Performance Pay, Trade and Inequality

Germán P. Pupato*
Ryerson University

December 2015

Abstract

This paper introduces moral hazard into a standard general equilibrium model with heterogeneous firms, to study the impact of trade liberalization on wage inequality between homogeneous workers. Trade liberalization operates on two margins of inequality, generating between- and within-firm wage dispersion. While the former channel has been the focus of numerous recent papers, the latter has remained largely overlooked in the literature. In the model, within-firm wage dispersion increases in firm productivity due to differential intensity in optimal performance pay across firms. International trade liberalization triggers labor reallocations towards high productivity firms that increase within-firm inequality. To motivate the theory, the paper documents cross-sectional patterns in residual wage dispersion and performance pay across and within firms using nationally representative, matched employer-employee data for Canada.

KEYWORDS: wage inequality, performance pay, moral hazard, heterogeneous firms, trade liberalization.

1 Introduction

Our understanding of the impact of international trade on wage inequality has evolved substantially over the last twenty years. In the early 1990's, most economists dismissed the role of

*This research was supported by the Social Sciences and Humanities Research Council of Canada. I am grateful to Alex Alvarez, Costas Arkolakis, Kyle Bagwell, 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 discussions. 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.

trade as a driving force behind the steep increases in wage inequality that had been observed in many countries around the world since the late 1970's. Standard factor proportions theory was not easily reconcilable with increasing inequality in developing countries, the absence of significant reallocations of labor across industries and evidence showing that standard human capital variables, such as education and experience, could account for only minor shares of the level and growth of inequality in both developed and developing countries.[1]

In recent years, however, a new generation of trade models has caught up to these empirical challenges by shifting its focus from industries to firms, as the basic units of analysis. This research agenda has been fueled by numerous studies documenting a set of stylized facts regarding heterogeneity in firm-level outcomes within industries, including systematic differences between exporting and non-exporting firms. Several recent trade theories are motivated by the empirical finding that more productive firms pay higher wages, on average, even after controlling for worker characteristics such as education, experience, occupation and industry affiliation.[2] Using Brazilian data, Helpman et al. (2014) report that 37% of the variance of log wages within sector-occupation cells in 1990 is accounted for by the variation in wage premia across firms. These facts are compatible with models of firm heterogeneity featuring search frictions and bargaining (Davidson et al. (2008), Helpman et al. (2010), Coçar et al. (forthcoming)), efficiency wages (Verhoogen (2008), Davis and Harrigan (2011)) or fair wage constraints (Egger and Kreickemeier (2009), Amiti and Davis (2011)), in which ex-ante identical workers receive higher wages in more productive firms and wages are systematically related to the export status of the firm.

With the exception of Verhoogen (2008) (discussed below), however, workers employed in the same firm receive identical wages in these models. This literature therefore cannot elucidate an equally sizable component of residual wage inequality (34%) reported in Helpman et al. (2014), namely, within-firm wage dispersion. This evidence is corroborated in recent empirical studies in the United States and several European countries, collected in Lazear and Shaw (2008). Summarizing the findings, they report that within-firm wage variation ranges from 60% to 80% of the total wage dispersion in each of those countries.[3]

The purpose of this paper is to develop a theoretical framework to study this important and largely unexplored dimension of wage inequality, emphasizing its links to international trade. To do so, I extend 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 identical co-workers as firms pay for performance to align the incentives of employees with their best interests. In particular, I study a sequential pro-

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

[2]Evidence of size and exporter wage premia is reported in Bernard and Jensen (1995), Amiti and Davis (2011) and Helpman et al. (2014) for US, Indonesian and Brazilian firms, respectively.

[3]Lazear and Shaw (2008), page 6.

duction process during which workers stochastically make mistakes that are detrimental to product quality. Workers can reduce the frequency of their mistakes by exerting costly effort at each production task. Firms, in turn, monitor the performance of their workers, observing outcomes (mistakes/successes) but not inputs (effort choices). Importantly, as the frequency of tasks increases, individual performance converges to a Brownian process. This feature of the model allows for a tractable characterization of optimal performance-pay contracts that builds on the seminal work of Holmström and Milgrom (1987).

Second, I introduce cross-firm differences in optimal contracting strategies by allowing for complementarity between firm productivity and the performance of workers in generating product quality. Each firm designs a set of contracts, providing incentives to implement desired effort levels. Because high productivity firms have a comparative advantage in generating quality, they find it optimal to 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. Importantly, this pattern of inequality across firms is entirely driven by differences in endogenous contracting strategies since, conditional on effort, the stochastic component of worker performance is invariant across firms.

Heterogeneity in performance-pay compensation generates implications for residual wage inequality that remain unexplored in the trade literature.[4] To illustrate these, consider the variance of log wages in any one of the two countries in the model, denoted $Var(\widetilde{w})$. The latter can be decomposed as

$$Var(\widetilde{w}) = Var\left[E(\widetilde{w}|\theta)\right] + E\left[Var(\widetilde{w}|\theta)\right], \tag{1}$$

where θ indexes the set of active firms in a given equilibrium. $E(\widetilde{w}|\theta)$ and $Var(\widetilde{w}|\theta)$ denote the mean and variance of log wages across workers employed in firms with productivity θ, respectively. In turn, $Var\left[\cdot\right]$ and $E\left[\cdot\right]$ integrate over the equilibrium distribution of workers across firms. According to (1), the total log wage variance is the sum of (i) the variance of firm-level mean log wages (*between*-firm inequality) and (ii) the mean of firm-level log wage variances (*within*-firm inequality).[5] Recent theoretical studies link trade liberalization to residual wage inequality through mechanisms that operate exclusively on the between-firm component of wage inequality, in which firms of different sizes pay different wages to identical

[4] To the extent that data on worker-specific performance histories at the firm level remain hidden to econometricians, wage variation generated by this model should, from an empirical perspective, be understood as residual (i.e. wage variation across workers of identical observable characteristics such as education, experience, occupation, industry affiliation, etc).

[5] Throughout the paper, *within-firm inequality* refers to the second term on the right-hand side of (1), or the analogous terms that arise from additively decomposable measures of wage inequality. In turn, *firm-level inequality* refers to wage inequality among co-workers in a given firm, such as $Var(\widetilde{w}|\theta)$.

workers but there is no wage dispersion inside firms. The model developed in this paper is, to the best of my knowledge, the first to link trade and residual wage inequality through both channels.

More specifically, the key features and implications of the model regarding the effect of international trade on wage inequality are summarized as follows:

(a) Performance pay generates wage dispersion within firms. By punishing or rewarding employees according to their performance, high-powered incentives amplify the effect of idiosyncratic performance on wages. This implies $Var(\widetilde{w}|\theta) > 0$ in every firm θ.

(b) Different firms design different performance-pay contracts, generating cross-firm variation in the first and second moments of firm-level wage distributions. In particular, more productive firms offer higher-powered incentives and hence $Var(\widetilde{w}|\theta)$ increases in θ. 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. This generates variation in $E(\widetilde{w}|\theta)$ across firms that, according to (1), translates into positive between-firm inequality.

(c) International trade liberalization (i.e., a reduction in bilateral variable trade costs) shapes the distribution of workers across firms through general equilibrium reallocations of labor towards high productivity firms. Specifically, under relatively mild restrictions on the distribution of firm productivity, I show that the distribution of workers across firms in a post-liberalization equilibrium first-order stochastically dominates the corresponding pre-liberalization equilibrium distribution.

The main result of the paper shows that a bilateral trade liberalization leads to monotonic increases in within-firm inequality in both countries, as it reallocates workers from low-inequality firms to high-inequality firms. This is valid for symmetric countries and for three additively decomposable inequality measures: the variance of log wages, the Theil index and the mean log deviation. 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.[6] In this case, within-firm inequality increases in the liberalizing country. As in Helpman et al. (2010) and Coçar et al. (forthcoming), however, the effects on between-firm inequality are non-monotonic and difficult to characterize analytically, even in symmetric equilibria.[7]

[6] A special case in which a unilateral trade liberalization indeed triggers firm selection in asymmetric equilibria is when the liberalizing country is a small open economy (as defined in Demidova and Rodriguez-Clare (2013)).

[7] 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 Coçar et al. (forthcoming), the effect of increased openess on between-firm inequality is determined by

To spark further interest in the topic and theoretical approach of this paper, I present empirical evidence from the 2003 Workplace and Employee Survey (WES), a matched employer-employee dataset for Canada. Within-firm inequality is a salient feature of the data. Conditional on worker observables (including education, experience, gender, union membership and tenure), the magnitude of within-firm wage inequality is approximately 80% of the magnitude of between-firm wage inequality, and accounts for almost 40% of the total variance of log weekly wages within industries and occupations in Canada. Moreover, to motivate the key driver of within-firm inequality in this paper, I rely on high quality information on performance pay for a nationally representative sample of workers, a unique feature of the WES. In 2003, 46% of the workers in the sample received strictly positive commissions, piece-rate payments or bonuses.[8] While performance pay accounts, on average, for a relatively modest fraction of individual wages, it makes a substantial contribution to overall wage inequality. For example, for workers receiving any form of performance pay, the latter represents 12% of weekly wages. Across all workers, however, the standard deviation of performance pay is two thirds of the standard deviation of performance-independent compensation.

In the absence of a clean episode of trade liberalization (during the time frame of the WES) to test the main result of the paper, I motivate the theory by providing empirical support for key cross-sectional predictions of the model. In any given equilibrium, the theory has clear-cut implications regarding the cross-sectional variation in firm-level wage distributions. These predictions, in turn, play crucial roles in the mechanism that links trade liberalization and inequality in this paper. First, the data reveal substantial variation in firm-level wage inequality. For unconditional wages, the coefficient of variation of $Var(\widetilde{w}|\theta)$ is almost three times greater than the coefficient of variation of $E(w|\theta)$. This finding hints at the potential of labor reallocations across firms to shape wage inequality through the within-firm channel, in addition to the usual between-firm channel emphasized in previous studies. Second, a robust empirical pattern emerges from the heterogeneity in firm-level wage distributions: high-wage firms are typically high-inequality firms, according to several inequality measures and controlling for differences in observable workforce composition. In the theory, a positive correlation between $Var(\widetilde{w}|\theta)$ and $E(w|\theta)$ is a general outcome of the optimal contracting problem when firms face different returns to worker effort. Third, I document positive cross-firm correlations between firm size (employment or revenue), mean and inequality of performance pay, controlling for differences in observable workforce composition. Complementarity between

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.

