclassical assumptions and is particularly tractable. Expected worker income is pinned down by the wage in the outside sector as long as countries are incompletely specialized across sectors. As a result, the opening of trade leaves expected worker income unchanged as long as countries remain incompletely specialized.

Second we consider a single-sector economy where expected worker income responds to the opening of trade. While the model's predictions for sectoral wage inequality are the same for both approaches to closing the model, the impact of opening the closed economy to trade on sectoral unemployment depends on what happens to expected worker income, as discussed above.

6.1. *Expenditure, Mass of Firms, and the Labor Force*

Before considering the above two approaches to closing the model in general equilibrium, we first provide some additional conditions for sectoral equilibrium that can be used to solve for other sectoral variables of interest. While

[36]While we assume no search frictions in the outside sector, Helpman and Itskhoki (2010) showed in a model without worker heterogeneity or screening that introducing search frictions in the outside sector generates an expected income ω_0 that is independent of features of the differentiated sector. Augmenting the model here to incorporate search frictions in the outside sector would generate a similar result.

these variables have not been needed for the analysis so far, they are used below when we embed the sector in general equilibrium.

Recall that the system of equilibrium conditions discussed in Section 2 allowed us to solve $(\theta_d, \theta_x, A, \theta_d^*, \theta_x^*, A^*)$ for each differentiated product sector. From utility maximization across varieties within each sector, the demand shifter (A) depends on total expenditure on the sector's varieties (E) and the sector's ideal price index (P) as $A = E^{1-\beta}P^{\beta}$. Additionally, utility maximization across sectors implies that total expenditure on the sector's varieties depends on the sectoral price indices and aggregate income Ω. Therefore, utility maximization implies that the demand shifter for each sector i is a function of sectoral price indices and aggregate income:

$$A_i = \tilde{A}_i(\mathbf{P}, \Omega), \tag{23}$$

where the function $\tilde{A}(\cdot)$ is derived from the structure of preferences, $\mathbf{P}$ is a vector of sectoral price indexes P_i, and Ω is aggregate income. Given the sectoral demand shifters for each sector determined above, equation (23) for each sector provides a system of equations that can be used to solve for the sectoral price indices as a function of aggregate income.

Next consider a particular sector i, for which we now have both A_i and P_i; we now drop the subscript i. In this sector, the real consumption index (Q) follows from utility maximization across varieties, which implies

$$Q = (A/P)^{1/(1-\beta)} \tag{24}$$

and yields total expenditure within the sector $E = PQ$. Similar relationships determine the foreign price index, real consumption index, and total expenditure within the sector (P^*, Q^*, E^*, respectively).

The mass of firms within the sector (M) can be determined from the market clearing condition that total domestic expenditure on differentiated varieties equals the sum of the revenues of domestic and foreign firms that supply varieties to the domestic market:

$$E = M\int_{\theta_d}^{\infty} r_d(\theta)\, dG_\theta(\theta) + M^* \int_{\theta_x^*}^{\infty} r_x^*(\theta)\, dG_\theta(\theta). \tag{25}$$

From $r_d(\theta) = r(\theta)/\Upsilon(\theta)$, $r_x(\theta) = r(\theta)(\Upsilon(\theta) - 1)/\Upsilon(\theta)$,[37] and total firm revenue (8), domestic and foreign revenue can be expressed in terms of variables that have already been determined $(\theta_d, \theta_x, \Upsilon(\theta))$. Therefore, we can solve for the mass of firms in each country (M, M^*) from (25) and a similar equation for foreign.

[37]See footnote 15.

The mass of workers searching for employment in the sector (L) can be determined by noting that total labor payments are a constant fraction of total revenue from the solution to the bargaining game:

$$\omega L = M \int_{\theta_d}^{\infty} w(\theta) h(\theta) \, dG_\theta(\theta) = M \frac{\beta\gamma}{1+\beta\gamma} \int_{\theta_d}^{\infty} r(\theta) \, dG_\theta(\theta),$$

where we have solved for the mass of firms (M) and total firm revenue ($r(\theta)$) above, and where a similar equation determines the sectoral labor force in foreign (L^*). Finally, we also require that the sectoral labor force is less than or equal to the supply of labor ($L \leq \bar{L}$), as discussed below.

