Performance Pay, Trade and Inequality

Germán P. Pupato*
Ryerson University

This draft: December 2012
First draft: May 2012
Preliminary Draft

Abstract

This paper introduces moral hazard into a standard general equilibrium model with heterogeneous firms, to study the impact of trade liberalization on wage inequality between identical workers. I show that trade liberalization operates on two margins of inequality, generating between- and within-firm wage dispersion. While the former channel has been studied in recent papers, the latter is novel in the literature. In the model, within-firm wage dispersion increases in firm productivity as a result of differential intensity in performance-pay compensation across firms. International trade liberalization generates labor reallocations towards high productivity firms that result in higher within-firm inequality.

1 Introduction

Our understanding of the impact of international trade on wage inequality has evolved substantially in the last twenty years. In the early 1990's, most economists discounted the role of trade as a driving force behind the steep increases in wage inequality that had been observed in many countries around the world since the late 1970's. Standard factor proportions theory was not easily reconcilable with increasing inequality in developing countries, the absence of significant reallocations of labor across industries and evidence showing that standard human capital variables like education and experience could account for only minor shares of the level and growth of inequality in both developed and developing countries.[1]

*I am grateful to Costas Arkolakis, Matilde Bombardini, Luis Braido, Eva Chalioti, Cecilia Fieler, Giovanni Gallipoli, Giovanni Maggi, Humberto Moreira and Marcelo Moreira for helpful comments and suggestions. The usual disclaimer applies.

[1] See Katz and Autor (1999) and Goldberg and Pavcnik (2007) for evidence on developed and developing countries, respectively.

In recent years, however, a new generation of trade models has caught up to these empirical challenges by shifting its focus from industries to firms, as the basic units of analysis. This research agenda has been fueled by numerous studies documenting a set of stylized facts regarding heterogeneity in firm-level outcomes within industries, including systematic differences between exporting and non-exporting firms. Recent theories place particular emphasis on the finding that more productive firms pay higher average wages, even after controlling for worker characteristics, including education, occupation and industry.[2] In a recent study using Brazilian data, Helpman et al. (2012) report that 38% of the variance of log wages within sector-occupation cells in 1990 can be accounted for by the variation in wage premia across firms. These facts are compatible with models of firm heterogeneity that feature search frictions and bargaining (Davidson et al. (2008), Helpman et al. (2010), Coçsar et al. (2011)), efficiency wages or fair wage constraints (Egger and Kreickemeier (2009), Davis and Harrigan (2011), Amiti and Davis (2011)), in which ex-ante identical workers receive higher wages in more productive firms and wages are systematically related to the export status of the firm.

However, Helpman et al. (2012) also report that there is an equally sizable component of residual wage inequality (34%) that none of these models can elucidate, namely, within-firm wage dispersion. This evidence is corroborated in recent empirical studies in the United States and several European countries, collected in Lazear and Shaw (2008). Overall, they report that within-firm wage variation ranges from 60 to 80 percent of the total wage dispersion in each of those countries. In a study of Mexican plants, Frías et al. (2012) find that an exogenous increase in the incentive to export, triggered by the peso devaluation in 1994, resulted in higher within-plant wage dispersion.

The purpose of this paper is to develop a theoretical framework to study this important and relatively unexplored dimension of wage inequality, emphasizing its links to international trade. To do so, I build on a standard two-country, general equilibrium model with heterogeneous firms, by adding two key ingredients. First, within-firm wage dispersion between identical workers arises as firms optimally respond to moral hazard, paying for performance in order to align the incentives of employees with their best interest. In particular, I study an environment in which workers can make mistakes that degrade product quality. Worker performance depends on exerting costly effort and thus a firm that wishes to improve product quality must offer adequate compensation to its employees for incurring these costs. However, effort is not contractible and firms can only relate compensation to performance, a noisy signal of effort, to alleviate the moral hazard problem. Second, I introduce cross-firm differences in optimal compensation policies by assuming that firms with high labor productivity have a comparative advantage in the production of quality. Each firm designs its optimal (log-linear) contract, providing incentives to attain a desired effort level. Because high productivity firms have a comparative advantage in producing

[2] Evidence of size and exporter wage premia is reported in Bernard and Jensen (1995), Amiti and Davis (2011) and Helpman et al. (2012) for US, Indonesian and Brazilian firms, respectively.

quality, they find it optimal to offer higher-powered incentives to their workers to attain higher product quality.[3] This implies that, in equilibrium, wages in more productive firms are relatively more unequally distributed.

Differences in performance-pay policies across firms generate implications for (residual) wage inequality that have not been captured in the literature on trade and inequality.[4] To illustrate these, consider the variance of wages in any one of the two countries in the model, denoted $Var(w)$, which can be decomposed as

$$Var(w) = Var\left[E(w/\theta)\right] + E\left[Var(w/\theta)\right]$$

where θ indexes the set of active firms in a given equilibrium. $E(w/\theta)$ and $Var(w/\theta)$ denote the mean and variance of wages across workers employed in firms with productivity θ, respectively. Total wage variance is thus the sum of (i) the variance of average wages across firms (*between*-firm inequality) and (ii) the average of within-firm wage variances (*within*-firm inequality). Recent theoretical studies link trade liberalization to residual wage inequality by proposing mechanisms that operate exclusively through the between-firm component of wage inequality, in which firms of different sizes pay different average wages but there is no wage dispersion inside firms. The model developed in this paper is, to the best of my knowledge, the first to link trade and residual wage inequality through both channels.

More specifically, the key features generating the effect of international trade on wage inequality in the model are the following:

(a) Performance pay generates within-firm wage inequality. By punishing or rewarding employees according to their performance, high-powered incentives amplify the effect of the idiosyncratic component of performance on wages.

(b) Different firms adopt different performance-pay policies and this leads to variation in both within-firm inequality and average wages across firms. Since more productive firms obtain higher returns from the effort of their employees, they optimally choose to offer higher-powered incentives. The equilibrium pattern of contracting strategies across firms thus implies that within-firm wage dispersion increases in firm productivity. Therefore, high productivity firms are also high wage dispersion firms in the model, thus generating variation in $Var(w/g)$ across firms. Because equilibrium requires workers to be indifferent between employment in any firm, high productivity firms also offer higher expected wages to compensate for higher effort levels, thus generating variation in $E(w/g)$ across firms.

(c) International trade liberalization triggers general equilibrium reallocations of labor towards high productivity firms, shaping the equilibrium firm productivity distribution and therefore contributing to wage inequality.

[3] This pattern is consistent with firm-level evidence in Bloom and Van Reenen (2007), who report a positive correlation between the extent to which firms reward performance and total sales in the United States, France, Germany, and the United Kingdom.

[4] Workers are identical, except for their ex-post income. Thus, from an empirical perspective, wage variation generated by the model should be understood as residual (or within-group) inequality (i.e. wage variation across workers of identical observable characteristics such as education, gender, experience, etc).

The focus on performance pay is also appealing from an empirical perspective. Lemieux et al. (2009) have documented an increasing prevalence of performance-pay compensation over time and shown how this trend can account for a significant share of the growth in wage inequality in the U.S.[5]

There are a number of studies in which within-firm wage dispersion is driven by workforce composition, such as Verhoogen (2008), Bustos (2011), Harrigan and Reshef (2011), Monte (2011), Burstein and Vogel (2012) and Caliendo and Rossi-Hansberg (2012).[6] In these models, workers are heterogeneous due to differences in ability or human capital, thus they can explain variation in skill premia, as opposed to (residual) wage dispersion within or between firms among identical workers. In addition, wages are determined in competitive labor markets and thus do not contain either firm- or match-specific components.

The outline of the paper is the following. The next section introduces the theoretical framework, sequentially describing individual preferences, entry, production technologies and the structure of the labor market. Section 3 studies the moral hazard problem, firms' optimal performance-pay policies and profit maximization. Section 4 analyzes the general equilibrium of the model, under free entry and trade balance conditions. Section 5 studies how trade liberalization affects the distribution of firm productivity, how labor is reallocated across firms and the implications of the theory for wage inequality between- and within-firms. The final section discusses extensions and topics for future versions of this paper.

2 Model Setup

There are two countries, Home and Foreign. To focus squarely on within-industry residual wage dispersion, I assume that each country is populated by identical workers that consume a single (differentiated) good. In addition, both countries are identical in terms of market structure and access to technology, although the size of their labor forces may differ. I thus focus on the description of the Home economy and use an asterisk to denote foreign variables.

2.1 Demand

In Home there is a continuum of identical risk-neutral workers of mass L. Individual preferences depend on the consumption of a differentiated product X_i

[5] In particular, using data from the PSID, Lemieux et al. (2009) show that the fraction of U.S. male workers on performance-pay jobs (i.e. workers earning piece rates, commissions, or bonuses) increased from about 30 percent in the late 1970s to over 40 percent in the late 1990s. They also show that wages are less equally distributed on performance-pay than non performance-pay jobs and conclude that the growth of performance-pay has contributed to about 25 percent of the increase in the variance of log wages between the late 1970s and the early 1990s.