[8] This figure is in line with empirical evidence in the United States. Using data from the PSID, Lemieux et al. (2009) show that the fraction of U.S. male workers on performance-pay jobs (i.e. workers earning strictly positive piece rates, commissions, or bonuses) increased from about 38% in the late 1970s to as much as 45% in the late 1990s. They also show that wages are less equally distributed on performance-pay than non performance-pay jobs and conclude that the growth of performance pay has contributed to about 25 percent of the increase in the variance of log wages between the late 1970s and the early 1990s.

effort and firm productivity in the model provides a simple rationale for these patterns.[9] Under this key property of technology, international trade liberalization triggers labor reallocations towards more productive, high-inequality firms, that result in monotonic increases in within-firm inequality.[10]

Naturally, some determinants of individual performance pay likely remain unmeasured in the data. In this case, dispersion in performance pay among co-workers with identical observable characteristics can, in principle, be rationalized by models in which a given firm offers either a single performance pay contract to homogeneous workers or a combination of contracts to workers of different, yet empirically unobservable, types. Performance pay, however, generates wage dispersion among identical co-workers in both models, under mild conditions discussed below. For conceptual clarity, I develop the mechanism linking trade and residual within-firm inequality in a framework without ex-ante skill heterogeneity. An exploration of the implications of this extension of the model is left for future work.

Related Literature. The effect of declining trade costs on inequality studied in this paper is distinct from, and complementary to, the mechanism studied in Verhoogen (2008). In the latter, an exchange-rate devaluation impacts firm-level wage variances in exporting firms, as the latter upgrade quality by paying relatively higher efficiency wages to (otherwise identical) workers employed in the export production line. Effort-wage schedules, however, are exogenous and a characterization of equilibrium changes in the distribution of workers across firms is not provided, thus preventing a general equilibrium analysis of the impact of trade on within-firm inequality.[11] The latter is the main goal of this paper.

Importantly, the theoretical results in this paper do not rely on the existence of trade-induced effects on firm-level wage distributions. In fact, in the model, there is no quality upgrading or downgrading associated to exporting and hence, given θ, $E(\widetilde{w}|\theta)$ and $Var(\widetilde{w}|\theta)$ do not change in response to trade liberalization. Heterogeneity in performance-pay con-

[9] This assumption has two additional implications in the model that receive empirical support. First, the intensity of performance pay increases in firm size. Bloom and Van Reenen (2007) report positive correlations between the extent to which firms reward performance and total revenue in the U.S., France, Germany, and the United Kingdom. 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.

[10] Evidence that trade liberalization induces market share reallocations towards high productivity firms is provided by Pavcnik (2002) and Trefler (2004), for Chile and Canada, respectively.

[11] These elements need not be essential for understanding firm-level responses to trade liberalization, the core of Verhoogen (2008). For present purposes, however, the importance of characterizing equilibrium changes in the distribution of workers across firms cannot be overstated. To illustrate this in a stark way, note that within-firm inequality in an economy could, in principle, *decrease* even in a situation in which firm-level wage variances *increase in every firm.* This would occur if trade liberalization induced labor reallocations towards firms with initially low firm-level wage variances. Observe that the latter are not necessarily low productivity, low effort, firms in Verhoogen (2008), since within-firm wage variances depend not only on the relative wage of high-effort workers but also on their employment shares. For example, for given wages, the firm-level variance of wages is non-monotonic in the share of high-effort workers and will decrease when the latter is sufficiently high.

tracts, however, ensures that reductions in variable trade costs will still increase within-firm inequality, purely through labor reallocations. Evidently, this mechanism will, in turn, be amplified by increases in firm-level variances driven by quality upgrading.[12]

There is a class of trade theories in which within-firm wage dispersion is driven by workforce composition, such as Bustos (2011), Monte (2011), Burstein and Vogel (2012), Caliendo and Rossi-Hansberg (2012), Emami Namini et al. (forthcoming) and Harrigan and Reshef (forthcoming). Workers are heterogeneous due to differences in skills or human capital. In these models, wages contain neither firm- nor match-specific components because they are determined in competitive labor markets.[13] Naturally, this literature has sought to analyze variation in skill premia rather than wage dispersion between identical workers. Firm-level wage dispersion in this class of models, however, can arguably be interpreted as residual if skill heterogeneity is assumed unobservable to the econometrician.[14] The implications of this approach for the link between trade and residual within-firm inequality have not yet been articulated. Unlike this literature, this paper explicitly seeks to analyze the effect of trade liberalization on the typical degree of firm-level inequality -defined by the second term on right-hand side of (1) or the analogous terms in additively decomposable measures of wage inequality. Moreover, the paper focuses on an empirically relevant channel, performance pay, that is not easily understood purely in terms of unobserved skill heterogeneity priced in competitive labor markets.

Wage dispersion among co-workers can also arise in dynamic models of firm heterogeneity with directed search, including Felbermayr et al. (2015) and Ritter (forthcoming). In the presence convex adjustment costs, workers with identical skills hired at different stages of a firm's life cycle earn different wages. At a given point in time, however, there is no wage dispersion among co-workers with identical tenure at the firm. These papers do not speak to cross-sectional variation in residual wages within firms that is conditional on employee tenure, such as the evidence presented in this paper.

A large empirical literature in labor economics and international trade studies sources of wage variation using matched employer-employee data. A number of recent papers, including Akerman et al. (2013), Card et al. (2013), Helpman et al. (2014), Barth et al. (2014) and Tito

[12] A footnote in page 25 discusses how quality upgrading in response to trade liberalization can be introduced in the model, along the lines of Verhoogen (2008). With this feature, firms in the model would optimally offer incentives to increase effort, leading to higher intensity of performance pay and firm-level wage inequality. Indeed, there is evidence that trade liberalization or greater openess lead to quality upgrading (Verhoogen (2008), Amiti and Khandelwal (2013) and Fan et al. (2015)), higher sensitivity of pay to performance (Cuñat and Guadalupe (2005, 2009)) and higher firm-level wage dispersion (Frías et al. (2012))

[13] In Bombardini et al. (2015), each firm hires workers of different skills but wages are bargained due to the existence of search frictions in the labor market. Identical co-workers earn identical wages.

[14] It is unclear whether a trivial reinterpretation of these models, in which skill heterogeneity is assumed fully unobservable to the econometrician, is likely to provide a useful lens to analyze residual wage dispersion empirically. After all, econometricians *can* predict individual skills, albeit imperfectly, using information routinely available in microdata sets (e.g. education, experience, occupation, industry, etc.). This complicates a direct interpretation of wage variation in these models as purely residual.

(2015), implement different versions of the decomposition (1) in Sweden, Germany, Brazil, U.S. and France, respectively.[15] Another branch of this literature documents variation in wage distributions across firms. Davis and Haltiwanger (1995) find that firm-level wage dispersion rises with firm size in the U.S. manufacturing industry. Lazear and Shaw (2008) report typically positive correlations between the standard deviation of log wages and the mean log wage across firms, for several countries and time periods.[16] The paper is also related to studies that investigate the link between performance pay and wage inequality. This literature, largely focused on top executive compensation in large firms, documents the key contribution of performance pay to the growth and dispersion of wages at the top end of the distribution (Piketty and Saez (2003) and Frydman and Saks (2010)). Lemieux et al. (2009) use PSID data to show that this phenomenon extends to a broader cross-section of the U.S. workforce. The paper contributes to this literature by documenting patterns of performance pay and wage inequality across firms, using nationally representative, matched employer-employee data.

Outline of the Paper. The next section presents the empirical evidence on residual wage dispersion and performance pay across firms in Canada. Section 3 introduces the theoretical framework, sequentially describing the timing of events, market structure, the production technology, the convergence of worker performance to a Brownian process, and consumer preferences. 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 the distribution of firm productivity and how labor is reallocated across firms. Section 7 analyzes the implications of the theory for wage inequality between and within firms. The final section discusses extensions and topics for future research. Appendix A contains proofs of lemmas, propositions and other theoretical results. Appendix B contains a description of the dataset and additional empirical results.

[15] These papers also report significant changes in within-firm inequality over time. Sharp increases are observed in Sweden, France and, to a minor extent, Germany. The opposite is true in Brazil and the U.S. I do not emphasize these findings because, unfortunately, they do not provide evidence either in favor or against the main result of this paper. The theory predicts that within-firm inequality should, ceteris paribus, increase in response to reductions in variable trade costs. The 'all else constant' condition, however, cannot be expected to hold in these countries during the periods analyzed in this literature.

[16] Figure I.8 in Lazear and Shaw (2008). Mean wage and wage inequality are always positively correlated across firms in the model. The latter, however, can still generate a negative correlation between mean *log* wage and wage inequality if inequality increases sufficiently fast in firm productivity. A footnote in page 36 provides the intuition.

2 Patterns of Wage Dispersion and Performance Pay

This section employs the 2003 Workplace and Employee Survey, a matched employer-employee dataset for Canada, to document a set of empirical facts that motivate the theory developed in the paper. The WES was designed and implemented by Statistics Canada. Background information on the survey and further details on the dataset are provided in Appendix B. Here I briefly discuss the salient 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 includes 14,265 workers and 3,540 firms. 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. This adjustment acknowledges that the sample of workers in each firm is drawn from a finite population (total employment in the firm) without replacement.[17] 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.[18]

The individual compensation measure used in this section is the average weekly wage before taxes and deductions and net of overtime payments. Performance pay is computed as weekly-equivalent tips, commissions, piecework payments and bonuses received by the worker. The performance-independent component is, in turn, computed as the weekly wage net of performance pay.

The vector of observable characteristics of worker 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. Tables B-I and B-II in Appendix B report descriptive statistics for the samples of workers and firms, respectively.

[17] Appendix B contains the details.

[18] This is not entirely surprising. Indeed, in a hypothetical, limiting case in which the dataset contained exactly one matched employee per firm, within- and between-firm inequality would mechanically equal 0 and 100 per cent of total wage inequality, respectively.

2.1 Between- and Within-firm Inequality

To gauge the contributions of between- and within-firm inequality across Canadian workers, I compute the ANOVA decomposition in equation (1). In row 1 of Table I, within-firm inequality accounts for 35% of the variance of log weekly wages in the full sample.[19] In the manufacturing sector, this share increases to 42% (row 2).[20] Table I also shows that quantitatively similar results are obtained by decomposing two alternative inequality measures, the Theil index and the mean log deviation (MLD) of wages.[21]

Table I - Within-firm Wage Inequality (%)

		Inequality Measure		
Row	Wage Decomposition	Var of logs	Theil	MLD
1	Unconditional	35	34	31
2	- Manufacturing	42	42	39
3	- Non-manufacturing	34	33	30
4	Within industries	40	38	35
5	Within industries and occupations	43	42	39
6	- Conditional on worker observables	45	44	42

Notes: Each row reports the percentage of within-firm inequality obtained by decomposing three measures of wage inequality -variance of logs, Theil Index and MLD- into its between-firm and within-firm components. In all cases, the corresponding percentage of between-firm inequality is 100 minus the percentage of within-firm inequality. Rows 1 to 3 report decompositions of unconditional wages in the full, manufacturing and non-manufacturing samples, respectively. Using the full sample, rows 4 to 6 report decompositions of residuals from regressions of log weekly wages on: industry fixed effects (row 4); industry and occupation fixed effects (row 5); worker observables, industry and occupation fixed effects (row 6). In rows 4 to 6, the Theil and MLD decompositions are applied to the exponential of the corresponding regression residuals.