6.2. *Economy With an Outside Sector*

Having solved for the remaining components of sectoral equilibrium, we now turn to our first approach to embedding the sector in general equilibrium. The aggregate consumption index ($\mathcal{C}$) is defined over consumption of a homogeneous outside good (q_0) and a real consumption index of differentiated varieties (Q):

$$\mathcal{C} = [\vartheta^{1-\zeta} q_0^{\zeta} + (1-\vartheta)^{1-\zeta} Q^{\zeta}]^{1/\zeta}, \quad 0 < \zeta < \beta,$$

where Q is modelled as in Section 2 and workers are assumed to be risk neutral. The parameter ϑ determines the relative weight of the homogeneous and differentiated sectors in consumer preferences.[38] While for simplicity we consider a single differentiated sector, the analysis generalizes in a straightforward way to the case of multiple differentiated sectors.

In the homogeneous sector, the product market is perfectly competitive and there are no labor market frictions. In this sector, one unit of labor is required to produce one unit of output and there are no trade costs. Therefore, as we choose the homogeneous good as the numeraire ($p_0 = 1$), the wage in this sector is equal to 1 in both countries.

To determine expected worker income in the differentiated sector, we use an indifference condition between sectors, which equates the expected utility of entering each sector in an equilibrium where both goods are produced. Under risk neutrality, this Harris–Todaro condition implies that expected worker income in the differentiated sector equals the certain wage of 1 in the homogeneous sector (see (11)),

$$xb = \omega = 1, \tag{26}$$

[38] While in the analysis here we assume that workers have CRRA-CES preferences and experience income risk, Helpman, Itskhoki, and Redding (2008a, 2008b) considered an alternative specification with quasilinear preferences and income insurance within families.

where incomplete specialization can be ensured by appropriate choice of labor endowments $(\bar{L}, \bar{L}^*)$, and relative preferences for the homogeneous and differentiated goods (ϑ). Positive unemployment occurs in the differentiated sector for a sufficiently large search friction α_0, such that $\alpha_0 > \omega = 1$ and, hence, $0 < x < 1$ in (12). Given an expected income of 1 in each sector, each country's aggregate income is equal to its labor endowment:

$$\Omega = \bar{L}. \tag{27}$$

To determine the price index in the differentiated sector (P), we use the functional relationship (23), which, with CES preferences between the homogeneous and differentiated sector, takes the form

$$A^{1/(1-\beta)} = \frac{(1-\vartheta)P^{(\beta-\zeta)/((1-\beta)(1-\zeta))}\Omega}{\vartheta + (1-\vartheta)P^{-\zeta/(1-\zeta)}}, \tag{28}$$

where the right-hand side is monotonically increasing in P. Therefore, this relationship uniquely pins down P given the demand shifter (A) and aggregate income (Ω).

To determine general equilibrium, we use the conditions for sectoral equilibrium in Sections 2 and 6.1 (where (28) replaces (23)), and combine them with the Harris–Todaro condition (26) and aggregate income (27). Together these relationships determine the equilibrium vector $(x, b, \theta_d, \theta_x, A, Q, P, M, L, \omega, \Omega)$. The model has a recursive structure and the equilibrium vector is unique as shown in the Technical Appendix. Having determined this equilibrium vector, the price index $\mathcal{P}$—dual to the aggregate consumption index $\mathcal{C}$—and consumption of the homogeneous good q_0 follow from CES demand. Finally, equilibrium employment in the homogeneous sector follows from labor market clearing ($L_0 = \bar{L} - L$, where incomplete specialization requires $L < \bar{L}$).