[6] In Yeaple (2005) and Sampson (2012), differences in workforce composition generate only between-firm wage inequality, since firms hire workers of a single type.

and on the level of effort μ_i exerted at work:

$$U(X_i, \mu_i) = \frac{X_i}{k(\mu_i)},$$

where i indexes individual workers and $k(\mu_i) \equiv \mu_i^\delta$, $\delta > 1$, is a strictly increasing and strictly convex cost-of-effort function.[7] The consumption of the differentiated product is an index of the consumption of a continuum of horizontally and vertically differentiated varieties, defined as

$$X_i \equiv \left[\int_{j \in J} (q(j) x_i(j))^{\frac{\nu-1}{\nu}} \, dj \right]^{\frac{\nu}{\nu-1}},$$

where j indexes varieties, J is the set of varieties available in the market, $x_i(j)$ and $q(j)$ denote the consumption and quality of variety j, respectively, and $\nu > 1$ is the elasticity of substitution across varieties. The quality-adjusted price index dual to X_i is denoted by P and depends on the prices $p(j)$ of individual varieties.[8] For a worker earning a wage w_i, the familiar two-stage budgeting solution yields $PX_i = w_i$ and individual demand $x_i(j) = w_i q(j)^{\nu-1} p(j)^{-\nu} / P^{1-\nu}$.

Other than for consumption purposes, the differentiated product X is also demanded as an intermediate input. In particular, as described below, firms invest units of the differentiated product to set up their production process and export activities (fixed costs). These activities are assumed to use the output of each variety in the same way as is demanded by final consumers. Therefore, denoting total expenditure on the differentiated good by E, the aggregate demand for variety j, denoted $x(j)$, has a constant price-elasticity and can be written as

$$x(j) = q(j)^{\nu-1} \frac{p(j)^{-\nu}}{P^{1-\nu}} E.$$

The equilibrium revenue of producer j, denoted $r(j)$, equals aggregate expenditure on variety j. Therefore,

$$r(j) = p(j) x(j) = A q(j)^\rho x(j)^\rho, \tag{1}$$

where $A \equiv P^{1-\nu} E^{\frac{1}{\nu}}$, $\rho \equiv (\nu - 1)/\nu$ and $0 < \rho < 1$.

For expositional purposes, it is convenient to simplify notation by setting the aggregate consumption index in the Home country to be the numeraire ($P = 1$) and express utility as a function of income and effort levels

$$U(w_i, \mu_i) = \frac{w_i}{k(\mu_i)}. \tag{2}$$

[7] The elasticity of the utility function with respect to the cost-of-effort is set to one, without loss of generality. An alternative specification of the utility function, often used in applied contract theory, is to assume separability in X_i and μ_i. Separability has the convenient property of eliminating income effects on the marginal cost of effort. On the other hand, the specification used in the paper is analytically more tractable when the moral hazard problem is embedded in general equilibrium.

[8] Specifically, $P \equiv \left[\int_{j \in J} \left(\frac{p(j)}{q(j)} \right)^{1-\nu} dj \right]^{\frac{1}{1-\nu}}$.

2.2 The Product Market

The product market is modelled in the same way as Melitz (2003). There is a competitive fringe of risk neutral firms that can choose to enter the differentiated sector and become a monopolistic producer of a variety of good X 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 $\theta \geq \theta_{\min}$ and $z > 1$. The Pareto distribution is not only tractable, but together with other assumptions in the model, implies a Pareto firm-size distribution which typically provides a reasonable approximation to observed data (Axtell (2001)). Since in equilibrium all firms with the same productivity behave symmetrically, I index firms and varieties 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.

The production technology of every variety is summarized by two functions, one describing the production of physical units and the other describing the production of quality.[9] Physical output of each variety (y) depends on the productivity of the firm and the number of employees hired (h):

$$y(\theta) = \theta h. \tag{3}$$

In turn, product quality (q) is described by a function $q(\theta, c)$ that depends on firm productivity and team performance, denoted c. Team performance is defined as the average performance of the employees of the firm, $c \equiv h^{-1} \int_0^h c_i di$, where c_i is worker i's performance in her production task, $i \in [0, h]$. The dependence of quality on c captures the notion that workers make mistakes during the production process that affect the final quality of the product. Although workers can reduce the impact (or frequency) of their mistakes by exerting effort, the latter is not the sole determinant of individual performance. In particular, I assume that the performance of worker i is a function of effort plus noise:

$$c_i = b(\mu_i) + \varepsilon_i, \tag{4}$$

where $b(\mu_i) \equiv \mu_i^{1-b}/(1-b)$, $b > 1$, is strictly increasing and concave in effort. Note that $b(\mu_i)$ is negative and can thus be interpreted as the expected number of mistakes for a worker whose effort is μ_i. The restriction $b > 1$ ensures that the elasticity of the expected number of mistakes with respect to effort is negative. In turn, ε_i is an i.i.d. draw from a cumulative distribution function $G_\varepsilon(\varepsilon)$ with mean zero and variance σ_ε^2 that captures unmodelled determinants of a worker's performance such as idiosyncratic skills, match-effects and variation in

[9] Endogenous quantity and quality choices of heterogeneous producers are also studied in Kugler and Verhoogen (2012).

the quality of inputs used in the production process. Under this specification, the Law of Large Numbers implies that the firm fully diversifies the impact of the idiosyncratic component of individual performance on output quality.[10] That is, equation (4) and the assumptions on ε_i imply that given effort levels μ_i for every i, then $c = E\left[b(\mu_i)\right]$ and the firm achieves product quality $q(\theta, E\left[b(\mu_i)\right])$.

The function $q(\theta, c)$ is assumed to be differentiable, concave, increasing in both of its arguments and strictly log-supermodular. The latter property implies that the marginal increase in product quality for a given increase in average performance is greater for high productivity firms.[11] Therefore, high productivity firms have a comparative advantage in producing high quality products. As we will see below, strict log-supermodularity of $q(\theta, c)$ is sufficient to characterize optimal quality choices across firms. However, to derive closed-form solutions I shall assume the following functional form for product quality:

$$q(\theta, c) = \left[\frac{1}{s}\theta^{s\lambda} + \frac{1}{w}\left(\frac{c_m}{c}\right)^{w\lambda}\right]^{\frac{1}{\lambda}}, \tag{5}$$

where c_m is the expected performance of a team in which every worker exerts the minimum effort; that is, $c_m \equiv \mu_m^{1-b}/(1-b)$. This specification ensures that quality is always positive. The parameter λ in equation (5) controls the degree of complementarity between productivity and average worker performance. I assume that $\lambda < 0$, which implies log-supermodularity. Setting $\lambda < 0$ also implies that quality is increasing and concave in team performance. The parameter $s \geq 0$ reflects the scope for quality differentiation. A higher s gives more productive firms a relatively greater incentive to produce high-quality outputs.

2.3 Labor Market

The key feature of the labor market in this model is the existence of moral hazard. Individual effort μ_i is unobservable to firms (or, more generally, noncontractible), thus they respond by designing wage contracts that tie compensation to individual performance c_i.[12]

More specifically, I assume that a firm and a worker can only write log-linear contracts of the form

$$\log w_i = \alpha + \beta c_i, \tag{6}$$

[10] This result relies on the assumption that firms hire a continuum of workers, which allows the application of the LLN. An advantage of this setup is that firm-level variables such as quantity, quality, employment and prices are non-stochastic, allowing the model to remain tractable.

[11] Because $q(\theta, c)$ is assumed to be differentiable, strict log-supermodularity is equivalent to a positive cross-partial derivative, $\frac{\partial^2 \log q}{\partial\theta\partial c} > 0$.

[12] Although potentially relevant to study within-firm wage variation, this paper does not deal with any form of group-based compensation schemes. The emphasis on individual incentives can be motivated empirically. Lazear and Shaw (2007) report that the share of large US firms that have more than 20 percent of their workforce working with some form of individual incentives, like a performance bonus, has grown from 38 percent in 1987 to 67 percent in 1999. The comparable share of firms using any form of 'gain-sharing' or group-based incentives was 7 percent in 1987 and 24 percent in 1999.

where w_i is the wage and (α, β) define the fixed and performance-related components of compensation. Importantly, this contract structure ensures that wages are positive for any realization of c_i. Because wages fuel the demand side of the model, this is an essential property for embedding the moral hazard problem in general equilibrium.

