Available online at www.sciencedirect.com

ScienceDirect

Journal of Economic Theory 172 (2017) 478–504

JOURNAL OF Economic Theory

www.elsevier.com/locate/jet

Performance pay, trade and inequality ☆

Germán Pupato

Ryerson University, Canada

Received 28 July 2016; final version received 15 September 2017; accepted 3 October 2017
Available online 14 October 2017

Abstract

This paper introduces moral hazard into a general equilibrium model with heterogeneous firms to study wage inequality between homogeneous workers. Optimal performance pay contracts yield non-degenerate wage distributions among co-workers, enabling the analysis of two conceptually distinct and empirically relevant dimensions of wage dispersion: between-firm and within-firm inequality. The latter remains virtually unexplored in the literature. As an application, I characterize analytically the impact of trade liberalization on within-firm inequality, highlighting a new channel through which international trade can contribute to residual wage dispersion. To motivate the theory, I show that the model is consistent with cross-firm empirical patterns in residual wage dispersion and performance pay using nationally representative, matched employer–employee data from Canada.
© 2017 Elsevier Inc. All rights reserved.

JEL classification: F16; J31; J33

Keywords: Performance pay; Within-firm wage inequality; Moral hazard; Firm heterogeneity; Trade liberalization

☆ This research was supported by the Social Sciences and Humanities Research Council of Canada, grant number 430-2015-1091. I am grateful to the Associate Editor, two anonymous referees, Alex Alvarez, Costas Arkolakis, Kyle Bagwell, Giuseppe Berlingieri, Bernardo Blum, Matilde Bombardini, Pierre Brochu, Lorenzo Caliendo, Eva Chalioti, Kunal Dasgupta, Svetlana Demidova, Pablo Fajgelbaum, Giovanni Gallipoli, Jason Garred, Elhanan Helpman, Michael Hoy, Esteban Jaimovich, Min Seong Kim, Thomas Lemieux, Ricardo López, Giovanni Maggi, Marc Melitz, Eduardo Morales, Humberto Moreira, Marcelo Moreira, Peter Neary, Andrés Rodríguez-Clare, John Romalis, Thomas Sampson, Mariano Somale, Derek Stacey, Daniel Trefler, Eric Verhoogen, Halis Yildiz, Haomiao Yu and participants in seminars and conferences for stimulating comments and suggestions. The usual disclaimer applies. Sara Rohany Tabatabai provided outstanding research assistance. Carmina Ng provided support with the Workplace and Employee Survey at the Toronto RDC.

E-mail address: gpupato@economics.ryerson.ca.

https://doi.org/10.1016/j.jet.2017.10.001
0022-0531/© 2017 Elsevier Inc. All rights reserved.

1. Introduction

Understanding the determinants of wage inequality is central to economics. Starting with the influential studies by Bound and Johnson (1992) and Katz and Murphy (1992) in the U.S., a large empirical literature has shown that residual wage inequality –i.e., wage dispersion within industries, occupations and skill groups– represents a major share of the level and growth of wage inequality in many countries. Economists have since become increasingly interested in understanding the role that individual firms play in shaping residual wages; e.g., Card et al. (2013), Bloom et al. (2016) and Helpman et al. (2017).

In this paper, I propose a theory of residual wage dispersion driven by differential intensity of performance pay across heterogeneous firms. Performance pay yields non-degenerate wage distributions among homogeneous co-workers, enabling the analysis of two conceptually distinct dimensions of residual wage dispersion: between-firm and within-firm inequality. Between-firm inequality –i.e., dispersion in firm-level average wages– has been extensively analyzed in recent empirical and theoretical studies, discussed below. Within-firm wage inequality –i.e., average firm-level wage dispersion–, however, remains a largely overlooked yet substantial component of residual wage dispersion. For example, using similar methodologies, Helpman et al. (2017) and Akerman et al. (2013) find that between-firm inequality accounts for 37% and 19% of the variance of residual log wages within sector-occupation cells in Brazil (1990) and Sweden (2001), respectively. In contrast, within-firm inequality accounts for 34% and 65%, respectively.[1]

This paper is, to the best of my knowledge, the first attempt in the literature to study within-firm wage inequality due to performance pay in a general equilibrium model with heterogeneous firms. The theory is tractable enough to enable an analytical decomposition of wage inequality and a comparative statics analysis of the within-firm component. As an application, I characterize analytically the impact of trade liberalization on within-firm inequality, highlighting a new channel through which international trade can contribute to residual wage dispersion. The key mechanism, however, is not application-specific and thus provides a useful tool to analyze alternative changes in the economic environment that trigger resource reallocations across firms, such as technical progress. To motivate the theory and its emphasis on performance pay, I show that the model is consistent with cross-sectional patterns in residual wage dispersion and performance pay using matched employer–employee data from Canada.

The theory extends a standard two-country, general equilibrium model with heterogeneous firms (Melitz, 2003), by adding two key ingredients. First, moral hazard, which generates wage dispersion among homogeneous co-workers as firms pay for performance to align the incentives of employees with their best interests. In particular, I study a sequential production process during which workers stochastically make mistakes that are detrimental to product quality. Workers can reduce the frequency of their mistakes by exerting non-contractible effort at each production task. Importantly, as the frequency of tasks increases, individual performance converges to a Brownian process. This feature of the model leads to a tractable characterization of optimal performance-pay contracts, a version of the seminal work of Holmström and Milgrom (1987) that I embed in general equilibrium.

Second, I introduce cross-firm differences in contracting strategies by allowing for complementarity between firm productivity and the performance of workers in generating product

[1] These findings are corroborated in recent empirical studies in the United States and nine European countries, collected in Lazear and Shaw (2008). Summarizing the evidence, they report that within-firm inequality ranges from 60% to 80% of the total wage dispersion in each of those countries (page 482).

quality. High productivity firms optimally offer higher-powered incentives. This implies that, in equilibrium, wages are relatively more dispersed in more productive firms, according to a rich class of inequality measures that includes the variance of log wages and all inequality measures that respect second-order stochastic dominance and scale independence. Moreover, because equilibrium in the labor market requires workers to be indifferent between employment in any two firms, high productivity firms also offer higher expected wages to compensate for higher effort levels, generating between-firm inequality. Remarkably, these patterns of inequality within and across firms are driven entirely by endogenous differences in contracting strategies since, conditional on effort, the stochastic component of worker performance is invariant across firms and workers.

In this environment, the impact of international trade liberalization on within-firm inequality depends on whether a reduction in bilateral variable trade costs induces general equilibrium reallocations of labor towards high or low inequality firms. As a preliminary step I show that, under relatively mild restrictions on the distribution of firm productivity,[2] the distribution of workers across firms in a post-liberalization equilibrium first-order stochastically dominates the corresponding pre-liberalization equilibrium distribution.[3] The assumption of complementarity between effort and productivity, moreover, implies that high productivity firms are also high inequality firms and hence a trade liberalization reallocates workers towards high inequality firms.[4]

As a corollary of this mechanism, I show that a bilateral trade liberalization leads to monotonic increases in within-firm inequality in both countries. The result holds for symmetric countries and for three additively decomposable inequality measures: the variance of log wages, the Theil index and the mean log deviation (MLD). Moreover, to the extent that a unilateral trade liberalization triggers firm selection (exit of the least productive firms), the result extends to equilibria in which countries have asymmetric parameterizations of labor endowments, trade costs, effort costs, technology and firm productivity distributions. In this case, within-firm inequality increases in the liberalizing country. As in Helpman et al. (2010) and Coşar et al. (2016), however, the effects on between-firm inequality are non-monotonic and difficult to characterize analytically, even in symmetric equilibria.

The empirical analysis relies on the 2003 Workplace and Employee Survey (WES), a matched employer–employee dataset from Canada. A unique feature of the WES is high quality information on performance pay for a nationally representative sample of workers. I find that within-firm wage inequality accounts for 45% of the total variance of residual log wages. Eliminating performance-pay jobs from the sample results in a 14% reduction in residual wage inequality, almost entirely explained by a 40% reduction in the within-firm component.

I document a number of stylized facts to provide empirical support for key cross-sectional features of the model. First, high wage firms are typically high inequality firms, which is consistent with the dynamic formulation of the moral hazard problem adopted in this paper, but not with other standard static formulations. Second, I document positive cross-firm correlations between size (employment or revenue), mean and inequality of performance pay, controlling

[2] The additional structure on the distribution of firm productivity is required to characterize changes in the employment distribution only when the pre-liberalization equilibrium is not autarkic.

[3] As discussed on page 494, this result is non-trivial and *not* implied by the standard form of labor reallocations established in Melitz (2003).

[4] Trade liberalization *will* have an impact on within-firm inequality in this model as long as the intensity of performance pay varies across firms. The complementarity assumption, however, is motivated by empirical evidence (discussed below) and allows me to unambiguously sign the impact.

for differences in observable workforce composition. Complementarity between effort and firm productivity in the model provides a simple rationale for these patterns.[5]

Outline of the paper. The next section provides a brief review of related research. Section 3 introduces the theoretical framework. Section 4 studies firms' optimal performance-pay contracts and profit maximization, embedding the moral hazard problem in a monopolistic competition model with heterogeneous firms. Section 5 analyzes the general equilibrium with two symmetric countries. Section 6 studies how trade liberalization affects labor reallocations and wage inequality. Section 7 presents the empirical evidence on residual wage dispersion and performance pay across firms in Canada. An online Appendix contains proofs, additional theoretical and empirical results, and an extended discussion of related research.

2. Related literature

The paper contributes to a literature that investigates the link between performance pay and wage inequality. The closest theoretical study is Raith (2003), who introduces moral hazard into an oligopoly model with free entry and shows that tougher product market competition typically induces firms to offer higher-powered incentives, increasing the dispersion of managerial compensation. In Raith (2003), however, firms are homogeneous and employ a single worker (manager) hence firm-level wage distributions are degenerate and identical across firms, and within-firm inequality is trivially nil.

Regarding empirical work in this literature, Lemieux et al. (2009) use data from the Panel Study of Income Dynamics and the National Longitudinal Survey of Youth to show that the fraction of U.S. male workers on performance-pay jobs increased from about 38% in the late 1970s to as much as 45% in the late 1990s. They conclude that the growth of performance pay during this period accounted for about 21% of the increase in the variance of log wages. Barth et al. (2012) use Norwegian matched employer–employee data to show that wage inequality among co-workers is higher in firms that adopt performance pay than in firms that offer fixed compensation, conditional on firm size.[6] Cuñat and Guadalupe (2009) find that import penetration leads to more incentive provision and higher firm-level wage dispersion among U.S. executives. The empirical analysis in section 7 complements these studies by quantifying the contribution of performance-pay jobs to within-firm inequality and documenting systematic variation in firm-level distributions of performance pay using nationally representative data from Canada.

In this paper, performance pay addresses an individual incentives problem in the tradition of Holmström (1979) and, in particular, Holmström and Milgrom (1987). In principle, however, firms may adopt performance pay for a variety of other reasons; chiefly among them, to provide incentives for teams (Holmström, 1982), to induce sorting (Lazear, 1986), and to retain employees (Oyer, 2004).[7] The model provides a tractable benchmark for future exploration of richer

[5] This assumption has two additional implications in the model that receive empirical support. First, Bloom and Van Reenen (2007) show that the prevalence of incentives (including performance pay) increases in firm size. Second, output quality increases in firm size. Kugler and Verhoogen (2012) and Manova and Zhang (2012) provide empirical evidence that is consistent with this prediction.