The share of within-firm inequality rises when decomposing residual wage dispersion. Table I reports decompositions of residuals obtained from linear regressions of log weekly wages on: industry fixed effects (row 4); industry and occupation fixed effects (row 5); and worker observables, industry and occupation fixed effects (row 6). Sequentially purging the variation in wage premia across industries, occupations and worker observables yields an increasing share of within-firm inequality that ranges from 40% to 45% of the variance of residual log wages. Section B.3 of Appendix B 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.

[19] 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 results in Table I are 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%.

[20] Results for industry-level decompositions (14 industries) are available upon request.

[21] These two measures, introduced by Theil (1967), belong to the generalized entropy class and can thus be decomposed into between and within components (see Shorrocks (1980)). The decomposition formulas for the MLD and the Theil index are given in equations (34) and (35), respectively.

2.2 Heterogeneity in Firm-level Wage Distributions

By definition, the within-firm component in equation (1) measures the degree of inequality within a typical Canadian firm, i.e. the average firm-level log wage variance. There is, however, substantial cross-firm variation around this average. For example, the coefficient of variation of the firm-level variance of log wages is 1.45, while the coefficient of variation of the firm-level mean wage is 0.55.[22] The existence of cross-firm heterogeneity in $Var\left[\widetilde{w}|\theta\right]$ indicates that, in principle, within-firm inequality can change not only through changes in $Var\left[\widetilde{w}|\theta\right]$, but also through worker reallocations across firms, for fixed $Var\left[\widetilde{w}|\theta\right]$. Simply put, if $Var\left[\widetilde{w}|\theta\right]$ were constant across firms, worker reallocations could not affect within-firm inequality.

Moreover, a clear pattern emerges from the heterogeneity in firm-level wage distributions: high-wage firms are typically high-inequality firms. The correlation between $E\left[w|\theta\right]$ and $Var\left[\widetilde{w}|\theta\right]$ across firms is 0.4, statistically significant at the 1% level.[23] Figure 1A illustrates this pattern. The horizontal axis sorts firms by increasing firm-level mean wage into 25 equally populated cells. The vertical axis measures the difference between the median firm-level variance of log wages in the corresponding cell and the median firm-level variance of log wages in cell 1. Figure B-1 in Appendix B shows similar patterns when inequality is measured by the Theil index or the MLD of wages.

The pattern documented in Figure 1A is robust to controlling for differences in the composition of observable skills. To show this, I estimate linear approximations to the mean and variance of the conditional log wage distribution for workers with identical observable skills employed in firm θ:

$$E\left[\widetilde{w}_i|e_i,\theta\right] \approx \phi_1 e_i + \psi_{\theta,1}, \tag{2a}$$

$$Var\left[\widetilde{w}_i|e_i,\theta\right] \approx \phi_2 e_i + \psi_{\theta,2}, \tag{2b}$$

where i indexes workers and $e_i = [\ldots]$.

I estimate $(\phi_1,\ \psi_{\theta,1})$ by regressing $\widetilde{w}_i$ 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 $(\phi_2,\ \psi_{\theta,2})$. The finite population correction is applied to $\psi_{\theta,2}$. The results again

[22] Here I'm interested in documenting cross-firm variation in firm-level wage inequality; hence I report the coefficient of variation of the firm-level variance of log wages. Unlike the variance of wages, the variance of log wages is a scale-independent measure of inequality. The finding is robust across sectors and inequality measures, and holds when the sample is restricted to firms with at least five matched employees. For example, in the manufacturing sector, the coefficients of variation of $Var\left[\widetilde{w}|\theta\right]$ and $E\left[w|\theta\right]$ are 1.80 and 0.45, respectively. In turn, the coefficients of variation of the firm-level MLD (Theil index) of wages in the full and manufacturing samples are 1.37 and 1.40 (1.35 and 1.35), respectively. When the sample is restricted to firms with at least five matched employees, the coefficients of variation of $Var\left[\widetilde{w}|\theta\right]$ and $E\left[w|\theta\right]$ are 1.06 and 0.45, respectively.

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

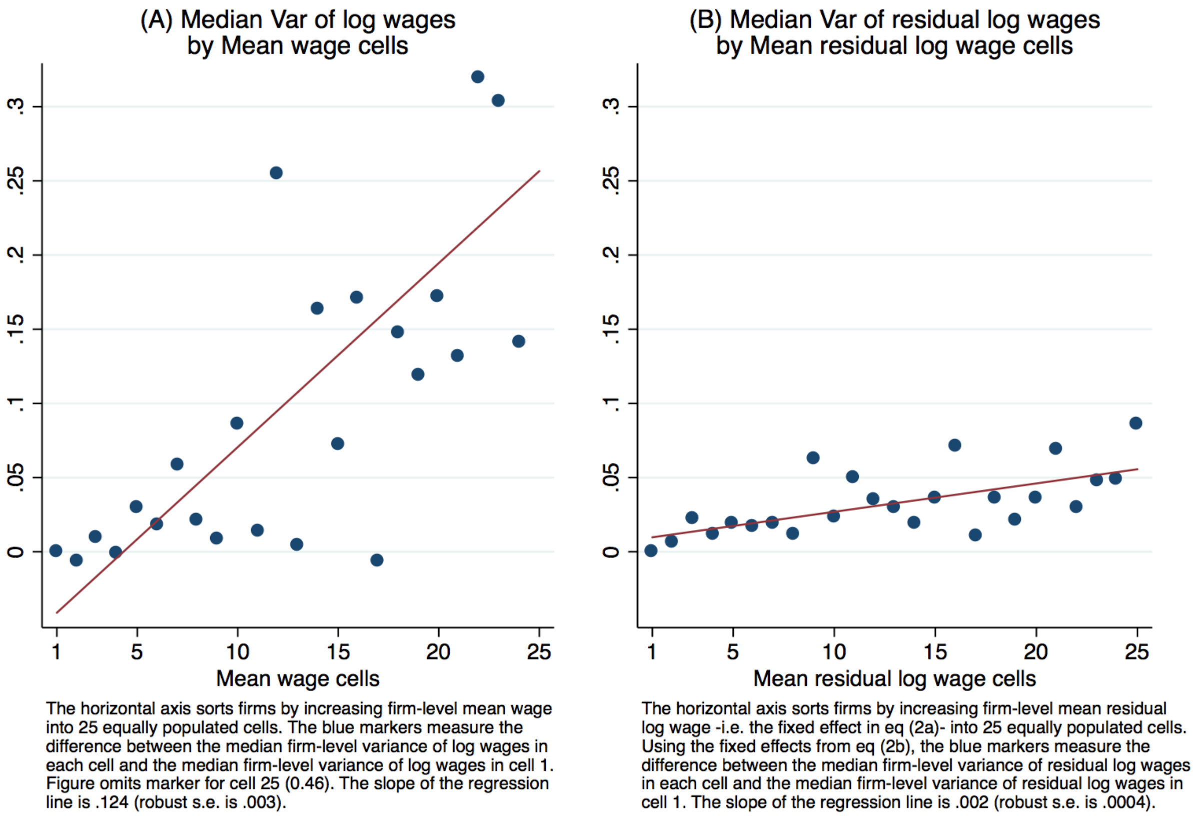


Figure 1: Heterogeneity in Firm-level Wage Distributions

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.[24] Figure 1B illustrates this pattern and Figure B-1 in Appendix B displays similar results when firm-level inequality is measured by the Theil index or the MLD of residual wages –obtained from (2a). A Wald test indicates that the estimated ϕ_2 is statistically different from zero (1% level). This provides evidence of composition effects in the residual variance, indicating that the degree of inequality between observationally equivalent co-workers depends on their observable characteristics.[25]

Below, in Corollary 1, I show that the patterns documented in Figure 1 can be rationalized in the context of a theory that features homogeneous workers and cross-firm heterogeneity in performance pay. The estimated $\psi_{\theta,1}$ and $\psi_{\theta,2}$, however, capture both variation in

[24] In the subsample of firms with at least five matched workers, the correlation is also 0.3 (1% level).

[25] In line with a long literature in labor economics, I find statistically significant wage premia for the individual characteristics e_i. The Wald test for the joint significance of e_i in equation (2a) rejects the null hypothesis that $\phi_1 = 0$ with p-value 0.000. The estimated firm fixed effects are jointly significant at the 1% level in both equations (2a) and (2b).

firm-specific compensation policies for workers with identical characteristics and variation in unobserved workforce composition across firms. The latter interpretation can be ruled out only in the absence of selection on observables, an admittedly strong assumption. Even in this case, this approach is silent on the specific compensation policies that contribute to the heterogeneity in firm-level wage distributions. To improve on some of these limitations, I turn to a different approach.

2.3 Performance Pay

An alternative approach to assess whether compensation policies shape firm-level wage distributions is to directly measure their effects. The WES provides a unique opportunity to gauge the incidence of a specific compensation policy, performance pay, on individual wages. For each worker i in the sample, I compute

$$w_i = fixed_i + pp_i,$$

where pp_i is i's performance pay and $fixed_i$ is the 'fixed' (performance-independent) component of w_i.

Prevalence of Performance Pay. Table II documents 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 performance-pay jobs (PP jobs), i.e. jobs in which workers received strictly positive performance pay. Dissecting this figure by segments of the wage distribution (rows 2-5), sector (rows 6-7), occupation (rows 8-9) or gender (rows 10-11) reveals that PP jobs span a broad cross-section of the workforce. Second, while performance pay accounts for a relatively modest fraction of total compensation, it makes a substantial contribution to overall wage inequality. For example, the mean share of performance pay in PP jobs is 12% of annual compensation. Yet the standard deviation of performance pay is two thirds of the standard deviation of performance-independent compensation across all jobs (row 1). Again, a similar picture emerges if these figures are computed for different segments of the wage distribution, sector, occupation or gender.[26]

[26] Descriptive statistics at the industry level (14 industries) are available upon request.

Table II - Prevalence of Performance Pay

Row	Sample	PP Jobs: Percent	PP Jobs: Mean $\frac{pp_i}{wage_i}$	All jobs: $\frac{std\ dev\ (pp_i)}{std\ dev\ (fixed_i)}$
1	Full sample	46	12	0.67
2	Quartiles 0 to 1 of wage distribution	40	8	0.53
3	Quartiles 1 to 2 of wage distribution	48	7	0.61
4	Quartiles 2 to 3 of wage distribution	48	7	0.73
5	Quartiles 3 to 4 of wage distribution	59	21	0.95
6	Manufacturing	39	7	0.77
7	Non-manufacturing	47	12	0.66
8	Managers	56	17	0.69
9	Non-managers	44	10	0.72
10	Females	47	11	0.77
11	Males	45	12	0.67

Notes: This table describes the prevalence of performance pay in various samples of workers (rows in the table). For each sample, the table reports the fraction of PP jobs, the mean share of performance pay in PP jobs and the ratio of the std. deviations of performance pay (pp_i) and performance-independent compensation ($fixed_i$) for all workers in the sample.