Each firm designs wage contracts to induce its employees to exert optimally chosen effort levels. However, because workers are homogeneous, a competitive equilibrium in the labor market requires that every contract offered by any firm should yield the same expected utility, denoted $\overline{u} > 0$.[13] Contracts yielding a lower expected utility than this outside option would fail to attract workers. Exceeding $\overline{u}$ would not be profit-maximizing. As shown in Section 4, the value of $\overline{u}$ is determined endogenously in the general equilibrium of the model.

3 The Firm's Problem

This section studies the profit maximization problem of firm θ, proceeding in several steps. The first step characterizes the link between individual effort supply and performance-pay incentives in the presence of moral hazard. The second step characterizes the cost-minimizing contracts that the firm designs to induce a given performance level from a set of h employees. The solution to this problem determines the cost function of product quality. The final step sets up the profit maximization problem, in which the firm determines employment, product quality and whether to export given demand in the domestic and foreign markets.

3.1 Performance Pay and Individual Effort

The combination of risk neutrality and log-linear contracts yields a simple setup of the moral hazard problem that allows a tractable characterization of the variation in the optimal provision of incentives across heterogeneous firms.[14] Qualitatively identical results to those in this section can also be derived under standard assumptions of separable utility and linear contracts. However, the current framework offers better tractability in the next section, where the moral hazard problem is embedded in general equilibrium.

By offering a contract (α, β), the firm indirectly presents worker $i \in [0, h]$ with a mapping linking individual effort to expected compensation, denoted $E[w_i(\mu_i)]$. The worker then chooses the effort level that maximizes the expected

[13] Positive wages and cost of effort together with equation (2) imply that, in equilibrium, $\overline{u} > 0$.

[14] Introducing a trade-off between risk insurance and incentives does not qualitatively alter the implications of this framework for wage inequality within and across firms. For example, both the current framework and the popular application of moral hazard with CARA preferences, normally distributed noise and linear contracts generate a first order condition that implies that the optimal effort level is increasing in the piece rate, see Bolton and Dewatripont (2005). A technical difficulty with the latter setup is that it generates negative wages.

value of (2),

$$\max_{\mu_i \geq \mu_{\min}} E\left[w_i(\mu_i)\right] k(\mu_i)^{-1}.$$

In an interior solution, the first-order condition for this problem is given by

$$E\left[w'(\mu_i)\right] = k'(\mu_i)\left[E\left[w(\mu_i)\right]/k(\mu_i)\right].$$

If the expected wage schedule shifts upwards, as is the case when the piece rate β increases, the term in squared brackets on the right-hand side generates an income effect that tends to reduce the optimal choice of effort, just as in the canonical neoclassical model of labor supply. This income effect is overcome by the substitution effect when the wage schedule is sufficiently concave. For consistency with numerous studies documenting performance gains from performance pay, I impose such condition for the case of log-linear contracts.[15] Using equations (4) and (6), $E\left[w_i(\mu_i)\right] = \exp\left[\alpha + \beta b(\mu_i)\right] E[\exp(\beta\varepsilon_i)]$. Under the assumed functional forms for $b(\cdot)$ and $k(\cdot)$, the optimal effort level satisfies

$$\mu_i = \left(\frac{\beta}{\delta}\right)^{\frac{1}{b-1}}, \quad b > 1 + \delta \tag{7}$$

The parameter configuration ensures that equation (7) characterizes the unique global maximizer of the worker's problem (see Appendix) and that individual effort is strictly increasing in the piece rate.[16]

3.2 Optimal Log-linear Contracts, Quality and Wages

The cost of attaining a given level of product quality $q(\theta, c)$ is determined by the cost of providing adequate incentives to guarantee a team performance c. There are infinitely many sets of log-linear contracts (α_i, β_i), $i \in [0, h]$, that the firm could write with its h employees in order achieve this goal. However, the optimal set of contracts minimizes the expected compensation subject to (i) inducing average performance c, (ii) incentive compatibility constraints and (iii) participation constraints:[17]

$$\min_{\{\alpha_i, \beta_i, \mu_i\}_{i=0}^{h}} h^{-1} \int_0^h E\left[w_i(\mu_i)\right] di$$

$$\begin{aligned} s.t \quad & (i)\ c = h^{-1} \textstyle\int_0^h b(\mu_i) di \\ & (ii)\ E\left[w'(\mu_i)\right] = k'(\mu_i)\left[E\left[w(\mu_i)\right]/k(\mu_i)\right] \\ & (iii)\ E\left[w_i(\mu_i)\right] k(\mu_i)^{-1} \geq \overline{u} \end{aligned}$$

[15] See, for example, Parent (1999), Lazear (2000) and references cited in Lazear and Shaw (2007).

[16] This parameter configuration also ensures that high productivity firms choose higher team performance in Section 3.3 (see Appendix).

[17] The 'first-order approach', which involves replacing the incentive compatibility constraint with the first-order condition of the worker's problem, is valid in the current framework because equation (7) identifies a unique global maximizer (see Appendix).

For the case of log-linear contracts and the assumed functional forms for $b(\cdot)$ and $k(\cdot)$, the previous problem can be written as:

$$\min_{\{\alpha_i,\beta_i,\mu_i\}_{i=0}^{h}} h^{-1}\int_0^h e^{\alpha_i+\beta_i\frac{\mu_i^{1-b}}{1-b}} E\left[e^{\beta_i\varepsilon_i}\right] di \tag{8}$$

$$\begin{aligned} s.t \quad & (i)\ c=[h(1-b)]^{-1}\int_0^h \mu_i^{1-b} di \\ & (ii)\ \beta_i = \delta\mu_i^{b-1} \\ & (iii)\ \mu_i^{-\delta} e^{\alpha_i+\beta_i\frac{\mu_i^{1-b}}{1-b}} E\left[e^{\beta_i\varepsilon_i}\right] \geq \overline{u} \end{aligned}$$

The following proposition characterizes the solution to this problem.

Proposition 1 (Cost-minimizing contracts) *Consider the effort level μ that satisfies $c=b(\mu)$; that is $\mu=[c(1-b)]^{1/(1-b)}$. Then, the solution to problem (8), denoted $(\alpha_i^*,\beta_i^*,\mu_i^*)$, is:*

(a) Effort: $\mu_i^*=\mu$ for all $i\in[0,h]$.

(b) Piece rate: $\beta_i^*=\delta\mu^{b-1}$ for all $i\in[0,h]$.

(c) Fixed compensation: $\alpha_i^*=\ln[\overline{u}k(\mu)]+\delta/(b-1)$ for all $i\in[0,h]$.

Proof. Appendix. ■

In principle, the firm could choose to offer different incentives to different workers. However, this is not cost-effective. The symmetry of optimal effort levels -part (a)- follows from the convexity of the effort cost function $k(\cdot)$. Intuitively, convexity implies that the cost of compensating a worker for a higher than average effort exceeds the cost reduction of inducing another worker to exert a lower than average effort level. Regarding parts (b) and (c), note that the firm offers incentive-compatible contracts and sets the fixed component of compensation to ensure that the participation constraint is satisfied with equality. Therefore, conditional on the effort level determined in part (a), the incentive and participation constraints pin down the contract.[18]

Individual wages depend on the compensation policy of the firm, defined in parts (b) and (c) of Proposition (1). Moreover, because noise is idiosyncratic with mean zero, the minimized value of the objective function in problem (8) is equal to the average wage that the firm pays for a team performance $c=b(\mu)$, denoted $\omega(\mu)$. This performance level allows a firm with productivity θ to achieve product quality $q(\theta,b(\mu))$. Alternatively, to achieve quality q, a firm with productivity θ needs to induce every employee to exert an effort level that satisfies $q=q(\theta,b(\mu))$, denoted $\mu(\theta,q)$. The average wage associated to such contract, $\omega(\mu(\theta,q))$, thus measures the per-employee cost of producing quality q in firm θ, denoted $c(\theta,q)$, which is henceforth referred to simply as the quality cost function of firm θ. These results are summarized in the following corollary.

[18] Note that $c<0$ and $b>1$ ensure strictly positive effort levels in part (a).

Corollary 2 (Cost of quality and wages) *Let $\omega(\mu)$ denote the average wage that firms pay to induce an average employee effort equal to μ and let $c(\theta, q)$ denote the per-employee cost of product quality q in firm θ. Then:*

(a) The average wage is equal to the minimized value of the objective function in problem (8),

$$\omega(\mu) = \overline{u}k(\mu).$$

(b) The wage of worker i is

$$w_i(\mu) = \omega(\mu)\frac{e^{\delta\mu^{b-1}\varepsilon_i}}{E_\theta\left[e^{\delta\mu^{b-1}\varepsilon}\right]},$$

for all $i \in [0, h]$, where $E_\theta[\cdot]$ denotes the mathematical expectation across workers in firm θ.