[6] Barth et al. (2012) also present a partial equilibrium model with linear contracts to study a firm's choice of compensation scheme and its impact on the dispersion of wages of its employees. In contrast, this paper introduces firm heterogeneity and optimal contracts to characterize economy-wide within-firm wage inequality in general equilibrium.

[7] Disentangling these alternative theories using individual compensation data is an interesting yet challenging empirical problem that I will not attempt to solve in this paper. See Parent (1999), Lazear (2000) and Oyer and Schaefer (2005).

effects of performance pay in general equilibrium. Moreover, I conjecture that the key mechanisms in the model are likely to survive in a richer framework because performance-pay contracts generate wage dispersion among identical co-workers under mild conditions (see page 502).

Theories studying the impact of trade liberalization on residual wage dispersion in models with heterogeneous firms and homogeneous workers propose mechanisms that operate exclusively on the between-firm component of wage inequality. Different firms pay different wages to otherwise identical workers as a consequence of labor market imperfections. The underlying source of between-firm inequality varies across studies, including search frictions and bargaining (Davidson et al., 2008; Helpman et al., 2010; Coşar et al., 2016), efficiency wages (Davis and Harrigan, 2011) and fair wage constraints (Egger and Kreickemeier, 2009; Amiti and Davis, 2011). There is, however, no wage dispersion among co-workers in these models. A notable exception in this literature is Verhoogen (2008), where each firm pays different efficiency wages to identical co-workers employed in different production lines. Effort-wage schedules, however, are exogenous and a characterization of trade-induced equilibrium changes in the distribution of workers across firms is not provided. This prevents a general equilibrium analysis of the impact of trade on within-firm inequality, which is the main goal of this paper.

A literature in organizational economics and international trade, including Garicano and Rossi-Hansberg (2004; 2006) and Caliendo and Rossi-Hansberg (2012), studies models in which firms hire heterogeneous workers in frictionless labor markets. Any two co-workers earn different wages if and only if they have different skill levels; hence these models generate neither between-firm nor within-firm inequality across workers of a given type. While this class of models has no implications for performance pay, they can still generate residual wage dispersion among co-workers if skills are assumed unobservable to the econometrician. However, this literature typically focuses on firm-level inequality or aggregate skill premia and thus falls short of tackling the general equilibrium effects of trade liberalization on within-firm inequality.[8] To illustrate this in a stark way, note that within-firm inequality in an economy could *decrease* even in a situation in which firm-level wage inequality *increases in every firm*, if trade liberalization triggers labor reallocations towards firms with initially low firm-level wage dispersion. I refer the reader to the Appendix for an extended discussion of this and other branches of related research.

3. Model

This section introduces the theoretical framework. 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. Labor endowments, trade costs, effort costs, technology and firm productivity distributions are allowed to be asymmetric across countries for much of the analysis, although section 5 analyzes symmetric equilibria. Throughout, I focus on the description of Home's economy and use an asterisk to denote Foreign's variables.

[8] Interpreting wage dispersion in these models as purely residual requires assuming that variation in skills or human capital in the model is orthogonal to education, experience, occupation and other variables typically observed by econometricians. This is a severe constraint on the applicability of this approach to the analysis of residual wage inequality. Moreover, even if a theory models skills that cannot be predicted by the econometrician, positive firm-level inequality is a necessary yet insufficient ingredient to undertake a comparative statics analysis of within-firm inequality. As section 6 shows, the latter also requires an understanding of changes in the distribution of workers across firms.

3.1. Setup

The timing of events in the model combines elements of Holmström and Milgrom (1987) and the static formulation of Melitz (2003). There is a competitive fringe of risk neutral potential entrants (firms) to the differentiated sector. Upon incurring a sunk entry cost f_e, a firm observes its productivity θ, independently drawn from a distribution $G_\theta(\theta)$ with positive and bounded support $[\theta_L, \theta_H]$. Firms then decide whether to exit, produce solely for the domestic market, or produce for both the domestic and export markets. A successful entrant becomes a monopolistic producer of a single variety of the differentiated good. Production requires a fixed cost f_d. In addition, exporting requires a fixed cost f_x and an iceberg variable trade cost, such that $\tau > 1$ units of the firm's output must be produced per unit that arrives in the foreign market. Fixed and entry costs are measured in units of the domestic differentiated good. Since all firms with the same productivity behave symmetrically in equilibrium, I index firms by θ from now onward. In the presence of moral hazard, each firm hires a mass of workers and designs performance-pay contracts to implement desired effort sequences. Workers accept or reject contracts prior to starting production.

3.2. Sequential production with stochastic performance

In every firm, the production process requires each worker to perform a continuum of tasks over a unit time interval. At each instant $t \in [0, 1]$, worker i chooses a possibly history-dependent effort level $\epsilon_i(t)$, where $\epsilon_i(t) \in [\epsilon_L, \epsilon_H] \subset \Re_{++}$. Effort choices generate a stochastic process of worker-specific performance outcomes, denoted $Z_i(t)$, such that:

$$Z_i(t) = \int_0^t \mu\left(\epsilon_i\left(t'\right)\right) dt' + B_i(t), \tag{1}$$

for $t \in [0, 1]$, where $\mu(\cdot)$ is a continuous and increasing function and $B_i(t)$ is a Wiener process on $0 \le t \le 1$, such that for all i, $B_i(0) = 0$ a.s. and $E\left[B_i(1)^2\right] = 1$. Note that individual performance is independent of firm productivity and, conditional on effort, independent across any two tasks and any two workers.[9]

In section A.1 of the Appendix, I show that this continuous-time setting can be interpreted as the limit of a sequential production process with a discrete number of tasks in which workers make mistakes stochastically and effort reduces the likelihood of a mistake. In particular, I show that, as task duration approaches zero, the cumulative performance of worker i up to task t (i.e. the number of successes minus the number of mistakes accumulated up to time t) converges to the Brownian process (1). Consequently, I interpret $-Z_i(1)$ as the *net* number of mistakes of worker i at the end of the production process and define the average net number of mistakes, denoted n, an input in the quality production function, as follows:

$$n \equiv h^{-1} \int_0^h -Z_i(1) di, \tag{2}$$

where h is the mass of workers allocated to the production process.

[9] The randomness of a task's outcome captures unmodeled determinants of a worker's performance such as unobserved skills, idiosyncratic variation in the quality of inputs used in the production process and subjective evaluations.

Because the Z_i's are conditionally independent across workers, the exact law of large numbers (LLN) implies that the firm fully diversifies the impact of idiosyncratic individual performance, $B_i(1)$, on n.[10] Moreover, in section 4.1, I show that firms optimally implements non-stochastic effort sequences. Therefore, if every worker in a firm exerts a constant effort ϵ, then $n = -\mu(\epsilon)$ almost surely and thus equilibrium firm-level variables such as quality, output price, revenue and profits are deterministic with probability one.

Output and quality. Output is horizontally and vertically differentiated. Physical output (y) in a firm with productivity θ is linear in the mass of workers,

$$y = \theta^s h, \quad s \geq 0. \tag{3}$$

The analysis thus nests the case of identical labor productivity across firms ($s = 0$). In turn, product quality (q) is non-negative and C^2, increasing in firm productivity and decreasing in the average net number of mistakes; that is,

$$q = q(\theta, n), \quad q_n < 0 < q_\theta.$$

In section 4.2, I make additional assumptions on $q(\cdot, \cdot)$. To ensure the existence and uniqueness of a solution to the profit maximization problem, I assume that $q(\theta, -\mu(\epsilon))$ is concave in ϵ and $q_\epsilon / q \to \infty$ as $\epsilon \to 0$.[11] To study comparative statics, I rely on submodularity of $q(\theta, n)$ in θ and n or, equivalently, supermodularity of $q(\theta, -\mu(\epsilon))$ in θ and ϵ. This additional structure, however, is not necessary to characterize optimal contracts in section 4.1.

3.3. Demand

Home is populated by a continuum of identical risk-neutral workers of mass L. The preferences of any worker i depend on the consumption of a differentiated product X_i and on the sequence of effort $\boldsymbol{\epsilon}_i \equiv \{\epsilon_i(t); t \in [0, 1]\}$ exerted during the production process:

$$U(X_i, \boldsymbol{\epsilon}_i) = \frac{X_i}{\exp\left(\int_0^1 k(\epsilon_i(t))\, dt\right)}, \tag{4}$$

where $k(\cdot)$ is an increasing and convex instantaneous cost-of-effort function. X_i indexes the consumption of a continuum of horizontally and vertically differentiated varieties, defined as

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

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.[12]

[10] I rely throughout on applications of the exact law of large numbers for a continuum of random variables; see Sun (2006).

[11] Henceforth, I use q_ϵ to denote the partial derivative of $q(\theta, -\mu(\epsilon))$ with respect to ϵ.

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

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 final consumption, differentiated products are also demanded by firms to set up production and export activities (sunk and fixed costs). These activities are assumed to use the output of each variety in the same way as is demanded by final consumers. Let E denote the total expenditure on the differentiated good by firms and consumers located in Home. The aggregate demand for variety j in Home, denoted $x(j)$, is

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

The aggregate expenditure on variety j in Home, denoted $r(j)$, is

$$r(j) \equiv p(j)x(j) = Aq(j)^{\rho} x(j)^{\rho}, \tag{5}$$

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

For expositional purposes, I simplify the notation by setting the aggregate consumption index in Home to be the numeraire ($P = 1$). The utility of domestic consumers can then be expressed solely as a function of income and effort choices; i.e., $U(X_i, \boldsymbol{\epsilon}_i) = U(w_i, \boldsymbol{\epsilon}_i)$.

4. The firm's problem

This section studies the problem of firm θ located in Home, in two steps. The first step takes employment and quality as given and focuses on the design of optimal contracts to attain the targeted quality at minimum cost. The second step sets up the profit maximization problem and determines employment, quality and the export decision.

4.1. *Optimal performance-pay contracts*

The cost of attaining a given quality $q_0 = q(\theta, n_0)$ per unit of output is determined by the cost of providing incentives such that the average net number of mistakes in the production process is n_0. A *performance-pay contract* for any worker i is an arbitrary function $w_i = w_i(Z_i^1)$, stipulating i's wage based on the realized path of individual performance Z_i^1; i.e., $Z_i^1 \equiv \{Z_i(t); t \in [0, 1]\}$. Workers accept or reject contracts prior to starting production at time $t = 0$, select effort in each task t having observed $\{Z_i(t'); t' \in [0, t)\}$ and receive wages upon completion of all tasks at time $t = 1$.

To attain q_0, a firm employing h workers designs a set of contracts and effort sequences $\{w_i, \boldsymbol{\epsilon}_i; i \in [0, h]\}$ that minimize expected total compensation subject to: (i) inducing no more than n_0 net mistakes per worker, (ii) the stochastic processes for individual performance, (iii) incentive compatibility constraints and (iv) participation constraints:

$$\min_{\{w_i, \boldsymbol{\epsilon}_i; i \in [0,h]\}} \int_0^h E\left[w_i(Z_i^1)\right] di, \tag{6}$$

$$\begin{aligned}
\text{s.t.}\quad & \text{(i) } n_0 \geq h^{-1} \textstyle\int_0^h E\left(-Z_i(1)\right) di, \\
& \text{(ii) } Z_i(t) = \textstyle\int_0^t \mu\left(\epsilon_i\left(t'\right)\right) dt' + B_i(t), \text{ for } i \in [0, h], \\
& \text{(iii) } \boldsymbol{\epsilon}_i \in \arg\max_{\widehat{\epsilon}_i} E\left[U(w_i, \widehat{\epsilon}_i)\right], \text{ for } i \in [0, h], \\
& \text{(iv) } E\left[U(w_i, \boldsymbol{\epsilon}_i)\right] \geq \overline{u}, \text{ for } i \in [0, h],
\end{aligned}$$

where $\overline{u}$ is the worker's outside option, endogenously determined in equilibrium.