Performance Pay and Firm Size. Next, I analyze the variation of pp_i across and within firms, restricting attention to PP jobs. This approach has merits and limitations that are worth discussing. 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 the analysis likely underestimates the prevalence of performance pay and its contribution to overall inequality since observations for which realized performance pay is zero are discarded.[27] More importantly, 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.[28] Because some determinants of individual performance pay likely remain unmeasured in the data, however, it is not possible to infer the exact degree to which measured dispersion of performance pay

[27] By restricting the ensuing analysis to pp_i, I'm also neglecting any effect that performance pay may have on the dispersion of the fixed component of compensation. However, this margin does not contribute to within-firm inequality in my model; there is no dispersion in $fixed_i$ within firms since workers are homogeneous (Corollary 1).

[28] 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 presented in the next section. 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 $Var(pp_i|s_i, \theta) > 0$.

among co-workers with identical observable characteristics reflects wage dispersion within and across unobservable worker types. I follow up on this point at the end of this section.

In search of systematic patterns of variation in performance pay across firms, I focus on two observable characteristics of firms that are relevant for the theory presented below: firm size and export status. In particular, I estimate the following linear approximations to the probability of observing positive performance pay, the mean and the variance of the conditional distribution of log performance pay for workers with identical observable skills:

$$P\left[pp_{i\theta} > 0 | Size_\theta, Ex_\theta, e_i\right] \approx \delta_A Size_\theta + \zeta_A Ex_\theta + \phi_A e_i, \tag{3a}$$

$$E\left[\ln pp_{i\theta} | Size_\theta, Ex_\theta, e_i\right] \approx \delta_B Size_\theta + \zeta_B Ex_\theta + \phi_B e_i, \tag{3b}$$

$$Var\left[\ln pp_{i\theta} | Size_\theta, Ex_\theta, e_i\right] \approx \delta_C Size_\theta + \zeta_C Ex_\theta + \phi_C e_i, \tag{3c}$$

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 III report estimates of (δ_A, ζ_A, ϕ_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 (δ_B, ζ_B, ϕ_B), obtained by regressing $\ln pp_{i\theta}$ on $Size_\theta$, Ex_θ and e_i in the sample of workers employed in PP jobs. The squared residuals obtained from this regression are subsequently regressed on $Size_\theta$, Ex_θ and e_i to estimate (δ_C, ζ_C, ϕ_C); the results are presented in columns (7) to (9). In all cases, standard errors are clustered at the firm level.

Table III - Performance Pay Across Firms

Outcome	$P\left[pp_{i\theta} > 0 \mid \cdot\right]$			$E\left[\ln pp_{i\theta} \mid \cdot\right]$			$Var\left[\ln pp_{i\theta} \mid \cdot\right]$		
Dep. Var.	$I_{[pp_{i\theta}>0]}$			$\ln pp_{i\theta}$			Sq. residuals from (4) to (6)		
	Basic	Add Exp Status	Add Controls	Basic	Add Exp Status	Add Controls	Basic	Add Exp Status	Add Controls
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Firm size	0.042	0.046	0.064	0.155	0.165	0.151	0.185	0.188	0.167
	$(0.009)^a$	$(0.010)^a$	$(0.010)^a$	$(0.042)^a$	$(0.040)^a$	$(0.031)^a$	$(0.034)^a$	$(0.038)^a$	$(0.034)^a$
Exporter		-0.048	-0.012		-0.122	-0.038		-0.042	-0.033
		(0.037)	(0.037)		(0.147)	(0.131)		(0.122)	(0.119)
Controls	No	No	Yes	No	No	Yes	No	No	Yes
R-sq	0.01	0.02	0.13	0.01	0.02	0.35	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} denotes statistical significance at the 1% level.

For each of the three outcomes considered, Table III 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 and the variance of log performance pay increase with firm total revenue.[29] 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.[30]

Echoing the findings in Table III, Table B-V in Appendix B reports a strong statistical relationship between firm size and wage distributions across firms. In particular, both the mean and the variance of log wage distributions are positively correlated with firm size at the 1% level, with and without controls for worker observables. Interestingly, these correlations are considerably weaker if performance-*independent* compensation is used instead of wages.

Altogether these findings suggest that firm productivity plays a role in explaining why high-wage firms are typically high-inequality firms –a driving force behind the patterns documented in Figure 1. Moreover, performance pay emerges as an empirically relevant channel through which firm heterogeneity shapes inequality patterns. The rest of the paper formally explores this channel in an open economy populated by homogeneous workers, in which each firm offers a single performance-pay contract to all its employees. As acknowledged in the introduction, however, the data does not reject an extension of the model in which each firm offers a combination of contracts to workers of different, yet empirically unobservable, types. Precisely because of the nature of this extension, it is unclear whether augmenting the theory with ex-ante worker heterogeneity can lead to distinct and testable patterns in residual wage dispersion and performance pay across and within firms. An exploration of this topic is left for future work.

3 Model

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. Countries have identical market structure. Firms endogenously choose the level and quality of output, and market(s) to serve in the presence of international trade costs. The main departure from the literature is the existence of moral hazard in the production process. Firms respond by tying compensation to individual performance, generating between- and within-firm wage inequality.

[29]The estimates in Table III should not be given a causal intepretation. In fact, in the model, firm size, mean and variance of log performance pay are functions of firm productivity. Still, under conditions stated below, the theory predicts that these variables should indeed be positively correlated.

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

Labor endowments, trade costs, effort costs, technology and firm productivity distributions are allowed to be asymmetric for much of the analysis, although section 5 and most of section 6 concentrate on symmetric equilibria. I return to asymmetric equilibria at the end of section 6. Throughout, I focus on the description of Home's economy and use an asterisk to denote Foreign's variables.

3.1 Setup

The timing of events in the model combines elements of Holmström and Milgrom (1987) and the static formulation of Melitz (2003). A competitive fringe of risk neutral firms are potential entrants to the differentiated sector. Upon incurring a sunk entry cost of $f_e > 0$ units of the differentiated good, 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 good X. Production requires a fixed cost of $f_d > 0$ units of the domestic differentiated good. In addition, exporting involves a fixed cost of $f_x > 0$ units of the domestic differentiated good and an iceberg variable trade cost, such that $\iota > 1$ units of the firm's output must be produced per unit that arrives in the foreign market. Since all firms with the same productivity behave symmetrically in equilibrium, I index firms and varieties by θ from now onward.

The quantity and quality of firm output depend on the mass and effort, respectively, of workers allocated to a sequential production process with stochastic performance. 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. In the former case, each worker chooses effort at each stage of the production process, as a function of its (publicly observed) personal history of realized performance in previous tasks. At the end of the production process, contracts are executed and consumption takes place. Because workers are homogeneous, in equilibrium, every contract offered by any firm should be individually rational, yielding the same expected utility, denoted $\overline{u}$.[31] The latter is endogenously determined in equilibrium.

3.2 Sequential Production with Stochastic Performance

Firm output is vertically and horizontally differentiated. Physical output of each variety (y) increases in firm productivity (θ) and the mass of workers (h) allocated to the production process; i.e., $y = y(\theta, h)$. In turn, product quality (q) increases in firm productivity and decreases in the average number of mistakes (n) that workers make during the production

[31] Contracts yielding a lower expected utility than the outside option would fail to attract workers. Exceeding $\overline{u}$ would not be profit-maximizing.

process; i.e., $q = q(\theta, n)$. I will differ imposing additional structure on $y(\cdot,\cdot)$ and $q(\cdot,\cdot)$ until section 4.2 (equations (9) and (18), respectively), since this is not necessary to characterize optimal contracts. The latter are determined by the performance of workers in the production process, as described in the remainder of this section.

In every firm, the production process requires each worker to perform a sequence of T symmetric tasks, indexed by $\tau = 1, ..., T$. Each task spans an interval of time of length $\Delta \equiv 1/T$. Worker i chooses a sequence $\left\{\epsilon_{i\tau}^{\Delta}\right\}_{\tau=1}^{T}$ of possibly history-dependent effort levels for each task τ, where $\epsilon_{i\tau}^{\Delta} \in [\epsilon_L, \epsilon_H] \subset \Re_{++}$.[32] This choice generates a stochastic sequence of worker-specific performance outcomes $\{z_{i\tau}\}_{\tau=1}^{T}$, where $z_{i\tau}$ is equal to 1 if worker i successfully completes task τ and equal to -1 in the event of a mistake, for $\tau = 1, ..., T$. For a fixed Δ, the probability of success in any task τ, denoted $\pi_{i\tau}^{\Delta}$, is given by

$$\pi_{i\tau}^{\Delta} \equiv P\left(z_{i\tau} = 1 | z_{i1}, ..., z_{i\tau-1}\right) = \frac{1}{2} + \mu(\epsilon_{i\tau}^{\Delta})\frac{\Delta^{1/2}}{2},$$

where $\mu(\cdot)$ is continuous and, since $\epsilon_{i\tau}^{\Delta} \in [\epsilon_L, \epsilon_H]$, bounded.[33] The expected performance of worker i in task τ is $E\left(z_{i\tau} | z_{i1}, ..., z_{i\tau-1}\right) = \mu(\epsilon_{i\tau}^{\Delta})\Delta^{1/2}$ and thus it is also natural to assume that $\mu(\cdot)$ is increasing. Note that $z_{i\tau}$ is independent of firm productivity and, conditional on effort, independent of $z_{i'\tau'}$ for any two tasks τ and τ' and any two workers i and i' (unless, of course, $\tau = \tau'$ and $i = i'$). The randomness of a task's outcome captures unmodeled determinants of a worker's performance such as unobserved skills and idiosyncratic variation in the quality of inputs used in the production process.

Let $Z_{i\tau}^{\Delta}$ denote the normalized cumulative performance of worker i up to task τ, $Z_{i\tau}^{\Delta} \equiv \Delta^{1/2}\sum_{\tau'=1}^{\tau} z_{i\tau'}$. Equivalently, $-Z_{i\tau}^{\Delta}$ is worker i's normalized number of mistakes in excess of successes up to task τ; i.e., the *net* number of mistakes. To characterize the convergence of the path of cumulative performance as the duration of tasks Δ approaches zero, I embed the discrete process $\left\{Z_{i\tau}^{\Delta}\right\}_{\tau=1}^{T}$ in continuous time by linearly interpolating between the points $(0,0)$, $\left(\Delta, Z_{i1}^{\Delta}\right)$, $\left(2\Delta, Z_{i2}^{\Delta}\right)$,...,$\left(1, Z_{iT}^{\Delta}\right)$. In other words, I construct a function $Z_i^{\Delta}(t)$ satisfying

$$Z_i^{\Delta}(t) = \left(1 - \frac{t}{\Delta} + \left\lfloor\frac{t}{\Delta}\right\rfloor\right) Z_{i\lfloor t/\Delta\rfloor}^{\Delta} + \left(\frac{t}{\Delta} - \left\lfloor\frac{t}{\Delta}\right\rfloor\right) Z_{i\lfloor t/\Delta\rfloor+1}^{\Delta}, \tag{4}$$

for $t \in [0,1]$ and the initial condition $Z_{i0}^{\Delta} = 0$, where $\lfloor x \rfloor$ is the integer part of $x \geq 0$. Note that $Z_i^{\Delta}(t)$ is a random element of the space of continuous functions, $C[0,1]$. Similarly, let $\epsilon_i^{\Delta}(t)$ denote a continuous-time representation of $\left\{\epsilon_{i\tau}^{\Delta}\right\}_{\tau=1}^{T}$. Endowing $C[0,1]$ with the uniform metric, I obtain the following result.