(c) The quality cost function is

$$c(\theta, q) = \omega(\mu(\theta, q)),$$

where $\mu(\theta, q)$ is the effort level that each employee has to exert to attain quality q in firm θ, implicitly defined by $q = q(\theta, b(\mu))$.

Proof. Appendix. ■

Importantly, $b > 1$ implies that the standard deviation of (log) wages paid by the firm, $\delta\mu^{b-1}\sigma_\varepsilon$, is increasing in effort. Monotonicity and convexity of $k(\cdot)$ guarantee that the average wage is increasing and convex in effort μ. The next section endogeneizes the choice of effort which, together with Corollary (2), provide a mapping between wages and firm productivity that can be used to analyze wage variation between and within firms and the implications of international trade for wage inequality.

3.3 Profit Maximization

The linearity of the production function (3) implies that, given a choice of product quality q, the marginal cost of physical output y is constant and equal to $c(\theta, q)/\theta$ for a firm with productivity θ. Together with the assumption that firms can price- and quality-discriminate between domestic and foreign buyers, this implies that the profit maximization problem of firm θ is separable and can thus be written as the sum of profits in the domestic and foreign markets.[19] As usual, conditional on the entry decision, the structure of the CES demand ensures that the firm finds it profitable to serve domestic consumers. In turn, the firm exports if and only if gross profits from foreign sales exceed the fixed cost of exporting.

[19] Allowing for quality discrimination, the firm can in principle choose to supply different product qualities in the home and foreign markets. If so, workers allocated to different 'production lines' will earn different expected wages. Note that, in equilibrium, workers are indifferent between employment in either production line because every contract generates the same expected utility.

As shown in the Appendix, the formulation of the profit maximization problem can be simplified by using two properties of the solution. First, the optimal choices of product quality in the domestic and foreign markets, denoted $q_d(\theta)$ and $q_x(\theta)$, respectively, coincide and therefore $q_d(\theta) = q_x(\theta)$. There is no product quality upgrading or downgrading associated to exporting in this model.[20] Intuitively, the firm can increase revenue in a given market by either expanding output or quality. Optimality requires that choices of output and quality in each market satisfy the equality of relative marginal revenue and relative marginal cost. According to (1), relative marginal revenue of output is q_m/y_m in market $m = \{d, x\}$. In turn, the relative marginal cost of output is given by $c(\theta, q_m) / (c_q(\theta, q_m) y_m)$. Note that transport costs increase the marginal costs of output and quality proportionally in market x, thus they do not distort the relative marginal cost of output across markets. Therefore,

$$c_q(\theta, q_m) = \frac{c(\theta, q_m)}{q_m}, \text{ for } m = \{d, x\}.$$

This leads to an identical choice of quality in each market, denoted $q(\theta)$.[21]

Second, for an exporting firm the optimal allocation of total output, denoted $y(\theta)$, between the domestic and foreign markets, denoted $y_d(\theta)$ and $y_x(\theta)$, respectively, satisfies the standard condition of equal marginal revenues in the two markets. From (1), this requires $[y_x(\theta)/y_d(\theta)]^{1-\rho} = \tau^{-\rho}(A^*/A)$, which implies that firm revenue can be written as a function of total output and product quality:

$$r(\theta) \equiv r_d(\theta) + r_x(\theta) = Aq(\theta)^\rho y(\theta)^\rho \Upsilon(\theta)^{1-\rho}, \tag{9}$$

The variable $\Upsilon(\theta)$ is a measure of foreign market access of firm θ that decreases in the variable trade cost τ. As in Helpman et al. (2010), $\Upsilon(\theta) \equiv 1 + I_x(\theta) \left[\tau^{-\rho}(A^*/A)\right]^{1/(1-\rho)}$.

The firm's problem can thus be formulated as choices of total output y, team effort μ and export decision I_x that solve

$$\pi(\theta) \equiv \max_{\substack{y \geq 0, \mu \geq \mu_{\min}, \\ I_x \in [0,1]}} \left\{ Aq(\theta, b(\mu))^{\frac{\rho}{\lambda}} y^\rho \left[1 + I_x \tau^{-\frac{\rho}{1-\rho}} \left(\frac{A^*}{A}\right)^{\frac{1}{1-\rho}}\right]^{1-\rho} - \frac{\omega(\mu)}{\theta} y - f_d - I_x f_x \right\},$$

where the indicator $I_x(\theta)$ equals 1 if firm θ exports and 0 otherwise.

[20] Quality upgrading induced by exporting can be easily introduced to the model by assuming that foreign consumers trade off quality and quantity differently than domestic consumers. For example, if $X_i^* = \left[\int_{j \in J^*} \left(q^*(j)^\chi x_i^*(j)\right)^{\frac{\nu-1}{\nu}} dj\right]^{\frac{\nu}{\nu-1}}$, and $\chi > 1$. Alternatively, letting $\chi < 1$ would lead exporters to downgrade quality. This extension is left for future versions of the paper.

[21] Since $c(\theta, q)$ is strictly convex in q, for $q \geq 0$, q_m is unique and thus $q(\theta) = q_d(\theta) = q_x(\theta)$ Geometrically, the marginal and average costs of quality intersect at $q(\theta)$. Therefore, $q(\theta)$ minimizes the average cost of quality in firm θ.

The existence of a fixed production cost implies that there is a zero-profit cutoff θ_d such that firms drawing a productivity $\theta < \theta_d$ exit without producing. Similarly, the existence of a fixed exporting cost implies that there is an exporting cutoff θ_x such that firms drawing a productivity $\theta < \theta_x$ do not find it profitable to serve the export market. For consistency with a large empirical literature that finds evidence of self-selection of the more efficient firms into the export market, I focus on values of trade costs for which $\theta_{\min} < \theta_d < \theta_x$.[22] This implies that the firm market access variable can be written as

$$\Upsilon(\theta) \begin{cases} \Upsilon_x & \text{if} \quad \theta \geq \theta_x, \\ 1 & \text{if} \quad \theta < \theta_x, \end{cases}$$

where $\Upsilon_x \equiv 1 + \tau^{-\frac{\rho}{1-\rho}} (A^*/A)^{\frac{1}{1-\rho}} > 1$.

For any choice of quality, the first-order condition for total output requires that the marginal revenue of output be equal to the constant marginal production cost. With CES demand, this implies that the variable cost equals a constant fraction of firm revenue, as in equation (10) below. Similarly, the optimal choice of team effort weighs the marginal revenue generated by improved quality against the marginal increase in compensation. In an interior solution, $\mu(\theta) > \mu_{\min}$, dividing the first-order conditions for output and quality yields an equality of relative marginal revenues and relative marginal costs. As equation (11) shows, this implies that the optimal choice of team effort is attained when the percentage increase in quality induced by a marginal increase in effort is equal to the percentage increase in compensation. As a result,

$$\rho r(\theta) = \frac{\omega(\mu(\theta))}{\theta} y(\theta), \tag{10}$$

$$\frac{q_c\left(\theta, b(\mu(\theta))\right) b'(\mu(\theta))}{q\left(\theta, b(\mu(\theta))\right)} = \frac{\omega'(\mu(\theta))}{\omega(\mu(\theta))}. \tag{11}$$

Note that equation (11) depends on a single unknown, $\mu(\theta)$. The assumed functional forms for product quality and team performance allow a closed-form solution for optimal team effort. Because product quality is log-supermodular in productivity and team performance, only firms with productivity above a cutoff θ_μ find it profitable to induce a team effort higher than the minimum $\mu_{\min}$. I assume that θ_m is sufficiently high, so that $\theta_d \geq \theta_\mu$ in equilibrium.[23] Therefore:

$$\mu(\theta) = \mu_{\min} \kappa_\mu \theta^{s/(b-1)}, \tag{12}$$

[22] In general, the exact condition that yields $\theta_d < \theta_x$ depends on both trade costs and relative demand shifters (see equation (23)). In the case of symmetric countries, the condition is $(f_x/f_d)\tau^{\frac{\rho}{1-\rho}} > 1$, as in Melitz (2003).

[23] In the case $\theta_d < \theta_\mu$, product quality and thus firm revenue cease to be power functions of firm productivity when $\theta_d \leq \theta < \theta_\mu$, which precludes a closed-form analysis of the general equilibrium. However, the model has a similar structure to the case developed in the main text. For example, quality, output and revenue increase in firm productivity. The main difference is that optimal contracts do not vary across firms with productivity $\theta \in [\theta_d, \theta_\mu]$. Therefore, firms in this range pay the same average wage (i.e. no between-firm inequality) and exhibit the same degree of within-firm wage dispersion.

where $\theta_\mu = (\kappa_\mu)^{(1-b)/s}$. Constants κ_μ, κ_q, κ_ω, κ_r, κ_y (introduced below) are positive and defined in the Appendix.