Having characterized general equilibrium, we are now in a position to examine the impact of the opening of trade on ex ante expected welfare. Note that differentiated sector workers receive the same expected indirect utility as workers in the homogeneous sector when both goods are produced:

$$\mathbb{V} = \frac{1}{\mathcal{P}} \quad \text{for} \quad \eta = 0.$$

Therefore, the change in expected welfare as a result of the opening of trade depends solely on the change in the aggregate price index ($\mathcal{P}$), which, with our choice of numeraire, depends solely on the change in the price index for the differentiated sector (P). These comparative statics are straightforward to determine. From the free entry condition (16), the opening of trade raises the zero-profit productivity cutoff (θ_d). Using the Harris–Todaro condition (26) and labor market tightness (12), search costs (b) remain constant as long as both goods are produced, because expected worker income equals 1. Therefore, from the zero-profit cutoff condition (13), the rise in θ_d implies a lower value of the demand shifter (A). Given constant aggregate income (Ω) and a lower value of A, CES demand (28) implies that the opening of trade reduces the price index for the differentiated sector (P), which implies higher expected welfare in the open economy than in the closed economy. We, therefore, can make the following statement.

PROPOSITION 6: *Let $\eta = 0$. Then in the two-sector economy, every worker's ex ante welfare is higher in the open economy than in the closed economy.*

With constant expected worker income (ω), labor market tightness (x) is unchanged as a result of the opening of trade and, hence, the change in the sectoral unemployment rate (u) depends solely on the change in the hiring rate (σ).

6.3. *Single-Sector Economy*

We now turn to our second approach to embedding the sector in general equilibrium. The aggregate consumption index ($\mathcal{C}$) is defined over consumption of a continuum of horizontally differentiated varieties,

$$\mathcal{C} = Q,$$

where Q again takes the same form as in Section 2.

All search, screening, fixed production, and fixed exporting costs are denominated in terms of the aggregate consumption index, which implies that these activities use the output of each differentiated variety in the same way as is demanded by final consumers. While for simplicity we again focus on a single differentiated sector, the analysis generalizes in a straightforward way to the case where the aggregate consumption index is defined over the consumption of many sectors, each of which contains a continuum of horizontally differentiated varieties.

To determine general equilibrium, we follow the same approach as for sectoral equilibrium in Sections 2 and 6.1, where these equations now apply to the economy as a whole and we also solve for expected worker income (ω) and aggregate income (Ω). With a single sector, the analysis in Section 6.1 is slightly modified. Expenditure and the mass of firms in the two countries are determined by three equations for each country: total expenditure is the sum of expenditure on home and foreign varieties, total revenue is the sum of domestic expenditure on home varieties plus foreign expenditure on home varieties, and total expenditure equals total revenue.

Labor market clearing requires that total employment in the sector equals the economy's supply of labor, and expected worker income is determined from the requirement that total labor payments are a constant share of total revenue. Aggregate income follows immediately from expected worker income times the economy's supply of labor. Finally, we choose the aggregate consumption index in one country as the numeraire ($P = 1$) and solve for the dual price index in the other country (P^*) using the sectoral demand shifter (A^*), total expenditure (E^*), and CES demand ($A^* = (E^*)^{1-\beta}(P^*)^{\beta}$).

In the case of symmetric countries, the single-sector model can be solved in closed form, as shown in the Technical Appendix. The closed-form solution for expected worker income is

$$\omega = \alpha_0^{\gamma\beta/((1-\beta)\Delta)} c^{\beta(1-\gamma k)(1+\alpha_1)/(\delta(1-\beta)\Delta)} \theta_d^{-\beta(1+\alpha_1)/((1-\beta)\Delta)} \bar{L}^{-(1+\alpha_1)/\Delta} \kappa_b^{(1+\alpha_1)/\Delta},$$

$$\Delta \equiv (1+\alpha_1) - \frac{\beta\gamma\alpha_1}{(1-\beta)} < 0,$$

where κ_b is defined in the Technical Appendix.[39]

As the opening of trade increases the zero-profit cutoff productivity below which firms exit (θ_d), it increases expected worker income and, hence, ex ante welfare.