[32] $\mu_{\min}$ is the lowest feasible effort level for any worker. For the purpose of describing technology, it suffices to take the sequence of effort levels as given. Optimal effort choices are analyzed in section 4.1 in the continuous-time limit of this production process.

[33] For any positive Δ, $|\mu(\cdot)| \leq \Delta^{-1/2}$ is necessary and sufficient for $0 \leq \pi_{it}^{\Delta} \leq 1$. When $\Delta \to 0$, the former condition becomes innocuous. However, boundedness of $\mu(\cdot)$ is still necessary to establish Lemma 1.

Lemma 1 *In the sequential production process with task duration $\Delta = 1/T$, consider a sequence of effort $\left\{\epsilon_{i\tau}^{\Delta}\right\}_{\tau=1}^{T}$ and the corresponding process of cumulative performance $\left\{Z_{i\tau}^{\Delta}\right\}_{\tau=1}^{T}$ for worker i. Suppose that $\epsilon_i^{\Delta}(t) \to \epsilon_i(t)$ a.s. as $\Delta \to 0$, for $t \in [0,1]$. If $\mu(\cdot)$ is continuous then, as $\Delta \to 0$, $Z_i^{\Delta}(t)$ converges in distribution to a stochastic process $Z_i(t)$, such that:*

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

for $t \in [0,1]$, where $B_i(t)$ is a Wiener process on $0 \leq t \leq 1$, such that for all i, $B_i(0) = 0$ a.s. and $E\left[B_i(1)^2\right] = 1$.

The crux of this result is showing that deviations of cumulative performance $\left\{Z_{i\tau}^{\Delta}\right\}_{\tau=1}^{T}$ from its expected value follow a martingale process. Convergence to a standard Brownian motion in the space $C[0,1]$ is then an application of martingale limit theory.[34] In particular, the proof of Lemma 1 relies on a result due to Brown (1971). The assumptions of Bernoulli task outcomes and unidimensional effort are not essential.[35]

Lemma 1 states that, when task duration approaches zero, the path of cumulative performance of worker i converges to a Brownian process whose (random) drift is a function of the worker's effort choices. The remainder of the paper restricts the attention to this limiting case to avoid the well-known intractability of optimal contracts in static, higher-dimensional settings (multiple outcomes and effort levels) of the moral hazard problem.[36] The value of Lemma 1 is to provide an economically relevant interpretation of this continuous-time environment as the limit of the sequential production process introduced above.

To map outcomes of the continuous-time production process into product quality, recall that $-Z_{iT}^{\Delta}$ is the net number of mistakes of worker i in the production process with task length Δ. Therefore, I define the net average number of mistakes as follows:

$$n \equiv h^{-1} \int_0^h -Z_i(1) di. \tag{5}$$

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.[37] Moreover, in section 4.1, I show that firms optimally implement non-stochastic effort sequences. These observations imply that n is almost surely a constant and thus equilibrium firm-level variables such as quality, output price, revenue and profits

[34] Hall and Heyde (1980) provide a comprehensive treatment of this literature.

[35] Hellwig and Schmidt (2002) generalize these assumptions in the context of a principal-agent model.

[36] See, for example, Holmström and Milgrom (1987) and chapter 4 in Bolton and Dewatripont (2005).

[37] I rely throughout on applications of the exact law of large numbers for a continuum of random variables. A precise statement can be found in Sun (2006). A less technical exposition of the subtle measurability issues that arise when extending the law of large numbers to a continuum of independent random variables is provided by Khan et al. (2015).

are deterministic with probability one.

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:

$$U(X_i, \boldsymbol{\epsilon}_i) = \frac{X_i}{\exp\left(\int_0^1 k(\epsilon_i(t))\, dt\right)}, \tag{6}$$

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

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

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.[38]

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 (fixed costs). These activities are assumed to use the output of each variety in the same way as is demanded by final consumers. Denoting total expenditure on the differentiated good by E, the aggregate demand for variety j, 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 equals the revenue of producer j in Home, denoted $r(j)$. Therefore,

$$r(j) \equiv p(j)x(j) = Aq(j)^{\rho}x(j)^{\rho}, \tag{7}$$

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., $U(X_i, \boldsymbol{\epsilon}_i) = U(w_i, \boldsymbol{\epsilon}_i)$.

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

4 The Firm's Problem

This section studies the problem of firm θ located in Home, in two steps. The first step takes firm employment and output quality as given, while seeking to characterize the design of optimal contracts to attain the targeted quality at minimum cost. The second step sets up the profit maximization problem, in which the firm determines employment, quality and whether to export given demand in the domestic and foreign markets.

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., $Z_i^1 \equiv \{Z_i(t); t \in [0,1]\}$.[39] 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 fewer than n_0 net mistakes per worker, (ii) the stochastic processes for individual performance, (iii) incentive compatibility constraints and (iv) participation constraints:

$$\min_{\{w_i, \boldsymbol{\epsilon}_i; i \in [0,h]\}} \int_0^h E\left[w_i(Z_i^1)\right] di \tag{8}$$

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

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

[39] Although potentially relevant to study within-firm wage variation, this paper does not deal with any form of group-based compensation schemes. The emphasis on individual incentives can be motivated empirically. In the 2003 WES, 33% of firms in my sample report having 'individual incentive systems' that reward individuals on the basis of output or performance. Only 10% of the firms in the sample, however, have 'group incentive systems' that reward individuals on the basis of group output or performance. Lazear and Shaw (2007) report that the share of large US firms in which more than 20 percent of their workforce is subject to some form of individual incentives, like a performance bonus, has grown from 38 percent in 1987 to 67 percent in 1999. The comparable share of firms using any form of 'gain-sharing' or group-based incentives was 7 percent in 1987 and 24 percent in 1999.

condition for the existence of a solution in (8). In turn, the supremum of N makes the moral hazard problem interesting, by requiring the firm to implement effort levels greater than ϵ_L.[40] 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 (8), 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

$$\begin{aligned}
\beta &= k'(\epsilon^*)/\mu'(\epsilon^*), \\
\alpha &= \ln \overline{u} + k(\epsilon^*) - \frac{k'(\epsilon^*)}{\mu'(\epsilon^*)}\mu(\epsilon^*) - \frac{1}{2}\left[\frac{k'(\epsilon^*)}{\mu'(\epsilon^*)}\right]^2.
\end{aligned}$$

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

The solution to problem (8) has several important features. First, 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. The model thus inherits the simple structure of contracts in 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 relative to Holmström and Milgrom (1987) is the specification of the objective functions of firms and workers. In particular, the firm's cost minimization problem (8) arises naturally in the context of the broader profit maximization problem studied in the next section. Moreover, the utility function (6) plays a key role in ensuring that wages are positive for all realizations of individual performance $Z_i(1)$.[41] Because wages fuel the demand side of the model, this is an appealing property when embedding the moral hazard problem in general equilibrium.

[40] 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.

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

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.[42] 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 sections 2.2 and 2.2. 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:

(i) the variance of log wages.

(ii) all inequality measures that respect second-order stochastic dominance and scale independence.

In line with the findings of section 2.2, 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 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 scale-independent inequality measures that includes the Theil index and the MLD of wages.

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

The next section endogeneizes the firm's choice of effort, establishes the conditions under which effort increases in firm productivity and shows that, conditional on productivity, optimal effort is independent of export status. With this additional structure, two implications immediately follow from Corollary (1). Both average wage and firm-level wage inequality: (i) increase in firm productivity and (ii) are independent of the firm's export status. These results are consistent with the findings of section 2.2 and will play a key role in the analysis of the impact of international trade on wage inequality.

4.2 Profit Maximization

If price and quality discrimination across markets is feasible, then by (7) revenues from domestic and foreign sales are given by $r_d = Aq_d^\rho y_d^\rho$ and $r_x = A^* q_x^\rho [y_x/\iota]^\rho$, respectively. r_m, q_m and y_m denote revenue, quality and output in market $m = \{d, x\}$, respectively. I assume that the firm's output for m is linear in the mass of workers allocated to that production line, h_m,

$$y_m = \theta^s h_m, \quad s \geq 0. \tag{9}$$

The ensuing analysis thus nests the case of identical labor productivity across firms ($s = 0$). Recall that fixed costs are measured in units of the domestic differentiated good, whose price is normalized to one. From these observations, 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

$$\Pi(\theta) \equiv \max_{\substack{\epsilon_m \in [\epsilon_L, \epsilon_H], \\ q_m \in [q_L, q_H], \\ y_m \geq 0,\ I_x \in \{0,1\}}} Aq_d^\rho y_d^\rho - \frac{\omega(\epsilon_d)}{\theta^s} y_d - f_d + I_x \left[A^* q_x^\rho [y_x/\iota]^\rho - \frac{\omega(\epsilon_x)}{\theta^s} y_x - f_x \right], \tag{10}$$

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.[43] 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 (10) 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

[43] 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.

$c^{\theta}(q)$ is the (factory) unit-cost function in firm θ when output quality is q. Assume $I_x = 1$ and express profits in market $m = \{d, x\}$ 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 $\eta_x \equiv A^* \iota^{-\rho}$.

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,

$$\eta_m q_m^{\rho} [y_m(q_m)]^{\rho-1} = c^{\theta}(q_m), \text{ for } m = \{d, x\}. \tag{11}$$

Solving for $y_m(q_m)$ from (11), substituting it in Π_m and rearranging yields

$$\Pi_m(q_m) = \left[(\rho)^{\rho/(1-\rho)} - (\rho)^{1/(1-\rho)}\right] (\eta_m)^{1/(1-\rho)} \left(\left[\frac{q_m}{c^{\theta}(q_m)}\right]^{\rho/(1-\rho)} \right) - f_m. \tag{12}$$

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

$$c_q^{\theta}(q_m(\theta)) = \frac{c^{\theta}(q_m(\theta))}{q_m(\theta)}, \text{ for } m = \{d, x\}. \tag{13}$$

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 ι. Therefore, trade liberalization does not induce quality upgrading or downgrading at the firm level, in either market.[44] 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. The latter is driven by the assumption of identical consumer preferences across countries.[45]

In this case, the firm's unit costs 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 (7), this requires $[y_x(\theta)/y_d(\theta)]^{1-\rho} = \iota^{-\rho}(A^*/A)$, which

[44] An alternative way to derive this result is the following. The firm can increase profits in a given market by either expanding output or quality. Optimality requires that choices of output and quality in each market satisfy the equality of relative marginal revenue and relative marginal cost. From (10), the variable trade cost decreases the marginal revenue of output and quality proportionally in the foreign market, thus ι does not distort the optimal quality per unit of output across markets.