Team effort determines optimal product quality, denoted (with a slight abuse of notation) $q(\theta) \equiv q(\theta, b(\mu(\theta)))$, and the average wage $\omega(\theta)$ according to Corollary (2):

$$\begin{aligned} q(\theta) &= \kappa_q \theta^s, \\ \omega(\theta) &= \kappa_\omega \overline{u} \theta^{ds/(b-1)}. \end{aligned}$$

Product quality increases in productivity, with elasticity equal to s. High productivity firms thus pay higher average wages to compensate their employees for the disutility of effort associated to the production of quality.

From the first-order condition for output (10), the expression for firm revenue (9) and the solution for the average wage $\omega(\theta)$, I solve for revenue and total output as functions of the demand shifters A and the reservation utility $\overline{u}$. Total employment, $h(\theta)$, follows from the production function (3). Therefore:

$$\begin{aligned} r(\theta) &= \kappa_r \Upsilon(\theta) \left(A\overline{u}^{-\rho}\right)^{1/(1-\rho)} \theta^{\Gamma}, && (13) \\ y(\theta) &= \kappa_y \Upsilon(\theta) \left(A\overline{u}^{-1}\right)^{1/(1-\rho)} \theta^{\Gamma - s(1-\rho)}, && (14) \\ h(\theta) &= \kappa_y \Upsilon(\theta) \left(A\overline{u}^{-1}\right)^{1/(1-\rho)} \theta^{\Gamma - 1 - s}, && (15) \end{aligned}$$

where $\Gamma \equiv [1 + s - ds/(b-1)]/(1-\rho)$. The condition $b \geq 1 + d/\rho$ ensures that $\Gamma > 1 + s$, so that revenue, output and employment increase in productivity for every $s \geq 0$. Note that, as usual in models with a fixed exporting cost and selection into export markets, firm revenue, output and employment increase discontinuously at the exporting cutoff as the marginal exporter incurs f_x. This is not the case for quality, team effort and average wage, since there is no motif for quality upgrading (or downgrading) associated to exporting in this model (see footnote in page 11). Another standard but useful implication of expression (13) is that the ratio of any two firms' revenues depends only on their relative productivities and relative market access.

Finally, the first-order condition (10) also implies that firm profits can be written as a function of revenue and the fixed costs,

$$\pi(\theta) = (1-\rho) r(\theta) - f_d - I_x(\theta) f_x. \quad (16)$$

4 Equilibrium

The general equilibrium of the model shares a common structure with the extensive literature that builds on Melitz (2003). This section explains how to solve for the remaining endogenous variables in the model. Further details can be found in the Appendix.

The zero-profit cutoff θ_d is the productivity level that leaves firms indifferent between exiting and producing for the domestic market. In turn, the exporting cutoff θ_x leaves firms indifferent between exporting and producing exclusively for the domestic market. From the expressions for revenue (13) and profits (16), these two conditions require

$$\kappa_r(1-\rho)\left(A\overline{u}^{-\rho}\right)^{1/(1-\rho)}\theta_d^{\Gamma} = f_d \tag{17}$$

and

$$\kappa_r(1-\rho)\left(\Upsilon_x - 1\right)\left(A\overline{u}^{-\rho}\right)^{1/(1-\rho)}\theta_x^{\Gamma} = f_x, \tag{18}$$

respectively.

Free entry implies that the expected profits of successful entrants should equal the sunk entry cost; that is, $\int_{\theta_d}^{\infty}\pi(\theta)dG_\theta(\theta) = f_e$. Using the Pareto productivity assumption, the expression linking revenue to firm productivity (13) and the conditions characterizing the productivity cutoffs (17) and (18), the free entry condition can be written as

$$\frac{f_d}{(z/\Gamma-1)}\left(\frac{\theta_{\min}}{\theta_d}\right)^z\left[1+\left(\frac{f_x}{f_d}\right)\left(\frac{\theta_d}{\theta_x}\right)^z\right] = f_e. \tag{19}$$

For future reference, note that the ratio of productivity cutoffs θ_x/θ_d to be inversely related to the domestic cutoff θ_d. Since equation (19) does not depend directly on the transport cost, it follows that changes in τ induce θ_d and θ_x/θ_d to change in opposite directions.

Equations (17), (18), (19) and their Foreign counterparts can be used to solve for the productivity cutoffs and demand shifters in Home and Foreign $(\theta_d, \theta_x, \theta_d^*, \theta_x^*, A, A^*)$ as functions of the reservation utilities $\overline{u}$ and $\overline{u}^*$. The demand shifters, in turn, determine firm market access variables $\Upsilon(\theta)$ and $\Upsilon^*(\theta)$.

The mass of firms and expenditure in Home and Foreign are determined by imposing market clearing and trade balance. First, note that total expenditure is proportional to the mass of firms in each country. This follows from the market clearing condition (1), which implies that aggregate expenditure on domestic varieties equals total revenues of domestic firms. In Home, this is written as

$$E = M\int_{\theta_d}^{\infty} r(\theta)dG_\theta(\theta), \tag{20}$$

where M denotes the mass of firms in Home. A similar equation applies in Foreign, linking E^* and M^*. Trade balance requires the equality of export sales of domestic and foreign firms. This is formally stated as

$$M\frac{\Upsilon_x-1}{\Upsilon_x}\int_{\theta_x}^{\infty} r(\theta)dG_\theta(\theta) = M^*\frac{\Upsilon_x^*-1}{\Upsilon_x^*}\int_{\theta_x^*}^{\infty} r^*(\theta)dG_\theta(\theta), \tag{21}$$

after using $r_x(\theta) = r(\theta)\left(\Upsilon_x - 1\right)/\Upsilon_x$ for $\theta \geq \theta_x$ and an analogous expression for export sales of foreign firms. Next, the definition of the demand shifter A and the choice of numeraire ($P = 1$) determine the expenditure in Home, $E = A^\nu$. Equation (20), its counterpart in Foreign and the trade balance condition (21) can then be used to solve for M, M^* and E^*. The price index in Foreign follows from $A^* = (P^*)^{1-\nu}(E^*)^{\frac{1}{\nu}}$.

Finally, the reservation utilities $\overline{u}$ and $\overline{u}^*$ are pinned-down by imposing labor market clearing in each country. In Home, this requires equating labor supply, L, and labor demand, $M \int_{\theta_d}^{\infty} h(\theta) dG_\theta(\theta)$. Substituting for firm employment using expression (15) and solving for $\overline{u}$ yields

$$\overline{u} = A \left[\frac{M}{L} \kappa_y \int_{\theta_d}^{\infty} \Upsilon(\theta) \theta^{\Gamma - 1 - s} dG_\theta(\theta) \right]^{1-\rho}. \tag{22}$$

In the same way, labor market clearing in Foreign yields $\overline{u}^*$.

5 Trade Liberalization, Selection and Inequality

This section begins by analyzing the impact of trade liberalization, modelled as a fall in the transport cost τ, on firm selection and labor reallocations across firms. This sets the stage for the analysis of wage inequality.

5.1 Firm Selection

There is substantial empirical evidence that episodes of trade liberalization shape the equilibrium distribution of firm productivity by inducing low productivity firms to exit and some firms to start exporting.[24] In the model, these findings are consistent with equilibria in which decreases in variable trade costs induce an increase in the domestic cutoff θ_d. Through the free-entry condition (19), a higher θ_d implies a lower ratio of productivity cutoffs θ_x/θ_d, increasing the fraction of exporting firms.[25] I will refer to the class of equilibria satisfying this property as equilibria in which trade liberalization leads to firm selection.

Definition 3 *An equilibrium exhibits* ***firm selection in response to trade liberalization*** *if a marginal fall in the transport cost increases the domestic cutoff* θ_d.

Besides its empirical relevance, this class of equilibria is of interest because, as I show below, the impact of trade liberalization on wage inequality can be sharply characterized in this setting. Under what conditions does trade liberalization lead to firm selection? As in other models in the trade literature, it is difficult to derive a necessary and sufficient condition under which this property holds. However, firm selection can be ensured in special cases that enhance the tractability of the equilibrium.

To see this, divide equation (18) by (17) to obtain

$$(\Upsilon_x - 1) \left(\frac{\theta_x}{\theta_d} \right)^{\Gamma} = \frac{f_x}{f_d}, \tag{23}$$

[24] See, for example, Pavcnik (2002), Trefler (2004) and Bustos (2011).