Proposition 1 characterizes the solution to this problem for the case in which $n_0 \in N \equiv [-\mu(\epsilon_H), -\mu(\epsilon_L))$. The infimum of N ensures that n_0 is technologically feasible, a necessary condition for the existence of a solution in (6). In turn, the supremum of N makes the moral hazard problem interesting, by requiring the firm to implement effort levels greater than ϵ_L.[13] To establish the result, I need:

Assumption 1. For all $\epsilon \in [\epsilon_L, \epsilon_H]$,

(a) $\mu(\epsilon)$ is C^1, strictly increasing and strictly concave.
(b) $k(\epsilon)$ is C^1, strictly increasing and convex.

Proposition 1 *(Cost-minimizing contracts). Suppose that $n_0 \in N$. Then, under Assumption 1, there exists an a.s. unique global minimizer in problem (6), denoted $\{w_i^*, \boldsymbol{\epsilon}_i^*; i \in [0, h]\}$, such that for all $i \in [0, h]$:*

(a) Effort: $\epsilon_i^*(t) = \epsilon^*$ for all $t \in [0, 1]$, where

$$\epsilon^* = \mu^{-1}(-n_0).$$

(b) Contract: $\log(w_i^*) = \alpha + \beta Z_i(1)$, where

$$\beta = k'(\epsilon^*)/\mu'(\epsilon^*),$$

$$\alpha = \log \overline{u} + k(\epsilon^*) - \frac{k'(\epsilon^*)}{\mu'(\epsilon^*)}\mu(\epsilon^*) - \frac{1}{2}\left[\frac{k'(\epsilon^*)}{\mu'(\epsilon^*)}\right]^2.$$

(c) Performance: $Z_i(1) \overset{d}{\sim} N\left[\mu(\epsilon^*), 1\right]$.

The optimal contract for worker i is a log-linear function of i's cumulative performance at time $t = 1$ and implements a constant effort in each task of the production process. It is straightforward to interpret the performance component of the contract, β, as defining either a bonus or a (non-linear) piece-rate/commission based on performance at the end of the production process.[14]

The solution to problem (6) has several important features. First, it inherits the simple structure of optimal contracts from Holmström and Milgrom (1987). As in that paper, tasks (time periods) are technologically independent and consumption takes place after production, eliminating any scope for improved statistical inference and for consumption smoothing throughout the production process. A conceptually significant departure from Holmström and Milgrom (1987) is the specification of the objective functions of firms and workers. In particular, the firm's cost minimization problem (6) arises naturally in the context of the broader profit maximization problem studied in the next section. Moreover, the utility function (4) plays a key role in ensuring that wages are positive for all realizations of individual performance $Z_i(1)$.[15] Because wages fuel

[13] If $n_0 \geq -\mu(\epsilon_L)$, the firm can satisfy (i) by simply offering a constant wage that ensures participation, trivializing the moral hazard problem.

[14] Under the former interpretation, the bonus is a fraction $\exp\left[\beta Z_i(1)\right] - 1$ of base wage $\exp[\alpha]$. Note that bonus size has worker-specific (Z_i) and firm-specific (α, β) components. Because the latter depend on firm productivity (next section), the model effectively allows the bonus to depend on both individual performance and firm profit.

[15] In Holmström and Milgrom (1987), firms and workers have negative exponential (CARA) objective functions defined over cumulative performance at $t = 1$ and compensation, respectively. Moreover, effort costs are measured in monetary

the demand side of the model, this is an appealing property when embedding the moral hazard problem in general equilibrium.

Second, the firm's cost minimizing strategy is to offer identical contracts to its h employees. In principle, the firm could offer different contracts to different workers, yet this is not cost-effective. The symmetry of optimal effort levels –part (a)– follows from the convexity of the effort cost function $k(\cdot)$ and the concavity of $\mu(\cdot)$. Intuitively, the convexity of $k(\cdot)$ 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. In addition, the concavity of $\mu(\cdot)$ implies that a higher-than-average effort of some worker does not compensate the mistakes incurred by a lower-than-average effort of another worker. The strict concavity of $\mu(\cdot)$ ensures the uniqueness of the optimal effort and contract, up to an almost sure equivalence.

Third, it is straightforward to verify that β is increasing in ϵ^* under Assumption 1. Incentive compatibility requires the intensity of performance pay (proxied by β) to increase in effort, which is consistent with empirical studies documenting performance gains from performance pay.[16] The firm adjusts the fixed component of compensation α to ensure that the participation constraint is satisfied with equality.

The model generates patterns of heterogeneity in firm-level wage distributions that are consistent with the empirical patterns documented in section 7. These core cross-sectional predictions build on the following implications of Proposition 1.

Corollary 1 *(Firm-level wages and inequality). Suppose that the firm implements a constant effort* $\epsilon \in [\epsilon_L, \epsilon_H]$ *such that* $\epsilon_i(t) = \epsilon$ *for all* $t \in [0, 1]$ *and* $i \in [0, h]$*. Then, under Assumption 1:*

(a) The average wage paid by the firm is a.s. $\omega(\epsilon) \equiv \overline{u}e^{k(\epsilon)}$*, increasing in* ϵ*.*
(b) Firm-level wage inequality is increasing in ϵ*, according to the variance of log wages and all inequality measures that respect second-order stochastic dominance and scale independence.*

Corollary 1 implies that high wage firms are also high inequality firms in the model. Part (a) follows from the LLN; hence $\omega(\epsilon)$ equals $E\left[w_i^*|\epsilon\right]$ almost surely. As with output quality, the firm fully diversifies the impact of idiosyncratic performance on the average wage paid to its employees. Without loss of generality, I treat the firm's profit maximization problem as deterministic in the next section. Part (b) builds on the observation that, by Proposition 1, the distribution of firm-level wages is log-normal. It states that the optimal provision of incentives generates higher firm-level inequality in high-effort firms, according to a large class of inequality measures.

The dynamic formulation of the moral hazard problem adopted here has desirable theoretical and empirical implications that simpler formulations do not possess: (i) it delivers tractable optimal contracts that can be embedded in general equilibrium; (ii) it predicts a positive correlation between the first and second moments of firm-level wage distributions –Corollary 1– that is consistent with the empirical patterns documented on page 500. Section A.4 of the Appendix argues that standard (static) formulations of the moral hazard problem fail to satisfy at least one of these two criteria.

units. The optimal contract is a linear function of a normally distributed random variable and thus the support of the wage distribution is $\mathbb{R}$.

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

4.2. Profit maximization

If price and quality discrimination across markets is feasible, then by (5) revenues from domestic and foreign sales are given by $r_d = Aq_d^\rho y_d^\rho$ and $r_x = A^* q_x^\rho [y_x/\tau]^\rho$, respectively. r_m, q_m and y_m denote revenue, quality and output in market $m = \{d, x\}$, respectively. Recall that fixed costs are measured in units of the domestic differentiated good, whose price is normalized to one. It follows that the profit maximization problem of firm θ located in Home is additively separable in domestic and foreign profits and can be written as

$$\Pi(\theta) \equiv \max_{\substack{\epsilon_m \in [\epsilon_L, \epsilon_H], \\ q_m \in [q_L, q_H], \\ y_m \geq 0,\ I_x \in \{0,1\}}} Aq_d^\rho y_d^\rho - \frac{\omega(\epsilon_d)}{\theta^s} y_d - f_d + I_x \left[A^* q_x^\rho [y_x/\tau]^\rho - \frac{\omega(\epsilon_x)}{\theta^s} y_x - f_x \right], \tag{7}$$

where $q_\ell = q(\theta, -\mu(\epsilon_\ell))$, $\ell = \{m, L, H\}$ and $m = \{d, x\}$, and I_x equals 1 if firm θ exports and 0 otherwise.[17] The average wage function, $\omega(\cdot)$, is obtained from part (a) of Corollary 1.

Profits are strictly concave in output and marginal revenue of output is infinite as output approaches zero. Therefore, for any market $m = \{d, x\}$: (i) the first-order condition with respect to y_m is necessary and sufficient to maximize profits in m, for any given quality $q_m \in [q_L, q_H]$; (ii) corner solutions for output ($y_m = 0$) are ruled out. I thus solve problem (7) in three steps. First, assuming $I_x = 1$, I compute the optimal output in each market for a given quality q_m, denoted, $y_m(q_m)$. Second, I characterize the optimal q_m, accounting for its effect on $y_m(q_m)$. Finally, I determine whether exporting is profit maximizing.

Let $c^\theta(q) \equiv \omega(\epsilon(\theta, q))/\theta^s$, where $\epsilon(\theta, q)$ is implicitly defined by $q = q(\theta, -\mu(\epsilon))$. Then $c^\theta(q)$ is the (factory) unit-cost function in firm θ when output quality is q. Profits in market $m = \{d, x\}$ can be written as $\Pi_m \equiv \eta_m q_m^\rho y_m^\rho - c^\theta(q_m) y_m - f_m$, where $\eta_d \equiv A$ and $\eta_x \equiv A^* \tau^{-\rho}$.

For any fixed $q_m \in [q_L, q_H]$, output in market m maximizes Π_m if and only if it equalizes the marginal revenue of output and the marginal cost of output,

$$\rho \eta_m q_m^\rho [y_m(q_m)]^{\rho - 1} = c^\theta(q_m), \text{ for } m = \{d, x\}. \tag{8}$$

Solving for $y_m(q_m)$ from (8), substituting it in Π_m and rearranging yields

$$\Pi_m(q_m) = \left[(\rho)^{\rho/(1-\rho)} - (\rho)^{1/(1-\rho)} \right] (\eta_m)^{1/(1-\rho)} \left(\left[\frac{q_m}{c^\theta(q_m)} \right]^{\rho/(1-\rho)} \right) - f_m. \tag{9}$$

The restriction $0 < \rho < 1$ implies $(\rho)^{\rho/(1-\rho)} > (\rho)^{1/(1-\rho)}$. Therefore, quality $q_m(\theta)$ is profit maximizing for firm θ in market m if and only if $q_m(\theta)$ minimizes the average cost of quality (per unit of output) in m, $c^\theta(q)/q$. By Weierstrass theorem, if c^θ is continuous then $q_m(\theta)$ exists, since $q \in [q_L, q_H]$. If, in addition, c^θ is differentiable and $q_m(\theta) \in (q_L, q_H)$, then $c_q^\theta(q_m(\theta)) = c^\theta(q_m(\theta))/q_m(\theta)$, for $m = \{d, x\}$. Geometrically, the marginal and average costs of quality intersect at $q_m(\theta)$.