[45] 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 (see Verhoogen (2008)). For example, letting $X_i^* = \left[\int_{j \in J^*} (q^*(j)^{\chi} x_i^*(j))^{\rho} dj\right]^{1/\rho}$, 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.

implies that the firm's total revenue is

$$r(\theta) \equiv r_d(\theta) + I_x(\theta) r_x(\theta) = A q(\theta)^{\rho} y(\theta)^{\rho} \Upsilon(\theta)^{1-\rho}, \tag{14}$$

where $y(\theta)$ and $q(\theta)$ are total output and product quality in firm θ, respectively. Following Helpman et al. (2010), $\Upsilon(\theta) \equiv 1 + I_x(\theta) \left[\iota^{-\rho} \left(A^*/A\right)\right]^{1/(1-\rho)}$ is a measure of market access for firm θ.

Expression (11) implies that variable costs are equal to a fraction ρ of revenue in each market. With identical quality across markets, the firm's total profits is a fraction $1-\rho$ of total revenue net of fixed costs,

$$\Pi(\theta) = (1-\rho) r(\theta) - f_d - I_x(\theta) f_x. \tag{15}$$

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 $\theta < \theta_x$.[46] 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 $\Upsilon_x \equiv 1 + \iota^{-\frac{\rho}{1-\rho}} \left(A^*/A\right)^{\frac{1}{1-\rho}} > 1$.

Closed-form Solutions. As shown in expression (12), product quality $q(\theta)$ is fully determined by the unit-cost function $c^{\theta}(\cdot)$. The following three functional form assumptions in turn determine $c^{\theta}(\cdot)$, guaranteeing existence, uniqueness and optimality of $q(\theta)$, and yielding closed-form solutions to the profit maximization problem:

$$\begin{aligned} k(\epsilon) &= k\epsilon, \quad k > 0, && (16) \\ \mu(\epsilon) &= -1/\epsilon, && (17) \\ \log q &= (\gamma \log \theta)^{z} (1/n)^{(1-z)}, \quad \gamma > 0, \; z \in (0,1). && (18) \end{aligned}$$

Note that (16) and (17) satisfy Assumption (1). By (17), (5) and Lemma 1, $n = -\mu(\epsilon) > 0$ almost surely.[47] This guarantees that quality is properly defined in (18). Under specification (18), quality is log-submodular in θ and n and thus the return to reducing the number of mistakes increases in firm productivity. This key property implies that high productivity firms have a comparative advantage in producing high quality output, which is consistent

[46] Note that $r(\theta)$ increases in θ if and only if $r_m(\theta)$ increases in θ, for any $m = \{d, x\}$.

[47] Under (17), $\mu(\cdot) \subset [\mu(\epsilon_L), \mu(\epsilon_H)]$, satisfying the requirements of Lemma 1.

with the empirical evidence in Kugler and Verhoogen (2012) and Manova and Zhang (2012).[48]

Minimizing the average cost of quality under (16)-(18) yields a closed-form solution for a unique optimal quality $q(\theta)$ (see Appendix A). 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,

$$q(\theta) = \theta^{\kappa_q}, \quad \kappa_q \equiv \gamma \left[(1-z)/k\right]^{(1-z)/z}, \tag{19}$$

$$\epsilon(\theta) = \kappa_\epsilon \log \theta, \quad \kappa_\epsilon \equiv \gamma \left[(1-z)/k\right]^{1/z}. \tag{20}$$

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 $\kappa_\epsilon [\log \theta_L, \log \theta_H] \subset [\epsilon_L, \epsilon_H]$. To avoid a taxonomic exercise, I will henceforth focus on this parameter configuration.[49]

From the first-order condition for output (11) and the expression for firm revenue (14), 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 (9). Therefore:

$$r(\theta) = \kappa_r \Upsilon(\theta) \left(A\overline{u}^{-\rho}\right)^{1/(1-\rho)} \theta^{\Gamma}, \quad \kappa_r \equiv \rho^{\rho/(1-\rho)}, \tag{21}$$

$$y(\theta) = \kappa_y \Upsilon(\theta) \left(A\overline{u}^{-1}\right)^{1/(1-\rho)} \theta^{\Gamma+s-k\kappa_\epsilon}, \quad \kappa_y \equiv \rho^{1/(1-\rho)}, \tag{22}$$

$$h(\theta) = \kappa_y \Upsilon(\theta) \left(A\overline{u}^{-1}\right)^{1/(1-\rho)} \theta^{\Gamma-k\kappa_\epsilon}, \tag{23}$$

where $\Gamma \equiv \rho(\kappa_q - k\kappa_\epsilon + s)/(1-\rho)$. 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 there is no motif for quality upgrading (or downgrading) associated to exporting in this model (see footnote in page 25).

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

[48] Formally, in this context, $q(\theta, N)$ is log-submodular if, for any $0 < \theta_0 < \theta_1$ and $0 < N_0 < N_1$, then $\log q(\theta_0, N_1) + \log q(\theta_1, N_0) > \log q(\theta_0, N_0) + \log q(\theta_1, N_1)$. For an in-depth analysis of the relationship between supermodularity, submodularity and comparative advantage, see Costinot (2009).

[49] 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 the probability of observing positive performance pay, reported in Table III. Moreover, the share of PP jobs will typically decrease with trade liberalization, as long as a lower variable trade cost leads to a higher θ_d.

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 (21) and profits (15), these two conditions require

$$\kappa_r(1-\rho)\,(\overline{u})^{-\rho/(1-\rho)}\,\theta_d^{\Gamma}E = f_d \tag{24}$$

and

$$\kappa_r(1-\rho)\,(\overline{u}\iota)^{-\rho/(1-\rho)}\,\theta_x^{\Gamma}E = f_x, \tag{25}$$

respectively.[50] Dividing (25) by (24) yields

$$\iota^{-\frac{\rho}{1-\rho}}\left(\frac{\theta_x}{\theta_d}\right)^{\Gamma} = \frac{f_x}{f_d}. \tag{26}$$

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 (21) and productivity cutoffs (24) and (25), the free entry condition is

$$f_d J\left(\theta_d\right) + f_x J\left(\theta_x\right) = f_e, \tag{27}$$

where $J\left(\theta_m\right) \equiv \int_{\theta_m}^{\theta_H}\left[\left(\theta/\theta_m\right)^{\Gamma} - 1\right]dG_\theta(\theta)$, $m = \{d, x\}$, 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 (26) and (27) 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.[51] As shown in Appendix A, 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 2). 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 $(f_x/f_d)\iota^{\frac{\rho}{1-\rho}} \geq 1$. 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

[50] 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 (25) also uses the fact that $\Upsilon_x - 1 = \iota^{-\frac{\rho}{1-\rho}}$ in any symmetric equilibrium.

[51] 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.

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 (15) and using the free entry condition.[52] The market clearing condition for the goods market thus becomes

$$(1-\rho)E = M_e \left[f_e + f_d \left[1 - G_\theta(\theta_d)\right] + f_x \left[1 - G_\theta(\theta_x)\right]\right]. \tag{28}$$

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 (23) to substitute for firm employment yields

$$L = \kappa_y (\overline{u})^{-1/(1-\rho)} E M_e \int_{\theta_d}^{\theta_H} \Upsilon(\theta)\theta^{\Gamma - k\kappa_\epsilon} dG_\theta(\theta). \tag{29}$$

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 (24), (25), (27), (28) and (29).

Proposition 2 *There exists a unique symmetric equilibrium.*

6 Trade Liberalization, Selection and Reallocations

This section studies the impact of trade liberalization on reallocations of workers and wage shares across firms. Because optimal performance-pay contracts differ across firms, labor reallocations have profound implications for the equilibrium distribution of wages in the economy, which are studied in the next section. While the analysis centers on symmetric equilibria, I close this section with a partial characterization of these effects in asymmetric equilibria. When needed, I use a subindex $j \in \{0, 1\}$ to indicate equilibria before ($j = 0$) and after ($j = 1$) trade liberalization.

I consider a decline in the (bilateral) variable trade cost, $\iota_0 > \iota_1$, for $\iota_1 \in [\underline{\iota}, \overline{\iota})$, holding the remaining parameters of the model constant. The lower bound for ι_1 is $\underline{\iota} \equiv \max\left\{1, (f_d/f_x)^{(1-\rho)/\rho}\right\}$, ensuring that the post-liberalization equilibrium features selection of the most productive firms into exporting. If $f_d > f_x$ then $\iota_1 = \underline{\iota}$ corresponds to an equilibrium in which every firm exports following trade liberalization. The supremum for ι_1, denoted $\overline{\iota}$, is the variable trade cost such that $\theta_{x,1} = \theta_H$, implicitly defined by equations (26) and (27). Note that $\iota_j \geq \overline{\iota}$ if and only if equilibrium j is autarkic. The restriction $\iota_1 < \overline{\iota}$ thus ensures that trade liberalization effectively opens the economy.

[52] In particular, integrating (15) and using the fact that, by free entry, average firm profits are equal to $f_e/[1 - G(\theta_d)]$, yields $(1-\rho)\overline{r} = f_e/[1 - G(\theta_d)] + f_d + f_x [1 - G(\theta_x)]/[1 - G(\theta_d)]$.

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 (23),

$$G_h(\theta) = \frac{\int_{\theta_d}^{\theta} \Upsilon(\theta')\left(\theta'\right)^{\Gamma - k\kappa_\epsilon} dG_\theta(\theta')}{\int_{\theta_d}^{\theta_H} \Upsilon(\theta')\left(\theta'\right)^{\Gamma - k\kappa_\epsilon} dG_\theta(\theta')}, \text{ for } \theta \in [\theta_d, \theta_H]. \tag{30}$$

Importantly, $G_h(\theta)$ depends on just two endogenous variables; namely, the productivity cutoffs.[53] 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 ι_j, respectively. From (26), $\iota_0 > \iota_1$ if and only if $\theta_{x,0}/\theta_{d,0} > \theta_{x,1}/\theta_{d,1}$. Moreover, the free entry condition (27) 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 $\iota_0 \geq \bar{\iota}$, then $G_{h,1}$ first-order stochastically dominates $G_{h,0}$.

(b) If $\iota_0 < \bar{\iota}$, consider $\widehat{\theta} \in [\theta_{x,0}, \theta_H)$:

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 $[\theta_{d,1}, \theta_H)$.

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.[54] Indeed, Lemma (2) implies that $G_{h,1}$ second-order stochastically dominates $G_{h,0}$ whenever trade liberalization reduces the mass of active firms.[55] For a sharp characterization of the effect of trade liberalization on within-firm inequality, however, I will rely on a stronger form

[53] To see this, note that $\Upsilon_x = 1 + \iota^{-\frac{\rho}{1-\rho}}$ in any symmetric equilibrium, thus $\Upsilon(\theta)$ depends only on the productivity cutoffs. This property, however, does not rely on symmetry. In particular, Appendix A 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.

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

[55] To see this, recall that first-order stochastic dominance implies second-order stochastic dominance. More-

of labor reallocation, 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 $-(\Gamma+1)$ (see section A.8 in Appendix A). Assumption (2) is sufficient to establish that $G_{h,1}(\widehat{\theta}) \leq G_{h,0}(\widehat{\theta})$, for any $\widehat{\theta} \in [\theta_{x,0}, \theta_H)$. In light of Lemma (2), I obtain the following result.