PROPOSITION 7: *Let* $\eta = 0$. *Then in the one-sector economy the opening of trade* (i) *increases expected worker income* (ω) *and*, *hence*, *ex ante welfare*, *and* (ii) *increases labor market tightness* (x) *and search costs* (b).

While the predictions of the model without the outside sector for wage inequality are the same as those of the model with the outside sector, there is a new general equilibrium effect for unemployment through the tightness of the labor market (x).

6.4. *Risk Aversion*

Income risk in the differentiated sector arises from random search, which implies a positive probability of unemployment and uncertainty over a worker's employer, and, hence, uncertainty over the wage received in employment. In this section, we introduce risk aversion by considering the case where utility is concave in the aggregate consumption index ($0 < \eta < 1$). In the model with an outside sector, this implies that workers will require a risk premium to enter the differentiated sector rather than receiving a certain wage of 1 in the outside sector.

To determine general equilibrium in the model with the outside sector, we follow a similar approach as in Section 6.2 above.[40] Risk aversion changes the

[39]The stability of the equilibrium requires $\frac{\beta\gamma}{1-\beta}\frac{\alpha_1}{1+\alpha_1} > 1$, which is satisfied for sufficiently convex search costs (sufficiently high α_1) and sufficiently high elasticities of substitution between varieties (β sufficiently close to but less than 1). For parameter values satisfying this inequality, the equilibrium vector is again unique.

[40]Introducing risk aversion in the model with a single differentiated sector has little effect, because there is no riskless activity to or from which resources can move.

equilibrium share of revenue received by workers in the bargaining game, but does not otherwise affect the determination of sectoral equilibrium within the differentiated sector.[41] Additionally, risk aversion modifies the Harris–Todaro condition for indifference between sectors, which equates expected utility in the differentiated sector to the certain wage of 1 in the outside sector:

$$x\sigma \mathbb{E}w^{1-\eta} = x\sigma \int_{w_d}^{\infty} w^{1-\eta}\, dG_w(w) = 1,$$

where expected utility in the differentiated sector equals the probability of being matched (x) times the probability of being hired conditional on being matched (σ) times expected utility conditional on being hired.[42]

To determine expected worker income in the differentiated sector ($\omega = xb$), we use the modified Harris–Todaro condition above and the search technology (10), which together imply

$$\begin{aligned} \omega = {} & (\alpha_0)^{\eta/(1+(1-\eta)\alpha_1)}[(1+\mu\eta)\phi_w^{\eta}]^{(\alpha_1+1)/(1+(1-\eta)\alpha_1)} \\ & \times \Lambda(\rho, Y_x)^{-(1+\alpha_1)/(1+(1-\eta)\alpha_1)}, \end{aligned} \tag{29}$$

where $\Lambda(\rho, Y_x)$ and ϕ_w are defined in the Technical Appendix, and satisfies $\Lambda(0, Y_x) = 1$ and $0 < \Lambda(\rho, Y_x) < 1$ for $0 < \rho \leq 1$. A sufficiently large search friction (α_0) ensures positive unemployment ($0 < x < 1$ in (12)) and a positive risk premium in the differentiated sector ($\omega - 1 > 0$).

As shown in the analysis of sectoral equilibrium in Sections 2–4 above, the opening of trade increases sectoral wage inequality and unemployment for a given value of ω. This increase in wage inequality and unemployment enhances income risk in the differentiated sector, which implies that risk averse workers require a higher risk premium to enter the differentiated sector. This "risk effect" raises expected worker income (ω) following the opening of trade (since $\Lambda(0, Y_x) = 1$ and $0 < \Lambda(\rho, Y_x) < 1$ for $0 < \rho \leq 1$ in (29)). This increase in expected worker income in turn increases labor market tightness (x) and search costs (b).

PROPOSITION 8: *Let* $0 < \eta < 1$. *Then in the two-sector economy*, *the opening of trade* (i) *increases expected worker income* (ω), *and* (ii) *increases labor market tightness* (x) *and search costs* (b).