[25] With Pareto-distributed productivity, it is straighforward to verify that the fraction of exporting firms, given by $[1 - G_\theta(\theta_x)] / [1 - G_\theta(\theta_d)]$, is an increasing one-to one function of the ratio of productivity cutoffs θ_x/θ_d.

and recall the definition of the market access measure, $\Upsilon_x \equiv 1+\tau^{-\frac{\rho}{1-\rho}} (A^*/A)^{\frac{1}{1-\rho}}$. Since free entry implies that θ_x/θ_d and θ_d are inversely related, expression (23) implies that the equilibrium exhibits firm selection if and only if a fall in variable trade costs translates into higher market access Υ_x. This is evidently the case when countries are symmetric and thus $A^*/A = 1$, as stated in part (a) of Proposition (4) below. More generally, it is necessary and sufficient that the direct effect of τ on Υ_x (i.e. holding relative demand A^*/A constant) is not overturned by the equilibrium response of A^*/A. Part (b) of Proposition (4) gives a sufficient condition limiting the elasticity of relative demand A^*/A with respect to τ.

Proposition 4 *An equilibrium exhibits firm selection in response to trade liberalization if one of the following conditions hold:*
(a) Countries are of equal size, i.e. $L = L^*$.
(b) The reservation utilities $\overline{u}$ *and* $\overline{u}^*$ *are exogenously determined and satisfy*

$$T\left(f_x/f_d, \tau\right)^{-1} < \frac{\overline{u}^*}{\overline{u}} < T\left(f_x/f_d, \tau\right),$$

where $T\left(f_x/f_d, \tau\right) \equiv \frac{(f_x/f_d)^{1-2(1-\rho)z/\Gamma}+\tau^{2\rho z/\Gamma}}{2(f_x/f_d)^{1-(1-\rho)z/\Gamma}\tau^{\rho z/\Gamma}}$.

Proof. Appendix. ■

To the best of my knowledge, variants of the Melitz (2003) model which analytically characterize the effect of variable trade costs on the domestic cutoff typically rely on at least one of these two conditions.[26] These ensure that the equilibrium displays a block structure that allows the productivity cutoffs and demand shifters to be determined solely by equations (17), (18) and (19) in each country.

Condition (a) states that trade liberalization always induces firm selection in the case of symmetric countries. Asymmetry is allowed under condition (b), which is usually introduced in the literature by assuming the existence of a homogeneous good that is produced in every country under perfect competition and constant returns to labor. In this case, expected wages (and thus the reservation utility) are proportional to labor productivity in the homogeneous sector.

Bounds on the admissible degree of asymmetry, however, are defined by $T\left(f_x/f_d, \tau\right)$. The appendix shows (i) $T\left(f_x/f_d, \tau\right) > 1$ for finite values of τ and $f_x \geq f_d$ and (ii) T is increasing in both arguments.[27] These bounds are necessary because of the existence of a home market effect in the model.[28] Intuitively,

[26] For example, countries are symmetric in Melitz (2003). Helpman et al. (2010) derive closed-form solutions for θ_d only under symmetry or with an outside sector. Their analysis focuses on how changes in the fraction of exporting firms shape inequality. An exception is Demidova and Rodriguez-Clare (2011), who show that unilateral trade liberalization induces firm selection in the context of a small open economy variant of Melitz (2003).

[27] $f_x \geq f_d$ is a stardard assumption that ensures only the most productive firms export in equilibrium, in line with the extensive evidence of selection into exporting.

[28] Home market effects are a standard feature in models of monopolistic competition with costly trade, dating back to Krugman (1980). See Helpman and Krugman (1985), chapter 10, for an example in a model with both differentiated and homogeneous sectors.

when countries are asymmetric, a fall in transport costs induces the differentiated product industry to concentrate disproportionately in the country with the larger domestic market, i.e. the country with a higher reservation utility. If the demand asymmetry is sufficiently high, this effect may overturn the direct effect of transport costs on firm selection, thereby reducing the domestic cutoff θ_d. This effect becomes stronger with lower transport costs, which explains why the admissible degree of asymmetry is increasing in τ.

5.2 Labor Reallocations Across Firms

Firm selection in response to trade liberalization leads to shifts in the distribution of firm productivity that trigger reallocations of labor towards high productivity firms. This section formalizes this argument by first deriving the distribution of employment across firms and then establishing how it is affected by trade liberalization. Since optimal compensation policies differ across firms, labor reallocations have implications for the equilibrium distribution of wages in the economy which are studied in the next section.

The distribution of employment across firms, denoted $G_h(\theta)$, measures the fraction of workers employed in firms with productivity below θ,

$$G_h(\theta) = \frac{\int_{\theta_d}^{\theta} h(\theta')dG_\theta(\theta')}{\int_{\theta_d}^{\infty} h(\theta')dG_\theta(\theta')}.$$

Provided that firm productivity is not too dispersed (i.e. z is large enough), the integral in the denominator of this expression will converge. In this case, it is possible to use the solution for firm employment (15) and the Pareto productivity assumption to obtain

$$G_h(\theta) = \begin{cases} 1 - \frac{(\theta/\theta_d)^\Lambda + (\Upsilon_x - 1)(\theta_x/\theta_d)^\Lambda}{(\Upsilon_x - 1)(\theta_x/\theta_d)^\Lambda + 1} & \text{if} \quad \theta_d \leq \theta \leq \theta_x, \\ 1 - \frac{\Upsilon_x(\theta/\theta_d)^\Lambda}{(\Upsilon_x - 1)(\theta_x/\theta_d)^\Lambda + 1} & \text{if} \quad \theta_x \leq \theta, \end{cases}$$

where $\Lambda \equiv \Gamma - 1 - s - z$ and $z > 2 + s + \Gamma$.

An important property of the model is that the distribution of employment across firms is fully determined by the productivity cutoffs and three parameters, Λ, Γ and f_x/f_d. To check this, note that equation (23) implies that market access Υ_x can be written as $\Upsilon_x = 1 + (f_x/f_d)(\theta_x/\theta_d)^{-\Gamma}$. This property allows me to characterize changes in the distribution of employment in terms of changes in the productivity cutoffs across equilibria. To do this, let subscripts 0 and 1 denote outcomes corresponding to two equilibria of the model.

Proposition 5 *Consider any two equilibria indexed by* 0 *and* 1 *such that:*
(i) $\theta_{d,0} < \theta_{d,1}$,
(ii) $\frac{\theta_{x,0}}{\theta_{d,0}} > \frac{\theta_{x,1}}{\theta_{d,1}}$,
(iii) parameters Λ, Γ *and* f_x/f_d *are the same in both equilibria.*

Then the distribution of employment across firms in equilibrium 1 *first-order stochastically dominates the distribution of employment across firms in equilibrium* 0*. That is, for all* θ*,*

$$G_{h,1}(\theta) \leq G_{h,0}(\theta), \text{ with strict inequality for some } \theta.$$

Proof. Appendix. ■

This result allows a comparison of employment distributions across equilibria in which cutoffs satisfy conditions (i) and (ii). A special case of interest is the class of equilibria that exhibit firm selection as a response to trade liberalization. In any such equilibrium, a fall in variable trade costs induces low productivity firms to exit and results in a higher proportion of exporting firms, in line with conditions (i) and (ii) of Proposition 5. This yields the following result.

Corollary 6 *Consider any equilibrium that exhibits firm selection as a response to trade liberalization. Then the employment distribution that follows a trade liberalization first-order stochastically dominates the initial employment distribution.*

Corollary (6) provides a sharp characterization of labor reallocations towards high productivity firms following trade liberalization. In the next section, we exploit this result to study the impact of trade liberalization on wage inequality.

5.3 Wage Inequality

There are two sources of heterogeneity in individual wages, a firm-specific component θ and a worker-specific component ε_i. The distribution of wages in the economy (and thus measures of wage inequality) will therefore depend on the underlying distributions of firm productivity θ and idiosyncratic performance ε.

To formalize this point, combine the firm's optimal choice of effort (12) with parts (a) and (b) of Corollary (2), to obtain the wage of worker i employed in firm θ,

$$w(\theta, \varepsilon_i) = \overline{u}\eta_0\theta^{\delta s/(b-1)} \frac{e^{\eta_1 \varepsilon_i \theta^s}}{E_\theta \left[e^{\eta_1 \varepsilon \theta^s}\right]}. \tag{24}$$

where $\eta_0 \equiv (\mu_{\min}\kappa_\mu)^\delta$ and $\eta_1 \equiv \delta (\mu_{\min}\kappa_\mu)^{b-1}$ are positive constants. Next, let $\int_{\underline{\varepsilon}}^{\varepsilon(\theta,w)} dG_\varepsilon(\varepsilon)$ denote the fraction of employees in firm θ with wages lower than w, i.e. $\varepsilon(\theta, w)$ satisfies $w = w(\theta, \varepsilon(\theta, w))$. Then the wage distribution, denoted $G_w(w)$, is given by

$$G_w(w) = \int_{\theta_d}^{\infty} \int_{\underline{\varepsilon}}^{\varepsilon(\theta,w)} dG_\varepsilon(\varepsilon) dG_h(\theta). \tag{25}$$

The distribution of wages is therefore a mixture of the distributions of θ and ε. One approach to studying wage inequality in this model is to make distributional assumptions for these variables and construct inequality measures from the resulting wage distribution, computed according to (25). An important drawback of this approach is that, while the Pareto distribution is known to provide a reasonable approximation to empirical measures of firm performance, I am not aware of comparable evidence that would justify a distributional assumption for ε.[29]

Instead, I focus on a specific measure of inequality, the variance of log wages.[30] Besides its widespread application in empirical studies of wage inequality, this approach yields analytical results without the need to rely on a particular distributional assumption for ε.[31] In addition, unlike other popular measures of inequality such as the Gini coefficient and the 90-10 wage gap, the variance can be decomposed into between- and within-firm components. As discussed in the Introduction, this property allows me to highlight different channels through which international trade can have an impact on wage inequality. At the end of the section, I verify the robustness of the results using an alternative measure of inequality, the mean log deviation.