Importantly, the average cost of quality is independent of the variable trade cost τ. Therefore, trade liberalization does not induce quality upgrading or downgrading at the firm level, in either

[17] 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. Still, in equilibrium workers are indifferent between employment in either production line because every contract yields the same expected utility.

market. Moreover, if $q_m(\theta)$ is the unique global minimizer of $c^\theta(q)/q$ then $q_d(\theta) = q_x(\theta)$, and therefore, conditional on exporting, firm θ offers products of identical quality in the domestic and foreign markets.[18] The following result summarizes these properties and provides a sufficient condition for the existence and uniqueness of an optimal quality when c^θ is C^2.[19]

Lemma 1. *Suppose that $c^\theta(q)$ is C^2. If Assumption 1 holds, $q(\theta, -\mu(\epsilon))$ is concave in ϵ and $q_\epsilon/q \to \infty$ as $\epsilon \to 0$, then there is a unique profit maximizing quality in firm θ, denoted $q(\theta)$, independent of both the variable trade cost and the export status of the firm.*

Under Lemma 1, the firm's unit costs (at the factory gate) are also identical across markets and thus the optimal allocation of its total output must in turn equalize the marginal revenue of output in the domestic and foreign markets. From (5), this requires $[y_x(\theta)/y_d(\theta)]^{1-\rho} = \tau^{-\rho}(A^*/A)$, which implies that the firm's total revenue is

$$r(\theta) \equiv r_d(\theta) + I_x(\theta)r_x(\theta) = Aq(\theta)^\rho y(\theta)^\rho \Upsilon(\theta)^{1-\rho}, \tag{10}$$

where $y(\theta)$ is the total output of firm θ, respectively. Following Helpman et al. (2010), $\Upsilon(\theta) \equiv 1 + I_x(\theta)\left[\tau^{-\rho}(A^*/A)\right]^{1/(1-\rho)}$ is a measure of market access for firm θ. In turn, expression (8) implies that the firm's total profits is a fraction $1-\rho$ of total revenue net of fixed costs,

$$\Pi(\theta) = (1-\rho)r(\theta) - f_d - I_x(\theta)f_x. \tag{11}$$

As long as total revenue increases in firm productivity, 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 $I_x(\theta) = 0$ if and only if $\theta < \theta_x$. This implies that the firm market access variable is

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

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

Comparative statics. It is possible to characterize how the solution to the firm's problem varies across firms using standard results from the theory of monotone comparative statics. The key assumption is submodularity of $q(\theta, n)$ in θ and n or, equivalently, supermodularity of $q(\theta, -\mu(\epsilon))$ in θ and ϵ. Complementarity between productivity and effort implies that high productivity firms have a comparative advantage in producing high quality output, which is consistent with the empirical evidence in Kugler and Verhoogen (2012) and Manova and Zhang (2012). Complementarity also implies that the average cost of quality per unit of output is lower in high productivity firms. Hence the latter optimally expand employment and output, earning higher revenue and profits, provided that demand is sufficiently elastic.

[18] This result is driven by the assumption of identical consumer preferences across countries. Quality upgrading induced by exporting can be easily introduced into the model by assuming that foreign consumers trade off quality and quantity differently than domestic consumers, as in Verhoogen (2008). For example, letting $X_i^* = \left[\int_{j\in J^*}\left(q^*(j)^\chi x_i^*(j)\right)^\rho dj\right]^{1/\rho}$ and $\chi > 1$. Alternatively, $\chi < 1$ induces domestic exporters to downgrade quality. This suggests that tastes in export destinations matter, as they may amplify or dampen the link between trade and inequality advanced in this paper.

[19] In Lemma 1, the condition $q_\epsilon/q \to \infty$ as $\epsilon \to 0$ ensures that the optimal quality is positive. Moreover, as shown in the Appendix, $q(\theta) \in [q_L, q_H]$ provided that effort is defined over a sufficiently large interval.

Proposition 2. *If, in addition to the hypotheses of Lemma 1, $q(\theta, n)$ is submodular in θ and n, then optimal quality and effort are increasing in firm productivity. Moreover, if ρ is large enough, then optimal output, employment, revenue and profits are increasing in firm productivity.*

By Corollary 1, key properties of the firm-level wage distribution depend exclusively on optimal effort or, equivalently, optimal quality. The invariance of quality with respect to the trade cost –Lemma 1– and the monotonic relationship between quality and measures of firm size (output, employment or revenue) –Proposition 2– immediately deliver testable predictions that are analyzed in section 7.

Corollary 2. *Under the hypotheses of Proposition 2, the following functions of the firm-level wage distribution are increasing in firm size if ρ is large enough: the average wage, the variance of log wages and all inequality measures that respect second-order stochastic dominance and scale independence. Conditional on firm size, however, the firm-level wage distribution is independent of both the variable trade cost and the export status of the firm.*

Closed-form solutions. The following three functional form assumptions yield closed-form solutions to the profit maximization problem:

$$k(\epsilon) = k\epsilon,\ k > 0, \tag{12}$$

$$\mu(\epsilon) = -1/\epsilon, \tag{13}$$

$$\log q = (\gamma \log \theta)^z (1/n)^{(1-z)},\ \gamma > 0, z \in (0, 1). \tag{14}$$

Note that (12) and (13) satisfy Assumption 1. By (2) and (13), $n = 1/\epsilon > 0$ almost surely.[20] This guarantees that quality is properly defined in (14). Under (14), quality is log-submodular in θ and n. The parameter γ in (14) can be interpreted as the scope for quality differentiation, as in Kugler and Verhoogen (2012).

Minimizing the average cost of quality under (12)–(14) yields a closed-form solution for a unique optimal quality $q(\theta)$ (see Appendix). With slight abuse of notation, optimal effort is in turn obtained by inverting $q(\theta) = q(\theta, -\mu(\epsilon(\theta)))$. In the case of interior solutions,

$$q(\theta) = \theta^{\kappa_q},\ \ \kappa_q \equiv \gamma \left[(1-z)/k\right]^{(1-z)/z}, \tag{15}$$

$$\epsilon(\theta) = \kappa_\epsilon \log \theta,\ \ \kappa_\epsilon \equiv \gamma \left[(1-z)/k\right]^{1/z}. \tag{16}$$

Note that $q(\theta)$ and $\epsilon(\theta)$ will in fact be interior for every firm $\theta \in [\theta_L, \theta_H]$ provided that individual effort is defined over a sufficiently large interval; that is, if $\kappa_\epsilon \left[\log \theta_L, \log \theta_H\right] \subset [\epsilon_L, \epsilon_H]$. To avoid a taxonomic exercise, I will henceforth focus on this parameter configuration.[21]

From the first-order condition for output (8) and the expression for firm revenue (10), I solve for total output and revenue as functions of the demand shifters and the reservation utility. Total employment, denoted $h(\theta)$, follows from the production function (3). Therefore:

---

[20] Under (13), $\mu(\cdot) \subset [\mu(\epsilon_L), \mu(\epsilon_H)]$, satisfying the requirements of Lemma A-1. The latter ensures the validity of the microfoundation for the production process presented in section A.1 of the Appendix.

[21] The model can deliver other, arguably interesting, types of equilibria. For example, if $\epsilon_L > \kappa_\epsilon \log \theta_d$ firms with sufficiently low productivity implement the minimum effort (and quality) by offering a flat wage that is independent of performance. In this equilibrium, only a fraction of the jobs in the economy are performance-pay jobs. This provides a simple rationale for the positive correlation between firm size and the probability of observing positive performance pay, reported in Table 1. Moreover, the fraction of performance-pay jobs will typically increase with trade liberalization, as long as a lower variable trade cost leads to a higher θ_d.

$$r(\theta) = \kappa_r \Upsilon(\theta) \left(A\overline{u}^{-\rho}\right)^{1/(1-\rho)} \theta^{\Gamma}, \quad \kappa_r \equiv \rho^{\rho/(1-\rho)}, \tag{17}$$

$$y(\theta) = \kappa_y \Upsilon(\theta) \left(A\overline{u}^{-1}\right)^{1/(1-\rho)} \theta^{\Gamma+s-k\kappa_\epsilon}, \quad \kappa_y \equiv \rho^{1/(1-\rho)}, \tag{18}$$

$$h(\theta) = \kappa_y \Upsilon(\theta) \left(A\overline{u}^{-1}\right)^{1/(1-\rho)} \theta^{\Gamma-k\kappa_\epsilon}, \tag{19}$$

where $\Gamma \equiv \left[\gamma z\left((1-z)/k\right)^{1/z-1} + s\right]\rho/(1-\rho)$. The condition $\rho > 1-z$ implies $\Gamma > k\kappa_\epsilon$, ensuring that revenue, output and employment increase in productivity for all $s \geq 0$. 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 and effort since, as discussed above, there is no motive for quality upgrading (or downgrading) associated to exporting in this model.

Finally, I characterize the distribution of wages in firm θ. Combining the firm's optimal choice of effort (16) with parts (a) and (b) of Proposition 1 and functional forms (12) and (13), yields the wage of worker i employed in firm θ,

$$w_i = \overline{u}\theta^{k\kappa_\epsilon} \exp\left[(\kappa_\epsilon \log\theta)^3 \left(B_i(1) - (\kappa_\epsilon \log\theta)^6/2\right)\right]. \tag{20}$$

The expected wage and variance of log wages,

$$E\left[w_i|\theta\right] = \overline{u}\theta^{k\kappa_\epsilon}, \tag{21}$$

$$Var\left[\log w_i|\theta\right] = (\kappa_\epsilon \log\theta)^6, \tag{22}$$

are increasing in firm productivity θ but independent of both the variable trade cost and the export status of the firm.

5. Equilibrium

This section explains how to compute the remaining endogenous variables of the model in symmetric equilibria, in which labor endowments, trade costs, effort costs, technology and firm productivity distributions are identical across countries. Moreover, I establish the existence and uniqueness of this class of equilibria.

The zero-profit cutoff θ_d is the productivity level that makes firms indifferent between exiting and producing for the domestic market. In turn, the exporting cutoff θ_x makes firms indifferent between exporting and producing exclusively for the domestic market. From the expressions for revenue (17) and profits (11), these two conditions require

$$\kappa_r(1-\rho)\left(\overline{u}\right)^{-\rho/(1-\rho)} \theta_d^{\Gamma} E = f_d \tag{23}$$

and

$$\kappa_r(1-\rho)\left(\overline{u}\tau\right)^{-\rho/(1-\rho)} \theta_x^{\Gamma} E = f_x, \tag{24}$$

respectively.[22] Dividing (24) by (23) yields

$$\tau^{-\frac{\rho}{1-\rho}} \left(\frac{\theta_x}{\theta_d}\right)^{\Gamma} = \frac{f_x}{f_d}. \tag{25}$$

[22] In deriving these expressions, I use $A = E^{(1-\rho)}$, which follows by definition of the demand shifter A and the choice of numeraire ($P = 1$). Expression (24) also uses the fact that $\Upsilon_x - 1 = \tau^{-\frac{\rho}{1-\rho}}$ in any symmetric equilibrium.

Free entry implies that the expected profit of successful entrants should equal the sunk entry cost; that is, $\int_{\theta_d}^{\theta_H} \Pi(\theta) dG_\theta(\theta) = f_e$. Using the expressions for revenue (17) and productivity cutoffs (23) and (24), the free entry condition is

$$f_d J(\theta_d) + f_x J(\theta_x) = f_e, \tag{26}$$

where $J(\theta_m) \equiv \int_{\theta_m}^{\theta_H} \left[(\theta/\theta_m)^\Gamma - 1\right] dG_\theta(\theta)$, $m = \{d, x\}$, is monotonically decreasing in $[\theta_L, \theta_H]$ for any $G_\theta(\theta)$. To ensure that J is finite, I henceforth assume that the distribution of firm productivity has a finite Γ-th uncentered moment.