[41]In the Technical Appendix, we derive the solution to the Stole and Zwiebel (1996a, 1996b) bargaining game when workers are risk averse. We show that with CRRA-CES preferences, the solution takes a similar form as when there are differences in bargaining weight between the firm and its workers.

[42]The terms in the price index ($\mathcal{P}$) and $1/(1-\eta)$ cancel from the Harris–Todaro condition equating expected utility in the two sectors.

While the predictions of the model with risk aversion for wage inequality are the same as those of the model with risk neutrality, there is again a new general equilibrium effect for unemployment through the tightness of the labor market (x).

7. CONCLUSION

The relationship between international trade and earnings inequality is one of the most hotly debated issues in economics. Traditionally, research has approached this topic from the perspective of neoclassical trade theory with its emphasis on specialization across industries and changes in the relative rewards of skilled and unskilled labor. In this paper, we propose a new framework that emphasizes firm heterogeneity, Diamond–Mortensen–Pissarides search and matching frictions, and ex post heterogeneity in worker ability.

In this framework, the participation of some, but not all, firms in international trade provides a new mechanism for trade to affect wage inequality. We derive two results that hold in a class of models for which firm wages and employment are power functions of firm productivity, exporting increases the wage paid by a firm with a given productivity, and firm productivity is Pareto distributed. We show that the opening of the closed economy to trade raises sectoral wage inequality for any measure of inequality that respects second-order stochastic dominance, because the opening of trade increases the dispersion of firm revenue, which in turn increases the dispersion of firm wages. Once the economy is open to trade, wage inequality is at first increasing and later decreasing in trade openness. Therefore, a given change in trade openness can either raise or reduce wage inequality, depending on the initial level of trade openness.

The relationship between firm wages and revenue in our framework is derived from search and matching frictions and heterogeneity in ex post worker ability. Larger firms screen workers more intensively and have workforces of higher average ability, who are more costly to replace in the bargaining game and are, therefore, paid higher wages. As the opening of the closed economy to trade reallocates resources toward more productive firms that screen more intensively, it also affects unemployment. While the fraction of matched workers that are hired necessarily falls, the fraction of workers searching for employment that are matched can either remain unchanged or rise. Therefore, in contrast to the model's unambiguous predictions for wage inequality, its predictions for unemployment are more nuanced, and measures of wage and income inequality can yield quite different pictures of the impact of trade liberalization.

While for most of our analysis we concentrate on within-group wage inequality among ex ante identical workers, we show that our results on within-group wage inequality are robust to introducing multiple worker types with different ex ante observable characteristics. While trade raises wage inequality within every group of workers, it may increase or reduce wage inequality between different groups of workers. But even when between-group wage inequality falls, the rise of within-group inequality can dominate so that the opening of trade raises overall wage inequality.

Our model thus provides a unified framework for analyzing the complex interplay between wage inequality, unemployment and income inequality, and their relation to international trade. In popular discussion, the effects of international trade on income distribution are sometimes viewed as largely transitory, because it takes time for resources to be reallocated. In contrast, our framework identifies a systematic mechanism through which the opening of trade can raise equilibrium wage inequality across workers within sectors. This mechanism is founded in the new view of foreign trade that emphasizes firm heterogeneity in differentiated product markets and is consistent with observed features of firm and worker data. The tractability and flexibility of our framework lend themselves to a variety of further applications.

REFERENCES

ABOWD, J. M., R. H. CREECY, AND F. KRAMARZ (2002): "Computing Person and Firm Effects Using Linked Longitudinal Employer–Employee Data," Mimeograph, Cornell University. [1255]

ABOWD, J. M., F. KRAMARZ, AND D. N. MARGOLIS (1999): "High Wage Workers and High Wage Firms," *Econometrica*, 67, 251–333. [1255]

ACEMOGLU, D. (1999): "Changes in Unemployment and Wage Inequality; An Alternative Theory and Some Evidence," *American Economic Review*, 89, 1259–1278. [1243,1244,1252]

ACEMOGLU, D., AND R. SHIMER (1999): "Efficient Unemployment Insurance," *Journal of Political Economy*, 107, 893–928. [1244,1252]