In the model, different firms select different compensation policies to reward their employees. This implies that within-firm wage distributions differ across firms and thus inequality measures will crucially depend on the equilibrium allocation of workers across firms. The variance of log wages depends on the employment distribution and just the mean and variance of the within-firm log wage distributions, denoted $w_M(\theta)$ and $w_V(\theta)$, respectively. Letting $\widetilde{w}(\theta, \varepsilon_i) = \log w(\theta, \varepsilon_i)$ and using the expression for individual wages (24) yields

$$w_M(\theta) = E_\theta\left[\widetilde{w}(\theta, \varepsilon_i)\right] = \kappa_M + \frac{\delta s}{b-1}\log\theta - \log E_\theta\left[e^{\eta_1 \varepsilon \theta^s}\right], \tag{26}$$

$$w_V(\theta) = Var_\theta\left[\widetilde{w}(\theta, \varepsilon_i)\right] = \left(\eta_1 \sigma_\varepsilon \theta^s\right)^2, \tag{27}$$

where κ_M is a constant term. Given the equilibrium employment distribution, $G_h(\theta)$, these expressions can be integrated across firms to obtain the standard decomposition of the total variance of log wages into between and within-firm components,

$$Var(\widetilde{w}(\theta, \varepsilon_i)) = Var\left[w_M(\theta)\right] + E\left[w_V(\theta)\right].$$

The between-firm component, $Var\left[w_M(\theta)\right]$, is equal to the variance of average log wages across firms. The within-firm component, $E\left[w_V(\theta)\right]$, is equal to the average within-firm variance. Henceforth, I will refer to this second component as the *residual* variance of log wages. This allows me to avoid confusion with the

[29] It is of course possible to select $G_\varepsilon(\cdot)$ with the goal of approximating empirical wage distributions, though the analysis of inequality based on this fitted wage distribution is unlikely to be tractable. For this reason, I do not further pursue this approach in this paper and leave it for future research.

[30] The logarithmic transformation ensures that this measure of inequality is invariant to proportional shifts in the wage distribution, e.g. changes in the reservation utility $\overline{u}$ in equation (24). That is, if Home's wage distribution in an initial equilibrium 0 is simply a scaled-up version of that in another equilibrium 1, then the variance of log wages is the same in both equilibria.

[31] For example, among recent empirical studies, Lemieux (2006), Helpman et al. (2012) and Card et al. (2012) use variance decompositions of log wages to analyze changes in inequality in the US, Brazil and Germany, respectively.

within-firm variances $w_V(\theta)$ and also to highlight implications of the analysis in this section for the empirical assessment of the impact of trade on inequality. In empirical studies such as Helpman et al. (2012), the between-firm component is the estimated variance of the firm-fixed effects in a regression of individual wages that also controls for observable worker characteristics. The within-firm component is the variance of the regression residuals.

As in the previous related literature, wage inequality across ex-ante identical workers in the model is partly driven by cross-firm variation in average wages, i.e. between-firm inequality. Earlier models have shown that this variation can be generated by search frictions, efficiency wages or fair wage considerations,[32] while in this model firms compensate their workers for exerting different effort levels.

Unlike other models in the literature, however, part of the wage variation arises from differences in within-firm variances across firms. As long as worker performance is only a noisy signal of effort, i.e. $\sigma_\varepsilon > 0$, firms deal with the moral hazard problem by paying for performance, which results in within-firm wage dispersion. Moreover, within-firm inequality varies across firms since high productivity firms offer higher-powered incentives that magnify the variance of log wages between their employees. Note that, for given σ_ε, $w_V(\theta)$ increases in firm productivity. Cross-firm variation in inequality is a necessary ingredient for trade liberalization to have an impact on inequality through the within-firm component. When combined with the labor reallocations towards high productivity firms that result from trade liberalization, this mechanism generates increasing residual wage inequality.

Next, I show that if the initial equilibrium exhibits firm selection in response to trade liberalization, then the change in the residual variance is necessarily positive. The change in the between-firm variance, however, cannot be signed without imposing more structure on the distribution of the idiosyncratic component of individual performance ε.

Formally, let subscripts 0 and 1 denote outcomes corresponding to equilibria before and after trade liberalization, respectively. Consider first the change in the residual variance, which can be written as

$$\begin{aligned}
\Delta E\left[w_V(\theta)\right] &= \int_{\theta_{\min}}^{\infty} w_V(\theta)\left[dG_{h,1}(\theta) - dG_{h,0}(\theta)\right], \\
&= \int_{\theta_{\min}}^{\infty} w_V'(\theta)\left[G_{h,0}(\theta) - G_{h,1}(\theta)\right]d\theta, \\
&> 0.
\end{aligned}$$

The first line uses the fact that, in any equilibrium of the model, the within-firm variance depends only on firm-productivity. The second line follows after integrating by parts. From equation (27), the within-firm variance increases in θ, thus $w_V'(\theta) > 0$. Moreover, if the initial equilibrium exhibits firm selection in response to trade liberalization, then $G_{h,0}(\theta) \geq G_{h,1}(\theta)$ for all θ, with strict

[32] See the discussion in the Introduction.

inequality for some θ. Intuitively, trade liberalization generates labor reallocations towards high inequality firms, and this results in an unambiguous increase in the residual variance of log wages.

In turn, the change in the between-firm variance is given by

$$\begin{aligned}\Delta Var\left[w_M(\theta)\right] &= \int_{\theta_{\min}}^{\infty}\left[w_M(\theta)\right]^2\left[dG_{h,1}(\theta)-dG_{h,0}(\theta)\right]-\Delta\left(\overline{w}^2\right), \\ &= 2\int_{\theta_{\min}}^{\infty}w_M'(\theta)w_M(\theta)\left[G_{h,0}(\theta)-G_{h,1}(\theta)\right]d\theta-\Delta\left(\overline{w}^2\right),\end{aligned}$$

where $\Delta\left(\overline{w}^2\right)\equiv\left(\overline{w}_1\right)^2-\left(\overline{w}_0\right)^2$ and $\overline{w}_q\equiv\int_{\theta_{\min}}^{\infty}w_M(\theta)dG_{h,q}(\theta)$ is the mean log wage in equilibrium $q=\{0,1\}$. As in the analysis of the residual variance, the second line is obtained after integrating by parts. However, the change in $Var\left[w_M(\theta)\right]$ cannot, in general, be signed. First, note from (26) that the mean log wage is not necessarily increasing in firm productivity.[33] Furthermore, labor reallocations towards high productivity firms also imply a rise in the mean log wage, $\overline{w}_1>\overline{w}_0$, that tends to reduce the between-firm variance in the aftermath of trade liberalization. I summarize these results in the following Proposition.

Proposition 7 *Trade liberalization leads to an increase in the residual variance of log wages if and only if the equilibrium exhibits firm selection as a response to trade liberalization. The change in the between-firm variance of log wages cannot, in general, be signed.*

Although the variance of log wages is a popular measure for inequality comparisons in applied work, it may conflict with the Lorenz criterion (Foster and Ok (1999)).[34] The latter, however, incorporates some principles that are generally regarded as fundamental to the theory of inequality measurement.[35] For this reason, I close this section by analyzing the impact of trade liberalization using a Lorenz-consistent measure, the mean log deviation (MLD). This index, introduced by Theil (1967), belongs to the class of generalized entropy measures and, as such, it can be decomposed into between and within components.[36] The definition and decomposition of the MLD are given by

$$\begin{aligned}MLD &\equiv E\left[\log\left(\frac{\omega(\theta)}{w(\theta,\varepsilon_i)}\right)\right] \\ &= \int_{\theta_{\min}}^{\infty}\log\left(\frac{\overline{w}}{\omega(\theta)}\right)dG_h(\theta)+\int_{\theta_{\min}}^{\infty}E_\theta\left[\log\left(\frac{\omega(\theta)}{w(\theta,\varepsilon_i)}\right)\right]dG_h(\theta)\end{aligned}$$

[33] Actually, it is possible to construct examples in which, when productivity is high enough, the mean log wage decreases in θ.