Equations (25) and (26) fully determine the productivity cutoffs. For the remainder of the paper, I restrict the analysis to a class of equilibria satisfying $\theta_L < \theta_d \leq \theta_x$. The condition $\theta_L < \theta_d$ is necessary for the existence of equilibrium.[23] As shown in the Appendix, given f_x, f_d, f_e, Γ and $G_\theta(\theta)$, this condition holds provided that firm productivity is defined over a sufficiently large interval $[\theta_L, \theta_H]$ (see proof of Proposition 3). In turn, $\theta_d \leq \theta_x$ is imposed for consistency with a large empirical literature documenting selection of the most productive firm into exporting. As in Melitz (2003), $\theta_d \leq \theta_x$ if and only if $(f_x/f_d)\tau^{\frac{\rho}{1-\rho}} \geq 1$. The ensuing analysis therefore encompasses closed economy equilibria ($\theta_x \geq \theta_H$) and equilibria in which all firms export ($\theta_d = \theta_x$).

Equilibrium in the differentiated goods market requires the equality of aggregate expenditure and aggregate revenue. The latter equals $M\overline{r}$, where M and $\overline{r}$ denote the mass and average revenue of active firms, respectively. Given a mass of entrants M_e, the mass of active firms is $[1 - G_\theta(\theta_d)] M_e$. In turn, the average revenue of producers can be written as a function of the productivity cutoffs by integrating firm profits (11) and using the free entry condition.[24] The market clearing condition for the goods market thus becomes

$$(1-\rho)E = M_e \left[f_e + f_d \left[1 - G_\theta(\theta_d)\right] + f_x \left[1 - G_\theta(\theta_x)\right]\right]. \tag{27}$$

Finally, labor market clearing requires equating labor supply, L, and labor demand, $M_e \int_{\theta_d}^{\infty} h(\theta) dG_\theta(\theta)$. Using expression (19) to substitute for firm employment yields

$$L = \kappa_y (\overline{u})^{-1/(1-\rho)} E M_e \int_{\theta_d}^{\theta_H} \Upsilon(\theta) \theta^{\Gamma - k\kappa_\epsilon} dG_\theta(\theta). \tag{28}$$

The general equilibrium with two symmetric countries is characterized by productivity cutoffs θ_d and θ_x, aggregate expenditure E, mass of entrants M_e and reservation utility $\overline{u}$ that solve equations (23), (24), (26), (27) and (28).

Proposition 3. *There exists a unique symmetric equilibrium.*

6. Comparative statics: trade liberalization

As an application of the model, this section studies the impact of trade liberalization on wage inequality between homogeneous workers. The analysis centers on symmetric equilibria, although I also provide a partial characterization of comparative statics in asymmetric equilibria.

[23] On the other hand, $\theta_d < \theta_H$ is always satisfied. Otherwise, if $\theta_H \leq \theta_d \leq \theta_x$ then the free entry condition would be violated.

[24] In particular, integrating (11) and using the fact that, by free entry, average firm profits are equal to $f_e/\left[1 - G(\theta_d)\right]$, yields $(1-\rho)\overline{r} = f_e/\left[1 - G(\theta_d)\right] + f_d + f_x \left[1 - G(\theta_x)\right]/\left[1 - G(\theta_d)\right]$.

As a first step, I study reallocations of workers and wage shares across firms. Because optimal performance-pay contracts differ across firms, resource reallocations have profound implications for the equilibrium distribution of wages in the economy. Finally, I focus on characterizing changes in within-firm wage dispersion analytically, highlighting a new channel through which international trade can contribute to residual wage inequality.

The comparative statics exercise is the following. Consider a decline in the (bilateral) variable trade cost, $\tau_0 > \tau_1$, for $\tau_1 \in [\underline{\tau}, \overline{\tau})$, holding the remaining parameters of the model constant. The lower bound for τ_1 is $\underline{\tau} \equiv \max\{1, (f_d/f_x)^{(1-\rho)/\rho}\}$, ensuring that the post-liberalization equilibrium features selection of the most productive firms into exporting. If $f_d > f_x$ then $\tau_1 = \underline{\tau}$ corresponds to an equilibrium in which every firm exports following trade liberalization. The supremum for τ_1, denoted $\overline{\tau}$, is the variable trade cost such that $\theta_{x,1} = \theta_H$, implicitly defined by equations (25) and (26). Note that $\tau_j \geq \overline{\tau}$ if and only if equilibrium j is autarkic. The restriction $\tau_1 < \overline{\tau}$ thus ensures that trade liberalization effectively opens the economy. When needed, I use a subindex $j \in \{0, 1\}$ to indicate equilibria before ($j = 0$) and after ($j = 1$) trade liberalization.

6.1. Labor reallocations

The mass of workers employed in firms with productivity lower than or equal to θ is $M \int_{\theta_d}^{\theta} h(\theta') dG_\theta(\theta'|\theta' \geq \theta_d)$, for $\theta \geq \theta_d$. In any equilibrium of the model, the distribution of employment across firms, denoted $G_h(\theta)$, measures the fraction of workers employed in firms with productivity less than or equal to θ. Using the expression for firm employment (19),

$$G_h(\theta) = \frac{\int_{\theta_d}^{\theta} \Upsilon(\theta') \left(\theta'\right)^{\Gamma - k\kappa_\epsilon} dG_\theta(\theta')}{\int_{\theta_d}^{\theta_H} \Upsilon(\theta') (\theta')^{\Gamma - k\kappa_\epsilon} dG_\theta(\theta')}, \text{ for } \theta \in [\theta_d, \theta_H]. \tag{29}$$

Importantly, $G_h(\theta)$ depends on just two endogenous variables; namely, the productivity cutoffs.[25] This key property enables an analytical characterization of changes in the distribution of employment in terms of changes in the productivity cutoffs across equilibria.

In symmetric equilibria, productivity cutoffs respond to trade liberalization as in Melitz (2003). To see this, let $\theta_{d,j}$ and $\theta_{x,j}$ denote the domestic and export productivity cutoffs in an equilibrium with variable trade cost τ_j, respectively. From (25), $\tau_0 > \tau_1$ if and only if $\theta_{x,0}/\theta_{d,0} > \theta_{x,1}/\theta_{d,1}$. Moreover, the free entry condition (26) and the monotonicity of $J(\cdot)$ imply that the productivity cutoffs $\theta_{d,j}$ and $\theta_{x,j}$ are inversely related, for $j \in \{0, 1\}$. Given the restrictions imposed on trade costs, it follows immediately that $\theta_{d,0} < \theta_{d,1} \leq \theta_{x,1} < \theta_{x,0}$. Trade liberalization leads to the exit of the least productive firms and, equivalently, to the entry of new exporters to the foreign market. Under this configuration of cutoffs, I can establish the following result.

Lemma 2. *Let $G_{h,0}$ and $G_{h,1}$ denote employment distributions corresponding to symmetric equilibria before and after a bilateral trade liberalization, respectively.*

(a) If $\tau_0 \geq \overline{\tau}$, then $G_{h,1}$ first-order stochastically dominates $G_{h,0}$.

---

[25] To see this, note that $\Upsilon_x = 1 + \tau^{-\frac{\rho}{1-\rho}}$ in any symmetric equilibrium, thus $\Upsilon(\theta)$ depends only on the productivity cutoffs. This property, however, does not rely on symmetry. In particular, Section A in the Appendix shows that, in any equilibrium of the model, $\Upsilon_x = 1 + (f_x/f_d)(\theta_d/\theta_x)^\Gamma$ in the Home country, where θ_d and θ_x are the productivity cutoffs in Home.

(b) *If* $\tau_0 < \overline{\tau}$, *consider* $\widehat{\theta} \in \left[\theta_{x,0}, \theta_H\right)$:
If $G_{h,1}(\widehat{\theta}) \leq G_{h,0}(\widehat{\theta})$, *then* $G_{h,1}$ *first-order stochastically dominates* $G_{h,0}$.
If $G_{h,1}(\widehat{\theta}) > G_{h,0}(\widehat{\theta})$, *then* $G_{h,1}$ *intersects* $G_{h,0}$ *once, from below, in* $\left[\theta_{d,1}, \theta_H\right)$.

Lemma 2 describes the two types of admissible changes in the employment distribution that result from a decline in variable trade costs. In either case, the result is reminiscent of the tendency towards higher concentration of workers in high productivity firms, following trade liberalization, inherent of models with firm heterogeneity and selection into exporting.[26] Indeed, Lemma 2 implies that $G_{h,1}$ second-order stochastically dominates $G_{h,0}$ whenever trade liberalization reduces the mass of active firms.[27] For a sharp characterization of the effect of trade liberalization on within-firm inequality, however, I will rely on a stronger form of labor reallocations, obtained by imposing additional structure on the distribution of firm productivity.

Assumption 2. For $\theta \in [\theta_L, \theta_H]$,

$$\frac{J'(\theta)}{g_\theta(\theta)} \text{ is non-decreasing in } \theta.$$

Assumption 2 is satisfied by a class of productivity distributions that includes Pareto, distributions with non-decreasing densities and, more generally, densities with elasticity greater than or equal to $-(\Gamma + 1)$. Moreover, Assumption 2 holds whenever the hazard function of the distribution of (rescaled) domestic revenue is non-increasing, an empirically relevant case.[28] See Sections A.9 and A.10 in the Appendix. Assumption 2 is sufficient to establish that $G_{h,1}(\widehat{\theta}) \leq G_{h,0}(\widehat{\theta})$, for any $\widehat{\theta} \in \left[\theta_{x,0}, \theta_H\right)$. In light of Lemma 2, I obtain Proposition 4. Note that this result is not implied by the standard form of labor reallocations established in Melitz (2003).[29]

Proposition 4. *Let* $G_{h,0}$ *and* $G_{h,1}$ *denote employment distributions corresponding to symmetric equilibria before and a bilateral after trade liberalization, respectively. If* $\tau_0 < \overline{\tau}$, *impose Assumption 2. Then* $G_{h,1}$ *first-order stochastically dominates* $G_{h,0}$.

[26] In Melitz (2003), for example, trade liberalization leads to higher employment in exporting firms.

[27] To see this, recall that first-order stochastic dominance implies second-order stochastic dominance. Moreover, note that the single-crossing property stated in part (b) of Lemma 2 is equivalent to second-order stochastic dominance if mean firm employment increases following trade liberalization (see, for example, Proposition 4.6 in Wolfstetter, 1999). Because the labor market clears, the latter condition holds if and only if trade liberalization reduces the mass of active firms.

[28] Di Giovanni et al. (2011) show that the distribution of domestic revenue of French firms is very well approximated by a Pareto distribution, which has a decreasing hazard function.

[29] Melitz (2003) shows that employment in every exporter (non-exporter) increases (decreases) when the economy opens from autarky (note that the inequality in p. 1714 of that paper also holds for firm employment because the latter is a linear function of firm revenue). This result, however, is not sufficient to rank employment distributions in terms of first-order stochastic dominance. For example, consider the set of firms with productivity greater than or equal to θ_0, where θ_0 is greater than the export cutoff in the open economy. While employment of every firm in this set increases, the measure of the set may decrease because the mass of firms declines in the open economy. The change in the fraction of workers employed in firms with productivity greater than or equal to θ_0 is therefore ambiguous. Moreover, the result I establish is valid for arbitrary changes in variable trade costs and, under additional qualifications, for asymmetric countries.