ACEMOGLU, D., P. ANTRÀS, AND E. HELPMAN (2007): "Contracts and Technology Adoption," *American Economic Review*, 97, 916–943. [1249]

ALTONJI, J. G., AND C. R. PIERRET (2001): "Employer Learning and Statistical Discrimination," *Quarterly Journal of Economics*, February, 313–350. [1246]

AMITI, M., AND D. R. DAVIS (2008): "Trade, Firms, and Wages: Theory and Evidence," Working Paper 14106, NBER. [1240,1243]

ATTANASIO, O., P. GOLDBERG, AND N. PAVCNIK (2004): "Trade Reforms and Wage Inequality in Colombia," *Journal of Development Economics*, 74, 331–366. [1241]

AUTOR, D. H., AND D. SCARBOROUGH (2008): "Does Job Testing Harm Minority Workers? Evidence From Retail Establishments," *Quarterly Journal of Economics*, 123, 219–277. [1247]

AXTELL, R. L. (2001): "Zipf Distribution of U.S. Firm Sizes," *Science*, 293, 1818–1820. [1240, 1245]

BARRON, J. M., J. BISHOP, AND W. C. DUNKELBERG (1985): "Employer Search: The Interviewing and Hiring of New Employees," *Review of Economics and Statistics*, 67, 43–52. [1247]

BARRON, J. M., D. BLACK, AND M. LOEWENSTEIN (1987): "Employer Size: The Implications for Search, Training, Capital Investment, Starting Wages, and Wage Growth," *Journal of Labor Economics*, 5, 76–89. [1247,1256]

BERNARD, A. B., AND J. B. JENSEN (1995): "Exporters, Jobs, and Wages in U.S. Manufacturing: 1976–1987," *Brookings Papers on Economic Activity: Microeconomics*, 1995, 67–112. [1250,1256]

——— (1997): "Exporters, Skill Upgrading and the Wage Gap," *Journal of International Economics*, 42, 3–31. [1256]

BLANCHARD, O., AND J. GALÍ (2010): "Labor Markets and Monetary Policy: A New-Keynesian Model With Unemployment," *American Economic Journal: Macroeconomics*, 2, 1–30. [1252]

BURDETT, K., AND D. MORTENSEN (1998): "Wage Differentials, Employer Size and Unemployment," *International Economic Review*, 39, 257–273. [1243]

BURSTEIN, A., AND J. VOGEL (2009): "Globalization, Technology and the Skill Premium," Mimeograph, Columbia University. [1243]

BUSTOS, P. (2009): "Trade Liberalization, Exports and Technology Upgrading: Evidence on the Impact of MERCOSUR on Argentinean Firms," Mimeo, CREI. [1243]

CAHUC, P., F. POSTEL-VINAY, AND J.-M. ROBIN (2006): "Wage Bargaining With On-the-Job Search: Theory and Evidence," *Econometrica*, 74, 323–364. [1243]

COSTINOT, A., AND J. VOGEL (2009): "Matching and Inequality in the World Economy," Mimeograph, MIT. [1243]

DAVIDSON, C., AND S. MATUSZ (2009): *International Trade With Equilibrium Unemployment*. Princeton, NJ: Princeton University Press. [1242,1265]

DAVIDSON, C., L. MARTIN, AND S. MATUSZ (1988): "The Structure of Simple General Equilibrium Models With Frictional Unemployment," *Journal of Political Economy*, 96, 1267–1293. [1243]

——— (1999): "Trade and Search Generated Unemployment," *Journal of International Economics*, 48, 271–299. [1243]

DAVIDSON, C., S. MATUSZ, AND A. SHEVCHENKO (2008): "Globalization and Firm-Level Adjustment With Imperfect Labor Markets," *Journal of International Economics*, 75, 295–309. [1244]

DAVIS, D., AND J. HARRIGAN (2007): "Good Jobs, Bad Jobs, and Trade Liberalization," Working Paper 13139, NBER. [1243]