[34] The Lorenz criterion states that a distribution F is more unequal that distribution F' if and only if the Lorenz curve of F lies below the Lorenz curve of F' everywhere in the domain.

[35] Atkinson (1970) showed that this criterion is equivalent to second-order stochastic dominance when the two distributions have equal mean.

[36] Generalized entropy measures have several desirable properties. Cowell (2011), chapter 3, shows that an inequality measure belongs to this class if and only if it simultaneously satisfies the weak principle of transfers, decomposability, scale independence and the population principle.

The second equality states that the MLD of wages can be decomposed into the MLD of mean wages across firms (between-firm MLD) and the average MLD of wages within firms (residual MLD). The impact of trade liberalization on the MLD index can be evaluated using the expression for individual wages (24) and Corollary (6). The results are qualitatively identical to those obtained for the variance of log wages.

Proposition 8 *Trade liberalization leads to an increase in the residual MLD of wages if and only if the equilibrium exhibits firm selection as a response to trade liberalization. The change in the between-firm MLD of wages cannot, in general, be signed.*

Proof. Appendix. ■

6 Concluding Remarks

Evidence from firm-level studies consistently show that wage dispersion within firms is a major component of wage inequality in many countries. This paper is, to the best of my knowledge, the first in the literature to develop a theoretical framework to study the determinants of within-firm wage dispersion, its variation across firms and links to changes in international trade costs. Moreover, in light of the magnitude and growth in residual wage dispersion, the focus is on modelling within-firm wage inequality between identical workers.

Several key mechanisms in the model are consistent with different pieces of empirical evidence. The emphasis on performance pay is motivated by evidence showing that its prevalence has grown considerably in the last 30 years. Lemieux et al. (2009) find that, by the late 1990s, performance-pay jobs account for as much as 45% of the jobs of male workers in the United States. Cross-firm differences in performance-pay policies across firms are, in turn, consistent with evidence from the managerial economics literature. Bloom and Van Reenen (2007) report that large firms tend to rely on incentive pay more intensively than smaller firms. Moreover, the empirical results in Kugler and Verhoogen (2012) support the assumption that larger firms have a comparative advantage in producing high-quality goods. Although the hypothesis that quality depends on employee performance appears to be a natural assumption, I do not regard it as an essential part of the mechanism linking trade liberalization to wage inequality. An interesting topic for future work is to think about alternative settings that would lead high productivity firms to offer higher-powered incentives. Finally, evidence that trade liberalization induces market share reallocations towards high productivity firms is provided by Pavcnik (2002) and Trefler (2004), for Chile and Canada, respectively.

A common feature in related studies in the literature that is absent in this framework are exporter wage premia. In the model, conditional on productivity, exporting does not induce firms to pay higher wages. As mentioned, however, this feature can be easily incorporated into the model by assuming that foreign buyers have a relatively higher preference for quality than domestic consumers. This extension would also generate higher within-firm inequality in exporting firms, conditional on productivity, which is consistent with the empirical evidence reported in Frías et al. (2012). Importantly, the analysis shows that exporter wage premia are not necessary to obtain a clear-cut prediction of the impact of trade liberalization on within-firm wage dispersion. Introducing exporter wage premia would reinforce the main results of the paper.

There are a number of additional topics worth exploring in future versions of this draft. One of them is the impact of trade liberalization on ex-post welfare. On one hand, lower trade costs lead to lower consumption prices and higher expected wages. However, labor reallocations towards high productivity firms can potentially hurt unlucky workers who, despite high effort levels, end up receiving very low wages due to poor ex-post performance.

References

Amiti, M., Davis, D., 2011. Trade, firms, and wages: Theory and evidence. The Review of Economic Studies 79 (1), 1–36.

Atkinson, A., 1970. On the measurement of inequality. Journal of economic theory 2 (3), 244–263.

Axtell, R., 2001. Zipf distribution of u.s. firm sizes. Science 293 (5536), 1818–1820.

Bernard, A., Jensen, J., 1995. Exporters, jobs, and wages in us manufacturing: 1976-1987. Brookings Papers on Economic Activity: Microeconomics 1995, 67–119.

Bloom, N., Van Reenen, J., 2007. Measuring and explaining management practices across firms and countries. The Quarterly Journal of Economics 122 (4), 1351–1408.

Bolton, P., Dewatripont, M., 2005. Contract theory. The MIT Press.

Burstein, A., Vogel, J., 2012. Globalization, technology, and the skill premium. Columbia University, mimeograph.

Bustos, P., 2011. The impact of trade liberalization on skill upgrading: Evidence from argentina. Economics Working Papers.

Caliendo, L., Rossi-Hansberg, E., 2012. The impact of trade on organization and productivity. The Quarterly Journal of Economics 127 (3), 1393–1467.

Card, D., Heining, J., Kline, P., 2012. Workplace heterogeneity and the rise of west german wage inequality. NBER Working Paper 18522.

Coçsar, A., Guner, N., Tybout, J., 2011. Firm dynamics, job turnover, and wage distributions in an open economy. NBER Working Paper 16326.

Cowell, F., 2011. Measuring inequality. OUP Oxford.

Davidson, C., Matusz, S., Shevchenko, A., 2008. Globalization and firm level adjustment with imperfect labor markets. Journal of International Economics 75 (2), 295–309.

Davis, D., Harrigan, J., 2011. Good jobs, bad jobs, and trade liberalization? Journal of International Economics 84, 26–36.

Demidova, S., Rodriguez-Clare, A., 2011. The simple analytics of the melitz model in a small open economy. NBER Working Paper 17521.

Egger, H., Kreickemeier, U., 2009. Firm heterogeneity and the labor market effects of trade liberalization. International Economic Review 50 (1), 187–216.

Foster, J., Ok, E., 1999. Lorenz dominance and the variance of logarithms. Econometrica 67 (4), 901–907.

Frías, J. A., Kaplan, D. S., Verhoogen, E., 2012. Exports and within-plant wage distributions: Evidence from mexico. American Economic Review: Papers and Proceedings 102 (3), 435–440.

Goldberg, P., Pavcnik, N., 2007. Distributional effects of globalization in developing countries. Journal of Economic Literature 45, 39–82.

Harrigan, J., Reshef, A., 2011. Skill biased heterogeneous firms, trade liberalization, and the skill premium. NBER Working Paper 17604.

Helpman, E., Itskhoki, O., Muendler, M., Redding, S., 2012. Trade and inequality: From theory to estimation.

Helpman, E., Itskhoki, O., Redding, S., 2010. Inequality and unemployment in a global economy. Econometrica 78 (4), 1239–1283.

Helpman, E., Krugman, P., 1985. Market structure and foreign trade: Increasing returns, imperfect competition, and the international economy. MIT press.

Katz, L., Autor, D., 1999. Changes in the wage structure and earnings inequality. Handbook of labor economics 3, 1463–1555.

Krugman, P., 1980. Scale economies, product differentiation, and the pattern of trade. The American Economic Review 70 (5), 950–959.

Kugler, M., Verhoogen, E., 2012. Prices, plant size, and product quality. The Review of Economic Studies 79 (1), 307–339.

Lazear, E., 2000. Performance pay and productivity. The American Economic Review 90 (5), 1346–1361.

Lazear, E., Shaw, K., 2007. Personnel economics: The economist's view of human resources. The Journal of Economic Perspectives 21 (4), 91–114.

Lazear, E., Shaw, K., 2008. The structure of wages: an international comparison. University of Chicago Press.

Lemieux, T., 2006. Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill? The American economic review 96 (3), 461–498.

Lemieux, T., MacLeod, W., Parent, D., 2009. Performance pay and wage inequality. The Quarterly Journal of Economics 124 (1), 1–49.

Melitz, M., 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. Econometrica 71 (6).

Monte, F., 2011. Skill bias, trade, and wage dispersion. Journal of International Economics 83 (2), 202–218.

Parent, D., 1999. Methods of pay and earnings: A longitudinal analysis. Industrial and labor relations review 53 (1), 71–86.

Pavcnik, N., 2002. Trade liberalization, exit, and productivity improvements: Evidence from chilean plants. The Review of economic studies 69 (1), 245–276.

Sampson, T., 2012. Selection into trade and wage inequality.

Theil, H., 1967. Economics and information theory. North-Holland publishing company Amsterdam.

Trefler, D., 2004. The long and short of the canada-us free trade agreement. The American Economic Review 94 (4), 870–895.

Verhoogen, E., 2008. Trade, quality upgrading, and wage inequality in the mexican manufacturing sector. The Quarterly Journal of Economics 123 (2), 489–530.

Yeaple, S., 2005. A simple model of firm heterogeneity, international trade, and wages. Journal of international Economics 65 (1), 1–20.