Characterizing the impact of trade liberalization on the Theil index of wages requires analyzing changes in the distribution of wages across *firms*. The wage bill paid by firms with productivity lower than or equal to θ is $L\int_{\theta_d}^{\theta}\omega(\theta')dG_h(\theta')$, for $\theta \geq \theta_d$. The distribution of wages across firms, denoted $G_w(\theta)$, measures the fraction of wages paid by firms with productivity less than or equal to θ. Therefore,

$$G_w(\theta) = \frac{\int_{\theta_d}^{\theta}\omega(\theta')dG_h(\theta')}{\int_{\theta_d}^{\theta_H}\omega(\theta')dG_h(\theta')}, \text{ for } \theta \in [\theta_d, \theta_H]. \tag{30}$$

Proposition 5. *Let $G_{w,0}$ and $G_{w,1}$ denote wage distributions across firms corresponding to symmetric equilibria before and after a bilateral trade liberalization, respectively. If $\tau_0 < \overline{\tau}$, impose Assumption 2. Then $G_{w,1}$ first-order stochastically dominates $G_{w,0}$.*

Asymmetric equilibria. Here I discuss the extension of the results derived in this section to the case of two countries with asymmetric parameterizations of labor endowments, trade costs, effort costs, technology and firm productivity distributions.[30] As noted above, the distribution of employment across firms in Home is fully determined by its productivity cutoffs. It is then straightforward to verify that the proofs of Lemma 2 and Propositions (4) and (5) continue to hold in asymmetric equilibria as long as changes in variable trade costs (in either country) lead to the same configuration of pre- and post-liberalization productivity cutoffs in Home, i.e. $\theta_{d,0} < \theta_{d,1} \leq \theta_{x,1} < \theta_{x,0}$. By the free entry condition, this condition holds if and only if changes in variable trade costs trigger firm selection in Home (i.e. $\theta_{d,0} < \theta_{d,1}$, exit of the least productive firms).

The analytical tractability of the link between changes trade costs and changes in productivity cutoffs, however, is lost in asymmetric equilibria. An exception is the case of a small open economy. Following Demidova and Rodriguez-Clare (2013), Home is assumed to be a small open economy if the zero-profit productivity cutoff, aggregate expenditure and price index in Foreign are not affected by Home variables. In this context, the Appendix shows that a unilateral trade liberalization in Home (i.e. a decline in the variable cost of exporting for foreign firms) leads to firm selection in Home.

In light of these remarks, it is possible to provide a partial characterization of the impact of a unilateral trade liberalization on reallocations of workers and wage shares across firms in asymmetric equilibria. In the following result, $\theta_{d,j}$ and G_θ should be understood as Home-specific zero-profit cutoff in equilibrium j and productivity distribution, respectively.

Proposition 6. *Let $G_{h,j}$ and $G_{w,j}$ denote employment and wage distributions across firms in Home corresponding to asymmetric equilibria before ($j = 0$) and after ($j = 1$) a unilateral trade liberalization in Home, respectively. If some firms export from Home in the initial equilibrium, impose Assumption 2 on G_θ. If Home is not a small open economy, suppose that $\theta_{d,0} < \theta_{d,1}$. Then $G_{h,1}$ first-order stochastically dominates $G_{h,0}$ and $G_{w,1}$ first-order stochastically dominates $G_{w,0}$.*

[30] A fully asymmetric parameterization of preferences, however, is not allowed here. While the effort cost may differ across countries, a constant and symmetric elasticity of substitution is needed to preserve the structure of the firm's profit maximization problem and its solutions. Section A in the Appendix provides further details.

6.2. Wage inequality

I start by analyzing the impact of trade liberalization on the variance of log wages.[31] This measure has featured frequently applied in recent empirical studies of wage inequality.[32] Unlike other popular measures of inequality such as the Gini coefficient and the 90–10 wage gap, the variance is additively decomposable into between-firm and within-firm components. This property is analytically convenient to highlight different channels through which international trade impacts wage inequality. At the end of the section, I verify the robustness of the results by analyzing two additional, decomposable measures of inequality, the Theil index and the MLD.

The following propositions anticipate the results in this section. The theory developed in the previous sections delivers a sharp link between international trade liberalization and within-firm inequality in symmetric equilibria. As in Helpman et al. (2010) and Coşar et al. (2016), however, the effects on between-firm inequality are non-monotonic and difficult to characterize analytically without further assumptions, except in a special case discussed at the end of this section.[33]

Proposition 7. *Consider any symmetric equilibrium with variable trade cost τ_0. If $\tau_0 < \overline{\tau}$, impose Assumption 2. Then a bilateral trade liberalization $\tau_1 < \tau_0$, for $\tau_1 \in [\underline{\tau}, \overline{\tau})$, increases within-firm wage inequality, according to the following inequality measures: Variance of log wages, Theil Index and Mean Log Deviation.*

While the exposition below focuses on symmetric equilibria, it is straightforward to extend the argument to asymmetric equilibria that satisfy the conditions of Proposition 6. In the following result, G_θ should be understood as the productivity distribution in Home.

Proposition 8. *Let $\theta_{d,0}$ and $\theta_{d,1}$ denote the zero-profit cutoffs in Home corresponding to asymmetric equilibria before and after a unilateral trade liberalization in Home, respectively. If some firms export from Home in the initial equilibrium, impose Assumption 2 on G_θ. If Home is not a small open economy, suppose that $\theta_{d,0} < \theta_{d,1}$. Then a unilateral trade liberalization in Home increases within-firm wage inequality in Home, according to the following inequality measures: Variance of log wages, Theil Index and Mean Log Deviation.*

The variance of log wages. In the model, different firms select different performance-pay contracts to reward their employees. This generates heterogeneity in wage distributions across firms, and thus inequality measures will crucially depend on the equilibrium allocation of workers to firms. In particular, the variance of log wages depends on the employment distribution and on the mean and variance of the firm-level log wage distributions, denoted $E(\widetilde{w}_i|\theta)$ and $Var(\widetilde{w}_i|\theta)$,

[31] 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 (20).

[32] For example, Lemieux (2006), Helpman et al. (2017) and Card et al. (2013) use variance decompositions of log wages to analyze changes in inequality in the U.S., Brazil and Germany, respectively.

[33] Under Pareto firm productivity, Helpman et al. (2010) show that between-firm inequality when only some firms export is higher than in both autarky and free trade (see Proposition 3), for all inequality measures that respect second-order stochastic dominance and scale independence. However, no analytical results are provided for changes in variable trade costs between two equilibria in which only some firms export. In Coşar et al. (2016), the effect of increased openness on between-firm inequality is determined by two countervailing forces: increasing wage dispersion across firms and worker reallocations towards high productivity firms. Their model predicts "little if any effect of increased openness" on between-firm inequality.

respectively, where $\widetilde{w}_i \equiv \log w_i$. Given $G_{h,j}(\theta)$, these two moments can be integrated across firms to obtain the standard decomposition of the total variance of log wages into between-firm and within-firm components in equilibrium j. This yields,

$$Var_j = Var_j^{between} + Var_j^{within}, \tag{31}$$

where

$$Var_j^{between} = \int_{\theta_L}^{\theta_H} \left[E(\widetilde{w}_i|\theta) - \widetilde{w}_j^* \right]^2 dG_{h,j}(\theta),$$

$$Var_j^{within} = \int_{\theta_L}^{\theta_H} Var(\widetilde{w}_i|\theta) dG_{h,j}(\theta),$$

and $\widetilde{w}_j^* \equiv \int_{\theta_L}^{\theta_H} E(\widetilde{w}_i|\theta) dG_{h,j}(\theta)$ is the aggregate mean log wage in equilibrium j.

As in 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, while in this model firms compensate their workers for exerting costly effort.

Unlike other models in the literature, however, part of the wage variation arises from differences in firm-level wage inequality across firms. As long as worker performance is only a noisy signal of effort, firms deal with the moral hazard problem by paying for performance. This implies $Var(\widetilde{w}_i|\theta) > 0$ for all active firms; i.e., positive firm-level wage dispersion. Moreover, firm-level wage variances vary across firms. Under complementarity between productivity and effort, high productivity firms offer higher-powered incentives that magnify idiosyncratic differences in performance and generate higher wage inequality among co-workers. In particular, part (b) of Corollary 1 and Proposition 2 imply that $Var(\widetilde{w}_i|\theta)$ increases in firm productivity even when the variance of idiosyncratic performance is identical in every firm. In the absence of quality upgrading associated to exporting, however, firm-level wage distributions are independent of the variable trade cost τ (Corollary 2). In this case, cross-firm variation in inequality is a necessary ingredient for trade liberalization to have an impact on within-firm inequality.

Next, I show that, in combination with the stronger form of labor reallocations implied by Assumption 2, this mechanism generates increasing within-firm wage inequality. The change in the between-firm variance, however, cannot be signed without imposing more structure on the distribution of firm productivity.

Formally, let subscripts 0 and 1 denote outcomes corresponding to equilibria before and after trade liberalization, respectively. Consider first the change in the within-firm variance,

$$\begin{aligned}
\Delta Var^{within} &= \int_{\theta_L}^{\theta_H} Var(\widetilde{w}_i|\theta) \left[dG_{h,1}(\theta) - dG_{h,0}(\theta) \right], \\
&= \int_{\theta_L}^{\theta_H} \frac{dVar(\widetilde{w}_i|\theta)}{d\theta} \left[G_{h,0}(\theta) - G_{h,1}(\theta) \right] d\theta, \\
&> 0.
\end{aligned}$$

The first line holds because firm-level wage distributions are independent of the variable trade cost. The second line requires integration by parts. Part (b) of Corollary 1 and Proposition 2 imply that the firm-level variance increases in θ, thus $dVar(\widetilde{w}_i|\theta)/d\theta > 0$. Moreover, Proposition 4 implies $G_{h,0}(\theta) \geq G_{h,1}(\theta)$ for all θ, with strict inequality for some θ. Intuitively, under Assumption 2, trade liberalization generates strong labor reallocations towards high inequality firms, resulting in an unambiguous increase in the residual variance of log wages.[34]

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

$$\Delta Var^{between} = 2\int_{\theta_L}^{\theta_H} \frac{dE(\widetilde{w}_i|\theta)}{d\theta} E(\widetilde{w}_i|\theta)\left[G_{h,0}(\theta) - G_{h,1}(\theta)\right]d\theta - \left[\left(\widetilde{w}_1^*\right)^2 - \left(\widetilde{w}_0^*\right)^2\right],$$

after integrating by parts. However, $\Delta Var_j^{between}$ cannot, in general, be signed. Note that the firm-level mean log wage is not necessarily increasing in productivity.[35] Even if it were, labor reallocations towards high productivity firms would then imply an increase in the aggregate mean log wage, $\widetilde{w}_1^* > \widetilde{w}_0^*$, that tends to reduce the between-firm variance in the aftermath of trade liberalization.