Proposition 3 *Let* $G_{h,0}$ *and* $G_{h,1}$ *denote employment distributions corresponding to symmetric equilibria before and a bilateral after trade liberalization, respectively. If* $\iota_0 < \bar{\iota}$, *impose Assumption (2). Then* $G_{h,1}$ *first-order stochastically dominates* $G_{h,0}$*. That is, for all* $\theta \in [\theta_L, \theta_H]$,

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

To characterize the impact of trade liberalization on the Theil index of wages, I need to analyze 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,

$$G_w(\theta) = \frac{\int_{\theta_d}^{\theta}\omega(\theta')dG_h(\theta')}{\int_{\theta_d}^{\theta_H}\omega(\theta')dG_h(\theta')}, \text{ for } \theta \in [\theta_d, \theta_H]\,. \tag{31}$$

Proposition 4 *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* $\iota_0 < \bar{\iota}$, *impose Assumption (2). Then* $G_{w,1}$ *first-order stochastically dominates* $G_{w,0}$*. That is, for all* $\theta \in [\theta_L, \theta_H]$,

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

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

over, 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.

costs, effort costs, technology and firm productivity distributions.[56] 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 (3) and (4) 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, Appendix A 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 5 *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 $G_{w,0}$.*

7 Trade Liberalization and Inequality

There are two sources of heterogeneity in individual wages, a firm-specific component θ and a worker-specific component $B_i = B_i(1)$. The distribution of wages in the economy (and thus measures of wage inequality) will therefore depend on the underlying distributions of employment across firms productivity and idiosyncratic performance.

More specifically, combining the firm's optimal choice of effort (20) with parts (a) and (b) of Proposition (1) and functional forms (16) and (17), yields the wage of worker i employed

[56] 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. Appendix A provides further details.

in firm θ,

$$w_i = w(\theta, B_i) = \overline{u} \exp\left[k(\theta) + \beta(\theta)(B_i - \beta(\theta)/2)\right]. \tag{32}$$

For any w_0, let $\Phi(B(\theta, w_0))$ denote the fraction of employees in firm θ with wages lower than or equal to w_0, where Φ is the standard normal c.d.f. and $B(\theta, w_0)$ satisfies $w_0 = w(\theta, B(\theta, w_0))$. Then the wage distribution, denoted $F_w(w)$, is given by

$$F_w(w) = \int_{\theta_d}^{\theta_H} \Phi(B(\theta, w)) dG_h(\theta). \tag{33}$$

The distribution of wages is therefore a mixture of the distributions of $h(\theta)$ and B.

I start by analyzing a specific inequality measure constructed from (33), the variance of log wages.[57] The latter has been frequently applied in recent empirical studies of wage inequality.[58] Unlike other popular measures of inequality such as the Gini coefficient and the 90-10 wage gap, the variance is additively decomposable into between- 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, however, 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 established in this section and constitute the main results of the paper. 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 Coçar et al. (forthcoming), 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.

Proposition 6 *Consider any symmetric equilibrium with variable trade cost ι_0. If $\iota_0 < \overline{\iota}$, impose Assumption (2). Then a bilateral trade liberalization $\iota_1 < \iota_0$, for $\iota_1 \in [\underline{\iota}, \overline{\iota})$, 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 (5). In the following result, G_θ should be understood as the productivity distribution in Home.

Proposition 7 *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.*

[57] 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 (32).

[58] For example, among recent empirical studies, Lemieux (2006), Helpman et al. (2014) and Card et al. (2013) use variance decompositions of log wages to analyze changes in inequality in the US, Brazil and Germany, 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.

7.1 The Variance of Log Wages

In the model, different firms select different performance-pay contracts to reward their employees. This implies that within-firm wage distributions differ across firms, and thus inequality measures will crucially depend on the equilibrium allocation of workers across firms. 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 $Var(\widetilde{w}_i|\theta)$, respectively, where $\widetilde{w}_i \equiv \log w_i$. Given $G_h(\theta)$, these two moments can be integrated across firms to obtain the standard decomposition of the total variance of log wages into between- and within-firm components. This yields,

$$Var_j = Var_j^{between} + Var_j^{within},$$

where

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

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

The between-firm variance is the variance of average log wages across firms, while the within-firm variance is the weighted average of firm-level variances. I will refer to the within-firm variance interchangeably as the *residual* variance of log wages. This label prevents confusion with the firm-level variances $Var(\widetilde{w}_i|\theta)$ and also highlights a link between the analysis in this section and empirical studies of trade and inequality. For example, in Helpman et al. (2014), the between-firm component is the estimated variance of the firm-fixed effects in a regression of individual wages that also controls for observable worker characteristics. The within-firm component is the variance of the regression residuals.

As in 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., within-firm wage dispersion. Moreover, firm-level wage variances vary across firms. High productivity firms offer higher-powered incentives that magnify idiosyncratic differences in performance and thus translate into higher wage inequality among co-workers. In particular, part (b)(i) of Corollary (1) and the expression for optimal effort (20) 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 ι. 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 residual 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 follows from the independence of the firm-level wage distributions and the variable trade cost. The second line requires integration by parts. Part (b)(i) of Corollary (1) and the expression for optimal effort (20) imply that the firm-level variance increases in θ, thus $dVar(\widetilde{w}_i|\theta)/d\theta > 0$. Moreover, Proposition (3) 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.[59]

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

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

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

As in the analysis of the residual variance, the second line applies integration by parts. However, the change in $Var_j^{between}$ cannot, in general, be signed. Note that the firm-level mean log wage is not necessarily increasing in productivity.[60] 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.

7.2 Lorenz-consistent Inequality Measures

Similar results follow from the analysis of alternative 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)).[61] The latter, however, incorporates some principles that are generally regarded as fundamental to the theory of inequality measurement.[62] For this reason, I close this section by analyzing the impact of trade liberalization on two Lorenz-consistent measures, the Theil index (T) and the MLD of wages.[63]

The definition and decomposition of the MLD and Theil measures in equilibrium $j \in \{0,1\}$ are given by

$$\begin{aligned} MLD_j &\equiv E_j\left[\log\left(\frac{w_j^*}{w_i}\right)\right], \\ &= \int_{\theta_L}^{\theta_H} \log\left(\frac{w_j^*}{\omega(\theta)}\right) dG_{h,j}(\theta) + \int_{\theta_L}^{\theta_H} MLD\left(w_i|\theta\right) dG_{h,j}(\theta), \end{aligned} \tag{34}$$

and

$$\begin{aligned} T_j &\equiv E_j\left[\log\left(\frac{w_i}{w_j^*}\right)\frac{w_i}{w_j^*}\right], \\ &= \int_{\theta_L}^{\theta_H} \log\left(\frac{\omega(\theta)}{w_j^*}\right)\frac{\omega(\theta)}{w_j^*} dG_{h,j}(\theta) + \int_{\theta_L}^{\theta_H} T\left(w_i|\theta\right) dG_{w,j}(\theta), \end{aligned} \tag{35}$$

[60] 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.

[61] 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.

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

[63] 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.

respectively, where $w_j^* \equiv \int_{\theta_L}^{\theta_H} E(w_i|\theta) dG_{h,j}(\theta)$ is the mean wage in equilibrium j.

Expressions (34) and (35) state that the MLD and Theil indices can be written as the sum of a component measuring inequality of mean wages across firms (between-firm inequality) and a component measuring average firm-level inequality (within-firm inequality). Importantly, within-firm inequality has a similar structure in both measures, which also resembles the structure of the residual variance of log wages. In particular, for measure $I = \{MLD, T\}$,

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

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

The impact of trade liberalization on I^{within} is then evaluated as in the case of Var^{within}. In particular,

$$\begin{aligned}
\Delta I^{within} &= \int_{\theta_L}^{\theta_H} I(w_i|\theta) \left[dG_{\ell(I),1}(\theta) - dG_{\ell(I),0}(\theta) \right], \\
&= \int_{\theta_L}^{\theta_H} \frac{dI(w_i|\theta)}{d\theta} \left[G_{\ell(I),0}(\theta) - G_{\ell(I),1}(\theta) \right] d\theta, \\
&> 0.
\end{aligned}$$

This result relies on part (b)(ii) of Corollary (1), which ensures $dI(w_i|\theta)/d\theta > 0$. Moreover, Propositions (3) and (4) 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).

8 Concluding Remarks

Evidence collected from matched employer-employee data in several countries consistently shows that wage dispersion between and within firms are major components of wage inequality. This paper links trade liberalization to wage inequality through both channels. To the best of my knowledge, this is the first attempt in the literature to develop a general equilibrium framework to study the determinants of within-firm wage dispersion and its links to international trade liberalization. Moreover, in light of the magnitude and growth of residual wage dispersion, the focus is on modeling within-firm wage inequality between identical workers. Trade liberalization triggers reallocations of workers towards firms that intensively rely on contracting strategies that generate higher wage dispersion among co-workers. The paper identifies conditions under which this mechanism delivers a monotonic effect of trade liberalization on within-firm inequality.

Motivated by empirical evidence documenting the prevalence of performance pay in Canada and the U.S., I have emphasized heterogeneity in optimal performance-pay contracts as the key source of within-firm inequality in the model. In doing so, I have abstracted from cross-firm differences in the composition of worker skills that, if unobservable to the econometrician, would constitute an additional source of residual within-firm inequality in the data. I do so primarily for analytical convenience and because (I conjecture that) introducing ex-ante skill heterogenety in the analysis will most likely operate as a complementary source of within-firm inequality. In such an extension of the model, a firm would design a contract for each type of worker. Wage dispersion within a firm would then be composed of wage variation between and within worker types.

A common feature in several studies in the literature, yet absent in this framework, are exporter wage premia. In the model, conditional on productivity, exporting does not induce firms to pay higher wages. As mentioned, however, one way in which this feature can be incorporated into the model is by assuming that foreign buyers have a relatively higher preference for quality than domestic consumers, as in Verhoogen (2008). Alternatively, introducing increasing marginal costs of output may lead exporters and non-exporters to upgrade and downgrade quality, respectively, as a result of trade liberalization. These extensions would also generate higher wage dispersion within exporting firms, conditional on productivity, which is consistent with the empirical evidence reported in Frías et al. (2012). The analysis shows, however, that exporter wage premia are not necessary for international trade to have an impact on within-firm wage inequality. Introducing exporter wage premia is likely reinforce the main results of the paper.

There are a number of additional topics worth exploring in future research. First, the impact of trade liberalization on ex-post welfare. On one hand, lower trade costs lead to lower consumption prices and higher expected wages. Labor reallocations towards high productivity firms, however, can potentially hurt unlucky workers who, despite high effort levels, end up receiving low wages due to poor ex-post performance. Second, vertical differentiation is not, per se, an indispensable part of the mechanism linking trade liberalization to wage inequality advanced in this paper. The hypothesis that effort enhances product quality provides a microfoundation for the complementarity between firm productivity and effort in the revenue function of firms. The latter is the key driver of cross-firm differences in both performance pay and wage distributions. Although the specific microfoundation adopted in this paper generate predictions that are consistent with empirical evidence, developing alternative settings that lead to similar contracting patterns across firms might constitute an important extension of the present framework.