DAVIS, S. J., AND J. HALTIWANGER (1991): "Wage Dispersion Between and Within U.S. Manufacturing Plants, 1963–1986," *Brookings Papers on Economic Activity: Microeconomics*, 1991, 115–200. [1259]

DE MELO, R. L. (2008): "Sorting in the Labor Market: Theory and Measurement," Mimeograph, Yale University. [1255]

DIAMOND, P. A. (1982a): "Demand Management in Search Equilibrium," *Journal of Political Economy*, 90, 881–894. [1243]

——— (1982b): "Wage Determination and Efficiency in Search Equilibrium," *Review of Economic Studies*, 49, 217–227. [1243]

EGGER, H., AND U. KREICKEMEIER (2009a): "Firm Heterogeneity and the Labour Market Effects of Trade Liberalization," *International Economic Review*, 50, 187–216. [1240,1243]

——— (2009b): "Fairness, Trade, and Inequality," Mimeograph, University of Nottingham. [1240,1243]

FAGGIO, G., K. G. SALVANES, AND J. VAN REENEN (2007): "The Evolution of Inequality in Productivity and Wages: Panel Data Evidence," Discussion Paper 821, CEPR. [1259]

FEENSTRA, R. C., AND G. H. HANSON (1996): "Foreign Investment, Outsourcing and Relative Wages," in *The Political Economy of Trade Policy*, ed. by R. C. Feenstra, G. M. Grossman, and D. A. Irwin. Cambridge, MA: MIT Press, Chapter 6. [1262]

FELBERMAYR, G., M. LARCH, AND W. LECHTHALER (2009): "Unemployment in an Interdependent World," Mimeograph, University of Stuttgart-Hohenheim. [1242,1243,1265]

FELBERMAYR, G., J. PRAT, AND H.-J. SCHMERER (2008): "Globalization and Labor Market Outcomes: Wage Bargaining, Search Frictions and Firm Heterogeneity," Discussion Paper 3363, IZA. [1242,1243,1265]

FRÍAS, J., D. KAPLAN, AND E. VERHOOGEN (2009): "Exports and Wage Premia: Evidence From Mexican Employer–Employee Data," Mimeograph, Columbia University. [1256]

GARICANO, L. (2000): "Hierarchies and the Organization of Knowledge in Production," *Journal of Political Economy*, 108, 874–904. [1246]

GOLDBERG, P., AND N. PAVCNIK (2007): "Distributional Effects of Globalization in Developing Countries," *Journal of Economic Literature*, 45, 39–82. [1241,1262]

HELPMAN, E., AND O. ITSKHOKI (2009): "Labor Market Rigidity, Trade and Unemployment: A Dynamic Model," Mimeograph, Harvard University. [1242]

——— (2010): "Labor Market Rigidities, Trade and Unemployment," *Review of Economic Studies*, 77, 1100–1137. [1239,1242,1243,1248,1265,1272]

HELPMAN, E., O. ITSKHOKI, AND S. J. REDDING (2008a): "Wages, Unemployment and Inequality With Heterogeneous Firms and Workers," Working Paper 14122, NBER. [1239,1255,1274]

——— (2008b): "Inequality and Unemployment in a Global Economy," Working Paper 14478, NBER. [1239,1260,1274]

——— (2010): "Supplement to 'Inequality and Unemployment in a Global Economy'," *Econometrica Supplemental Material*, 78, http://www.econometricsociety.org/ecta/Supmat/8640_extensions.pdf. [1244]

LENTZ, R. (2010): "Sorting by Search Intensity," *Journal of Economic Theory* (forthcoming). [1243]

LUCAS JR., R. E. (1978): "On the Size Distribution of Business Firms," *Bell Journal of Economics*, 9, 508–523. [1246]

MELITZ, M. J. (2003): "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71, 1695–1725. [1239,1245,1255,1256]

MENEZES-FILHO, N. A., M.-A. MUENDLER, AND G. RAMEY (2008): "The Structure of Worker Compensation in Brazil, With a Comparison to France and the United States," *Review of Economics and Statistics*, 90, 324–346. [1241]