Lorenz-consistent inequality measures. 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).[36] The latter, however, incorporates some principles that are generally regarded as fundamental to the theory of inequality measurement. For this reason, I close this section by analyzing the impact of trade liberalization on two Lorenz-consistent and additively separable measures, the Theil index (T) and the MLD of wages.[37]

The definition and decomposition of the MLD and Theil measures are provided in section A.13 of the Appendix. Importantly, within-firm inequality has a similar structure in both measures, which also resembles the structure of the variance of log wages. In particular, for measure $I = \{MLD, T\}$ in equilibrium j,

$$I_j^{within} = \int_{\theta_d}^{\theta_H} I(w_i|\theta)dG_{\ell(I),j}(\theta),$$

where $\ell(I) = h$ if $I = MLD$ and $\ell(I) = w$ if $I = T$.

The impact of trade liberalization on ΔI^{within} is then evaluated as in the case of ΔVar^{within}. Part (b) of Corollary 1 and Proposition 2 ensure $dI(w_i|\theta)/d\theta > 0$. Moreover, Propositions (4)

[34] Recall that, by Lemma 2, Assumption 2 is not needed when the initial equilibrium is autarky.

[35] Intuitively, expected firm-level wages increase in productivity but so does wage dispersion. These two forces operate in opposite directions on average log wages, since the log transformation is both increasing and concave. When productivity is high enough, the mean log wage decreases in θ.

[36] 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.

[37] As members of the generalized entropy class, these measures have several desirable properties. Theorem 5 in Shorrocks (1980) shows that an inequality measure simultaneously satisfies the weak principle of transfers, decomposability, scale independence and the population principle only if it belongs to the class of generalized entropy measures. Moreover, Shorrocks (1980) points out that MLD and T enjoy two analytical advantages relative to any other generalized entropy measure. First, the total within-firm contribution to inequality is a weighted average of inequality across firms only for MLD and T. Second, the decomposition coefficients are independent of the between-group contribution only for MLD and T.

and (5) imply $G_{\ell(I),0}(\theta) \geq G_{\ell(I),0}(\theta)$ for $I = \{MLD, T\}$ and all θ, with strict inequality for some θ.

As anticipated, the effect of trade liberalization on between-firm inequality cannot be signed without further assumptions. Still, some progress can be made under the assumption that productivity follows an unbounded Pareto distribution. In this case, it is straightforward to verify that between-firm inequality in the open economy when all firms export is the same as in autarky, according to both the MLD and T measures. This result is reminiscent of Proposition 3(ii) in Helpman et al. (2010).

7. Patterns of wage dispersion and performance pay

This section employs the 2003 Workplace and Employee Survey, a matched employer–employee dataset from Canada, to document a set of empirical facts that motivate the theory developed in the paper. The survey was designed and implemented by Statistics Canada. Background information and further details on the dataset are provided in Appendix B. Here I briefly discuss the main features of the data.

The baseline sample for the empirical analysis is composed of firms with at least two matched, full-time, adult employees. I exclude non-profit firms and workers with missing values in the vector of individual characteristics (described below). The sample consists of 14,265 workers and 3,540 firms.[38]

The wage of worker i, denoted w_i, is the average weekly wage before taxes and deductions and net of overtime payments. Performance pay, denoted pp_i, is computed as weekly-equivalent tips, commissions, piecework payments and bonuses received by the worker. The performance-independent component, denoted $fixed_i$, is computed as the weekly wage net of performance pay. By definition, $w_i = fixed_i + pp_i$. Worker i is employed in a performance-pay job (PP job) if $pp_i > 0$.

The vector of observable characteristics of i, denoted e_i, contains 14 industry dummies, 47 occupation dummies, tenure with current employer, a full set of interactions between 5 education dummies and 4 experience dummies, and indicators for the following binary variables: union membership, gender, language mismatch between home and work, and foreign-born worker. The analysis also employs firm characteristics, including total annual revenue, total employment of full-time workers and export status. Summary statistics for workers and firms are reported in Tables B-I and B-II of Appendix B, respectively.

I start by documenting the importance of within-firm inequality in Canada. Next, I quantify the prevalence of performance pay and its contribution to wage inequality and the within-firm component. Finally, I report several empirical findings that support key cross-sectional implications of the model: (i) high wage firms are typically high inequality firms; and (ii) both average wage and firm-level wage inequality are increasing in firm size and independent of export status.

[38] While the WES provides a unique opportunity to analyze performance pay in a nationally representative sample, it features a relatively low number of matched employees per firm. I attempt to address potential concerns with the precision of inequality measurements at the firm level in two ways. First, I apply a finite population correction to construct unbiased estimates of firm-level log wage variances. The adjustment acknowledges that the sample of workers in each firm is drawn from a finite population (total employment in the firm) without replacement. Appendix B contains the details. Second, I verify that key findings hold when the sample is restricted to firms with at least 5 matched, full-time, adult workers. The latter actually results in higher estimated shares of within-firm inequality.

Between-firm and within-firm inequality. To gauge the relative sizes of between-firm and within-firm inequality across Canadian workers, I decompose raw and residual wage dispersion following equation (31). Within-firm inequality accounts for 35% of the variance of log weekly wages in the sample.[39] Interestingly, the share of within-firm inequality rises when decomposing residual wage dispersion. Consider the residuals obtained from a linear regression of log weekly wages on worker observables e_i, industry and occupation fixed effects. In this case, within-firm inequality accounts for 45% of the variance of residual log wages.

Table B-IV in Appendix B shows that quantitatively similar results are obtained when decomposing the Theil index and the mean log deviation of raw and residual wages.[40]

Performance pay: prevalence and contribution to inequality. The data reveal two salient facts about performance pay in Canada. First, performance pay is widespread in the labor market. In 2003, 46% of the workers in the sample were employed in PP jobs.[41] Table B-III in Appendix B dissects these figures by segments of the wage distribution, sector, occupation and gender, confirming in all cases that PP jobs span a broad cross-section of the workforce. Second, in line with the findings of Lazear and Shaw (2007) for the U.S., the data suggest that performance pay is most often determined by individual, rather than team, performance.[42]

A simple way of assessing the contribution of performance pay to wage inequality is by comparing wage decompositions in the baseline sample and the subsample of non-PP jobs. Eliminating all PP jobs from the baseline sample yields a reduction of 17% in the variance of log wages, operating largely through within-firm rather than between-firm inequality (41% and 0.4% reductions, respectively). Repeating this exercise for residual wages yields a reduction of 14% in the variance of residual log wages, reflecting a large fall in within-firm inequality (40%) and a small increase in between-firm inequality (0.1%).

High wage firms are high inequality firms. In line with the implications of Corollary 1, the data reveal a clear pattern across firm-level wage distributions: high wage firms are typically high inequality firms. The correlation between the mean wage, $E[w|\theta]$, and the variance of log wages, $Var\left[\log w|\theta\right]$, across firms is 0.4, statistically significant at the 1% level.[43]

This pattern is robust to controlling for differences in the composition of observable skills. To show this, I estimate linear approximations to the mean wage level and the variance of log wages of the conditional wage distribution for workers with identical observable skills employed in firm θ:

$$E[w_i|e_i,\theta] \approx \phi_1 e_i + \psi_{\theta,1}, \tag{32a}$$

[39] I implement this decomposition by regressing log weekly wages on firm fixed-effects. The share of within-firm inequality in the variance of log wages is one minus the R-squared of this regression. Not surprisingly, the result is sensitive to the number of matched workers per firm. In the subsample of firms with at least five matched workers, the share of within-firm inequality in the variance of log wages is 50%.

[40] In addition, section B.3.3 in the Appendix shows that the share of within-firm inequality in total log wage inequality within sectors and occupations in Canada echoes inequality decompositions in Brazil, Sweden and France documented in recent studies.

[41] Note that this figure likely underestimates the true prevalence of PP jobs since observations for which realized performance pay is zero are coded as non-PP jobs.

[42] Indeed, 33% of firms in the sample report having explicit ‘individual incentive systems’ that reward employees on the basis of individual output or performance. In contrast, only 10% of the firms in the sample have explicit ‘group incentive systems’ that reward employees on the basis of group output or performance.

[43] In the subsample of firms with at least five matched workers, the correlation increases to 0.5 (1% level).

$$Var\left[\log w_i|e_i,\theta\right]\approx\phi_2 e_i+\psi_{\theta,2}, \tag{32b}$$

where i indexes workers and e_i is the vector of observable characteristics defined at the beginning of this section. I estimate $(\phi_1, \psi_{\theta,1})$ in (32a) by regressing w_i on e_i and firm fixed effects. For (32b), I first regress $\log w_i$ on e_i and firm fixed effects. The squared residuals obtained from this regression are subsequently regressed on e_i and firm fixed effects to estimate $(\phi_2, \psi_{\theta,2})$.[44] The results again reveal that high wage firms are also typically high inequality firms: the correlation between the estimated $\psi_{\theta,1}$ and $\psi_{\theta,2}$ across firms is 0.3, statistically significant at the 1% level.[45] Figure B-1 in Appendix B illustrates these patterns, while Figure B-2 verifies their robustness when inequality is measured by the Theil index or the MLD of wages.

Performance pay, firm size and export status. Next, I analyze the variation of pp_i across and within firms, restricting attention to PP jobs. This approach has merits and limitations. Among the former, I focus on wage variation that is, *by definition*, linked to performance pay, precluding the possibility that the empirical patterns I document are driven purely by unobserved skill heterogeneity unrelated to performance pay. The flip side is that this approach does not allow me to identify exactly how performance pay operates at the firm level. In principle, performance-pay contracts can provide incentives for workers to exert more effort or serve as a screening device (or both) in sorting heterogeneous workers across firms. Still, both forms of performance pay *will* generate wage dispersion among identical co-workers under arguably mild conditions.[46]

In search of systematic patterns of variation in performance pay across firms, I focus on two observable characteristics of firms that are singled out in Corollary 2: firm size and export status. In particular, I estimate the following linear approximations to the probability of observing positive performance pay, the mean level and the variance of logs of the conditional distribution of performance pay for workers with identical observable skills in firm θ:

$$P\left[pp_{i\theta}>0|Size_\theta, Ex_\theta, e_i\right]\approx\delta_A Size_\theta+\zeta_A Ex_\theta+\phi_A e_i, \tag{33a}$$

$$E\left[pp_{i\theta}|Size_\theta, Ex_\theta, e_i\right]\approx\delta_B Size_\theta+\zeta_B Ex_\theta+\phi_B e_i, \tag{33b}$$

$$Var\left[\log pp_{i\theta}|Size_\theta, Ex_\theta, e_i\right]\approx\delta_C Size_\theta+\zeta_C Ex_\theta+\phi_C e_i, \tag{33c}$$

where $pp_{i\theta}$ is performance pay of worker i employed in firm θ, $Size_\theta$ is the natural log of total annual revenue in θ, Ex_θ is a dummy equal to one if θ exported in 2003 and e_i is the vector of i's observable characteristics.

Columns (1) to (3) of Table 1 report estimates of $(\delta_A, \zeta_A, \phi_A)$, obtained by regressing $I_{[pp_{i\theta}>0]}$, an indicator function of positive performance pay, on $Size_\theta$, Ex_θ and e_i using the full sample of workers. In turn, columns (4) to (6) report estimates of $(\delta_B, \zeta_B, \phi_B)$, obtained by regressing $pp_{i\theta}$ on $Size_\theta$, Ex_θ and e_i in the sample of workers employed in PP jobs. For (33c), I first regress $\log pp_{i\theta}$ on $Size_\theta$, Ex_θ and e_i. The squared residuals obtained from this regression are subsequently regressed on $Size_\theta$, Ex_θ and e_i to estimate $(\delta_C, \zeta_C, \phi_C)$; the results are presented in columns (7) to (9). In all cases, standard errors are clustered at the firm level.

44 The finite population correction is applied to $\psi_{\theta,2}$.

45 In the subsample of firms with at least five matched workers, the correlation is 0.5 (1% level).

46 It suffices to assume that individual performance is not a deterministic function of effort. To fix ideas, it is useful to think of pp_i as a firm-specific function or contract $pp_i=g(s_i, z_i, \theta)$ of worker i's type s_i (defined by observed and unobserved characteristics) and unobserved individual performance z_i. The dependence of pp_i on s_i allows performance pay to operate as a screening device, a mechanism that is absent in the model. Type s_i can, in principle, also affect pp_i indirectly through z_i. However, as long as z_i is imperfectly correlated with s_i (e.g. if performance is a stochastic function of type-specific effort), then $Var(pp_i|s_i,\theta)>0$.

Table 1
Performance pay across firms.

Outcome	$P\left[pp_{i\theta} > 0 \vert \cdot\right]$			$E\left[pp_{i\theta} \vert \cdot\right]$			$Var\left[\log pp_{i\theta} \vert \cdot\right]$		
Dep. Var.	$I_{\left[pp_{i\theta}>0\right]}$			$pp_{i\theta}$			See table notes		
	Basic (1)	Add Exp Status (2)	Add Controls (3)	Basic (4)	Add Exp Status (5)	Add Controls (6)	Basic (7)	Add Exp Status (8)	Add Controls (9)
Firm size	0.042	0.046	0.064	29.654	30.049	31.622	0.185	0.188	0.167
	(0.009)[a]	(0.010)[a]	(0.010)[a]	(11.107)[a]	(10.056)[a]	(9.224)[a]	(0.034)[a]	(0.038)[a]	(0.034)[a]
Exporter		−0.048	−0.012		−4.701	24.400		−0.042	−0.033
		(0.037)	(0.037)		(41.410)	(49.579)		(0.122)	(0.119)
Controls	No	No	Yes	No	No	Yes	No	No	Yes
R-sq	0.01	0.02	0.13	0.01	0.01	0.17	0.03	0.03	0.10
Obs	14,265	14,265	14,265	6,775	6,775	6,775	6,775	6,775	6,775

Notes: this table reports OLS estimates of the right-hand side parameters of (3a) in columns 1 to 3, (3b) in columns 4 to 6 and (3c) in columns 7 to 9. 'Firm size' is the natural log of firm total revenue. 'Exporter' is a dummy equal to 1 if the firm exports. 'Controls' is a vector of worker characteristics that includes industry and occupation fixed effects. Standard errors (in parentheses) are clustered at the firm level.

[a] Statistical significance at the 1% level.

For each of the three outcomes considered, Table 1 sequentially introduces firm size, export status and worker characteristics into the analysis. Firm size is positively correlated with all three outcomes: the probability of observing positive performance pay, the mean level of performance pay and the variance of log performance pay increase with firm total revenue. The estimates are significant at the 1% level, with and without worker-level controls, including industry and occupation fixed effects. On the other hand, there is no evidence of exporter premia conditional on firm size for any of the outcome variables, although the coefficients are imprecisely estimated.[47] Importantly, the estimates in Table 1 should not be given a causal interpretation. In fact, in the model, firm size, mean and variance of log performance pay are functions of firm productivity. Still, the theory predicts that these variables should indeed be positively correlated.

Echoing the findings in Table 1, Table B-VII in Appendix B reports a strong statistical relationship between firm size and wage distributions across firms. In particular, in line with Corollary 2, both the mean wage level and the variance of log wages are positively correlated with firm size at the 1% level, with and without controls for worker observables.

Altogether these findings suggest that firm productivity plays a role in explaining why high wage firms are typically high inequality firms. Moreover, performance pay emerges as an empirically relevant channel through which firm heterogeneity shapes inequality patterns.

8. Concluding remarks

This paper studies residual wage inequality due to performance pay in a tractable general equilibrium model with heterogeneous firms. The theory is consistent with empirical patterns in wage dispersion and performance pay across firms in Canada. As an application, the paper provides an analytical characterization of the impact of trade liberalization on within-firm inequality, highlighting a new channel through which international trade can contribute to residual wage dispersion.

[47] Table B-VI in Appendix B reports similar results using an alternative proxy for firm size, full-time employment.

Within-firm inequality is an employment-weighted average of firm-level residual wage inequality measures. Empirically, it accounts for a typically substantial, yet largely neglected, $1 - R$-squared in a linear regression of log wages on worker characteristics and firm fixed-effects. The paper tackles its analysis by focusing on wage dispersion among *homogeneous* co-workers. An alternative theoretical approach is to think of within-firm inequality as arising from skill heterogeneity that is observable to firms and workers but not to econometricians. The latter can immediately rationalize the existence of residual wage dispersion among co-workers in the data, i.e., R-squared < 1. Moreover, it is not possible to rule out the existence of unobserved skill heterogeneity and measurement error in residual wages. Paradoxically, this lack of discipline is arguably the main limitation of pursuing the alternative approach. What kind of empirical evidence would lead researchers to reject a theory of wage dispersion based on unconstrained and unobservable heterogeneity? How would observable skill dispersion in such models map into residual wage dispersion in the data?

The main virtue of the approach followed in the paper is that wage dispersion generated by the model can be readily interpreted as residual, and its predictions contrasted with the data. A key takeaway of the analysis is that, far from being an unfathomable regression residual, within-firm inequality can be driven by systematic patterns in wage and performance pay distributions across firms. As shown in the paper, these patterns exist in the data, inform theory and open new avenues to study the determinants of wage inequality.

Appendix. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jet.2017.10.001.

References

Akerman, A., Helpman, E., Itskhoki, O., Muendler, M.-A., Redding, S., 2013. Sources of wage inequality. Am. Econ. Rev. Pap. Proc. 103 (3), 214–219.

Amiti, M., Davis, D., 2011. Trade, firms, and wages: theory and evidence. Rev. Econ. Stud. 79 (1), 1–36.

Barth, E., Bratsberg, B., Hægeland, T., Raaum, O., 2012. Performance pay, union bargaining and within-firm wage inequality. Oxf. Bull. Econ. Stat. 74 (3), 327–362.

Bloom, N., Guvenen, F., Price, D.J., Song, J., Von Wachter, T., 2016. Firming up inequality. Mimeo. Stanford University.

Bloom, N., Van Reenen, J., 2007. Measuring and explaining management practices across firms and countries. Q. J. Econ. 122 (4), 1351–1408.

Bound, J., Johnson, G., 1992. Changes in the structure of wages in the 1980's: an evaluation of alternative explanations. Am. Econ. Rev. 82 (3), 371–392.

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

Card, D., Heining, J., Kline, P., 2013. Workplace heterogeneity and the rise of west German wage inequality. Q. J. Econ. 128 (3), 967–1015.

Coşar, A.K., Guner, N., Tybout, J., 2016. Firm dynamics, job turnover, and wage distributions in an open economy. Am. Econ. Rev. 106 (3), 625–663.

Cuñat, V., Guadalupe, M., 2009. Globalization and the provision of incentives inside the firm: the effect of foreign competition. J. Labor Econ. 27 (2), 179–212.

Davidson, C., Matusz, S., Shevchenko, A., 2008. Globalization and firm level adjustment with imperfect labor markets. J. Int. Econ. 75 (2), 295–309.

Davis, D., Harrigan, J., 2011. Good jobs, bad jobs, and trade liberalization. J. Int. Econ. 84 (1), 26–36.

Demidova, S., Rodriguez-Clare, A., 2013. The simple analytics of the Melitz model in a small open economy. J. Int. Econ. 90 (2), 266–272.

Di Giovanni, J., Levchenko, A.A., Ranciere, R., 2011. Power laws in firm size and openness to trade: measurement and implications. J. Int. Econ. 85 (1), 42–52.
Egger, H., Kreickemeier, U., 2009. Firm heterogeneity and the labor market effects of trade liberalization. Int. Econ. Rev. 50 (1), 187–216.
Foster, J., Ok, E., 1999. Lorenz dominance and the variance of logarithms. Econometrica 67 (4), 901–907.
Garicano, L., Rossi-Hansberg, E., 2004. Inequality and the organization of knowledge. Am. Econ. Rev. Pap. Proc. 94 (2), 197–202.
Garicano, L., Rossi-Hansberg, E., 2006. Organization and inequality in a knowledge economy. Q. J. Econ. 121 (4), 1383–1435.
Helpman, E., Itskhoki, O., Muendler, M., Redding, S., 2017. Trade and inequality: from theory to estimation. Rev. Econ. Stud. 84 (1), 357–405.
Helpman, E., Itskhoki, O., Redding, S., 2010. Inequality and unemployment in a global economy. Econometrica 78 (4), 1239–1283.
Holmström, B., 1979. Moral hazard and observability. Bell J. Econ. 10 (1), 74–91.
Holmström, B., 1982. Moral hazard in teams. Bell J. Econ. 13 (2), 324–340.
Holmström, B., Milgrom, P., 1987. Aggregation and linearity in the provision of intertemporal incentives. Econometrica 55 (2), 303–328.
Katz, L.F., Murphy, K.M., 1992. Changes in relative wages, 1963–1987: supply and demand factors. Q. J. Econ. 107 (1), 35–78.
Kugler, M., Verhoogen, E., 2012. Prices, plant size, and product quality. Rev. Econ. Stud. 79 (1), 307–339.
Lazear, E.P., 1986. Salaries and piece rates. J. Bus. 59 (3), 405–431.
Lazear, E., 2000. Performance pay and productivity. Am. Econ. Rev. 90 (5), 1346–1361.
Lazear, E., Shaw, K., 2007. Personnel economics: the economist's view of human resources. J. Econ. Perspect. 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? Am. Econ. Rev. 96 (3), 461–498.
Lemieux, T., MacLeod, W., Parent, D., 2009. Performance pay and wage inequality. Q. J. Econ. 124 (1), 1–49.
Manova, K., Zhang, Z., 2012. Export prices across firms and destinations. Q. J. Econ. 127 (1), 379–436.
Melitz, M.J., 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. Econometrica 71 (6), 1695–1725.
Oyer, P., 2004. Why do firms use incentives that have no incentive effects? J. Finance 59 (4), 1619–1650.
Oyer, P., Schaefer, S., 2005. Why do some firms give stock options to all employees?: An empirical examination of alternative theories. J. Financ. Econ. 76 (1), 99–133.
Parent, D., 1999. Methods of pay and earnings: a longitudinal analysis. Ind. Labor Relat. Rev. 53 (1), 71–86.
Raith, M., 2003. Competition, risk, and managerial incentives. Am. Econ. Rev. 93 (4), 1425–1436.
Shorrocks, A.F., 1980. The class of additively decomposable inequality measures. Econometrica 48 (3), 613–626.
Sun, Y., 2006. The exact law of large numbers via Fubini extension and characterization of insurable risks. J. Econ. Theory 126 (1), 31–69.
Verhoogen, E., 2008. Trade, quality upgrading, and wage inequality in the Mexican manufacturing sector. Q. J. Econ. 123 (2), 489–530.
Wolfstetter, E., 1999. Topics in Microeconomics: Industrial Organization, Auctions, and Incentives. Cambridge University Press.