MORETTI, E. (2004): "Workers' Education, Spillovers and Productivity: Evidence From Plant-Level Production Functions," *American Economic Review*, 94, 656–690. [1246]

MORTENSEN, D. T. (1970): "A Theory of Wage and Employment Dynamics," in *The Microeconomic Foundations of Employment and Inflation Theory*, ed. by E. S. Phelps et al. New York: Norton. [1243]

MORTENSEN, D. T., AND C. A. PISSARIDES (1994): "Job Creation and Job Destruction in the Theory of Unemployment," *Review of Economic Studies*, 61, 397–415. [1243]

MUNCH, J. R., AND J. R. SKAKSEN (2008): "Human Capital and Wages in Exporting Firms," *Journal of International Economics*, 75, 363–372. [1256]

OHNSORGE, F., AND D. TREFLER (2007): "Sorting It Out: International Trade With Heterogeneous Workers," *Journal of Political Economy*, 115, 868–892. [1243]

OI, W. Y., AND T. L. IDSON (1999): "Firm Size and Wages," in *Handbook of Labor Economics*, Vol. 3, ed. by O. Ashenfelter and D. Card. Amsterdam: Elsevier, Chapter 33, 2165–2214. [1255]

PELLIZZARI, M. (2005): "Employers' Search and the Efficiency of Matching," Discussion Paper 1862, IZA. [1247,1256]

PISSARIDES, C. A. (1974): "Risk, Job Search, and Income Distribution," *Journal of Political Economy*, 82, 1255–1268. [1243]

POSTEL-VINAY, F., AND J.-M. ROBIN (2002): "Equilibrium Wage Dispersion With Worker and Employer Heterogeneity," *Econometrica*, 70, 2295–2350. [1243]

ROBERTS, M. J., AND J. TYBOUT (1997): "The Decision to Export in Colombia: An Empirical Model of Entry With Sunk Costs," *American Economic Review*, 87, 545–564. [1250]

ROSEN, S. (1982): "Authority, Control and the Distribution of Earnings," *Bell Journal of Economics*, 13, 311–323. [1246]

SAEZ, E. (2001): "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 68, 205–229. [1240]

SCHANK, T., C. SCHANBEL, AND J. WAGNER (2007): "Do Exporters Really Pay Higher Wages? First Evidence From German Linked Employer–Employee Data," *Journal of International Economics*, 72, 52–74. [1256]

SHIMER, R., AND L. SMITH (2000): "Assortative Matching and Search," *Econometrica*, 68, 343–369. [1243]

STOLE, L. A., AND J. ZWIEBEL (1996a): "Organizational Design and Technology Choice Under Intrafirm Bargaining," *American Economic Review*, 86, 195–222. [1249,1278]

——— (1996b): "Intra-Firm Bargaining Under Non-Binding Contracts," *Review of Economic Studies*, 63, 375–410. [1249,1278]
VAN REENEN, J. (1996): "The Creation and Capture of Economic Rents: Wages and Innovation in a Panel of UK Companies," *Quarterly Journal of Economics*, 111, 195–226. [1255]
VERHOOGEN, E. (2008): "Trade, Quality Upgrading and Wage Inequality in the Mexican Manufacturing Sector," *Quarterly Journal of Economics*, 123, 489–530. [1242,1264]
YEAPLE, S. R. (2005): "A Simple Model of Firm Heterogeneity, International Trade, and Wages," *Journal of International Economics*, 65, 1–20. [1243]

Dept. of Economics, Harvard University, 217 Littauer Center, Cambridge, MA 02138, U.S.A. and CIFAR; ehelpman@harvard.edu,

Dept. of Economics, Princeton University, Fisher Hall 306, Princeton, NJ 08544-1021, U.S.A.; itskhoki@princeton.edu,

and

Dept. of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, United Kingdom; s.j.redding@lse.ac.uk.

Manuscript received June, 2009; final revision received February, 2010.