References

Akerman, A., Helpman, E., Itskhoki, O., Muendler, M.-A., Redding, S., 2013. Sources of wage inequality. The American Economic Review: Papers and Proceedings 103 (3), 214–219.

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

Amiti, M., Khandelwal, A. K., 2013. Import competition and quality upgrading. Review of Economics and Statistics 95 (2), 476–490.

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

Barth, E., Bryson, A., Davis, J., Freeman, R., 2014. It's where you work: Increases in earnings dispersion across establishments and individuals in the u.s. NBER Working Paper 20447.

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

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

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

Bombardini, M., Orefice, G., Tito, M., 2015. Does exporting improve matching? evidence from french employer-employee data. NBER Working Paper 21225.

Brown, B. M., 1971. Martingale central limit theorems. The Annals of Mathematical Statistics 42 (1), 59–66.

Burstein, A., Vogel, J., 2012. International trade, technology, and the skill premium. Columbia University, Mimeo.

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

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

Card, D., Heining, J., Kline, P., 2013. Workplace heterogeneity and the rise of west german wage inequality. The Quarterly Journal of Economics 128 (3), 967–1015.

Coçar, A., Guner, N., Tybout, J., forthcoming. Firm dynamics, job turnover, and wage distributions in an open economy. American Economic Review.

Costinot, A., 2009. An elementary theory of comparative advantage. Econometrica 77 (4), 1165–1192.

Cuñat, V., Guadalupe, M., 2005. How does product market competition shape incentive contracts? Journal of the European Economic Association, 1058–1082.

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

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

Davis, D., Harrigan, J., 2011. Good jobs, bad jobs, and trade liberalization. Journal of International Economics 84 (1), 26–36.

Davis, S. J., Haltiwanger, J., 1995. Employer size and the wage structure in u.s. manufacturing. Tech. rep.

Demidova, S., Rodriguez-Clare, A., 2013. The simple analytics of the melitz model in a small open economy. Journal of International Economics 90 (2), 266–272.

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

Emami Namini, J., Facchini, G., Lopez, R., forthcoming. A model of firm heterogeneity in factor intensities and international trade. Canadian Journal of Economics.

Fan, H., Li, Y. A., Yeaple, S., 2015. Trade liberalization, quality, and export prices. Review of Economics and Statistics 97 (5), 1033–1051.

Felbermayr, G., Impullitti, G., Prat, J., 2015. Firm dynamics and residual inequality in open economies. Mimeo.

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

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

Frydman, C., Saks, R. E., 2010. Executive compensation: A new view from a long-term perspective, 1936–2005. Review of Financial Studies 23 (5), 2099–2138.

Goldberg, P., Pavcnik, N., 2007. Distributional effects of globalization in developing countries. Journal of Economic Literature 45 (1), 39–82.

Hall, P., Heyde, C. C., 1980. Martingale limit theory and its application. Academic press, New York.

Harrigan, J., Reshef, A., forthcoming. Skill biased heterogeneous firms, trade liberalization, and the skill premium. Canadian Journal of Economics.

Hellwig, M. F., Schmidt, K. M., 2002. Discrete–time approximations of the holmström–milgrom brownian–motion model of intertemporal incentive provision. Econometrica 70 (6), 2225–2264.

Helpman, E., Itskhoki, O., Muendler, M., Redding, S., 2014. Trade and inequality: From theory to estimation. Mimeo.

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

Holmström, B., Milgrom, P., 1987. Aggregation and linearity in the provision of intertemporal incentives. Econometrica 55 (2), 303–328.

Karatzas, I., Shreve, S. E., 1988. Brownian motion and stochastic calculus. Springer.

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

Khan, M. A., Rath, K. P., Sun, Y., Yu, H., 2015. Strategic uncertainty and the ex-post nash property in large games. Theoretical Economics 10 (1), 103–129.

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

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

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

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

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

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

Levy, H., 1973. Stochastic dominance among log-normal prospects. International Economic Review 14 (3), 601–614.

Manova, K., Zhang, Z., 2012. Export prices across firms and destinations. The Quarterly Journal of Economics 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.

Menezes-Filho, N. A., Muendler, M.-A., 2011. Labor reallocation in response to trade reform. NBER Working Paper 17372.

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

Øksendal, B., 2003. Stochastic differential equations. Springer.

Parent, D., 1999. Methods of pay and earnings: A longitudinal analysis. Industrial and Labor Relations Review 53 (1), 71–86.

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

Piketty, T., Saez, E., 2003. Income inequality in the united states, 1913–1998. The Quarterly Journal of Economics 118 (1), 1–41.

Ritter, M., forthcoming. Inequality and international trade: The role of skill-biased technology and search frictions. Macro Dynamics.

Schaettler, H., Sung, J., 1993. The first-order approach to the continuous-time principal–agent problem with exponential utility. Journal of Economic Theory 61 (2), 331–371.

Shorrocks, A. F., 1980. The class of additively decomposable inequality measures. Econometrica 48 (3), 613–626.

Statistics Canada, 2003. Guide to the analysis of workplace and employee survey. Catalogue, 71–221–GIE.

Sun, Y., 2006. The exact law of large numbers via fubini extension and characterization of insurable risks. Journal of Economic Theory 126 (1), 31–69.

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

Tito, M. D., 2015. Worker-firm matching in a global economy. University of British Columbia, Mimeo.

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

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

Wolfstetter, E., 1999. Topics in microeconomics: Industrial organization, auctions, and incentives. Cambridge University Press.

Appendices

(to appear as supplementary online materials)

A Technical Appendix

A.1 Proof of Lemma 1

For any worker i and task duration $\Delta \equiv 1/T$, let $\left\{\mathcal{F}^{\Delta}_{i\tau'}\right\}_{\tau'=0}^{T}$ be a sequence of σ-fields on the underlying probability space, where $\mathcal{F}^{\Delta}_{i0}$ is the trivial σ-field and $\mathcal{F}^{\Delta}_{i1}$,...,$\mathcal{F}^{\Delta}_{iT}$ is the filtration generated by the random variables z_{i1},...,z_{iT}. Let $Z^{\Delta}_{i\tau}$ denote the (normalized) cumulative performance of worker i up to task $\tau \in \{1, ..., T\}$ with initial condition $z_{i0} = 0$; i.e., $Z^{\Delta}_{i\tau} \equiv \Delta^{1/2} \sum_{\tau'=1}^{\tau} z_{i\tau'}$.[64] Let $\overline{z_{i\tau}} \equiv E\left(z_{i\tau} | \mathcal{F}^{\Delta}_{i\tau-1}\right) = \mu(\epsilon^{\Delta}_{i\tau})\Delta^{1/2}$ for $\tau \in \{1, ..., T\}$, where effort is allowed to be history-dependent; i.e., $\epsilon^{\Delta}_{i\tau} = \epsilon^{\Delta}_{i\tau}(z_{i1}, ..., z_{i\tau-1})$. Then, for $\tau \in \{0, ..., T\}$,

$$Z^{\Delta}_{i\tau} \equiv \Delta \sum_{\tau'=1}^{\tau} \mu(\epsilon^{\Delta}_{i\tau'}) + B^{\Delta}_{i\tau}, \tag{A-1}$$

where

$$B^{\Delta}_{i\tau} \equiv \Delta^{1/2} \sum_{\tau'=1}^{\tau} \left(z_{i\tau'} - \overline{z_{i\tau'}}\right)$$

denotes the cumulative deviation of worker i's performance from its expected value up to task τ. Equation (A-1) is an identity, by definitions of $B^{\Delta}_{i\tau}$ and $Z^{\Delta}_{i\tau}$. For the remainder of the proof, I suppress the subscript i to simplify notation.

Let $Z^{\Delta}(t)$, $t \in [0, 1]$, be the piecewise linear interpolation defined in equation (4). Then, using equation (A-1),

$$\begin{aligned}
Z^{\Delta}(t) &= \left(1 - \frac{t}{\Delta} + \left\lfloor \frac{t}{\Delta} \right\rfloor\right) \left[\Delta \sum_{\tau'=1}^{\lfloor t/\Delta \rfloor} \mu(\epsilon^{\Delta}_{\tau'}) + B^{\Delta}_{\lfloor t/\Delta \rfloor}\right] + \\
&\quad + \left(\frac{t}{\Delta} - \left\lfloor \frac{t}{\Delta} \right\rfloor\right) \left[\Delta \sum_{\tau'=1}^{\lfloor t/\Delta \rfloor + 1} \mu(\epsilon^{\Delta}_{\tau'}) + B^{\Delta}_{\lfloor t/\Delta \rfloor + 1}\right], \\
&= \Delta \sum_{\tau'=1}^{\lfloor t/\Delta \rfloor} \mu(\epsilon^{\Delta}_{\tau'}) + B^{\Delta}(t) + \left(\frac{t}{\Delta} - \left\lfloor \frac{t}{\Delta} \right\rfloor\right) \Delta b(\epsilon^{\Delta}_{\lfloor t/\Delta \rfloor + 1}), \\
&= \Delta \sum_{\tau'=1}^{\lfloor t/\Delta \rfloor} \mu(\epsilon^{\Delta}_{\tau'}) + B^{\Delta}(t) + o(\Delta), \\
&= \int_{\Delta}^{\lfloor t/\Delta \rfloor \Delta} \mu(\epsilon^{\Delta}(t'))dt' + B^{\Delta}(t) + o(\Delta).
\end{aligned} \tag{A-2}$$

In the third line, $B^{\Delta}(t)$ is the piecewise linear interpolation between the points $(0, 0)$, $\left(\Delta, B^{\Delta}_1\right)$, $\left(2\Delta, B^{\Delta}_2\right)$,...,$\left(1, B^{\Delta}_T\right)$. More specifically,[65]

$$B^{\Delta}(t) = \left(1 - \frac{t}{\Delta} + \left\lfloor \frac{t}{\Delta} \right\rfloor\right) B^{\Delta}_{\lfloor t/\Delta \rfloor} + \left(\frac{t}{\Delta} - \left\lfloor \frac{t}{\Delta} \right\rfloor\right) B^{\Delta}_{\lfloor t/\Delta \rfloor + 1}, \tag{A-3}$$

[64] Throughout, I adhere to the convention that the value of an empty sum of numbers is zero. Thus, for example, $Z^{\Delta}_{i0} = 0$.

[65] This interpolation is equivalent to that employed in Brown (1971). Unlike the definition in equation (A-3), the procedure in Brown (1971) includes adjustments by the martingale's variance at different points of the process. For the case considered in this paper, however, $E\left[\left(\varepsilon^{\Delta}_{\tau}\right)^2\right] = \Delta\tau$ for all τ, by equation (A-8). It is then straightforward to show that the interpolation of $\left\{\varepsilon^{\Delta}_{\tau}\right\}_{\tau=0}^{T}$ following Brown's procedure is equivalent to equation (A-3).

In the fourth line above, $o(\Delta) \to 0$ a.s. as $\Delta \to 0$, which follows from the boundedness of $\mu(\cdot)$ and the fact that $(t/\Delta - \lfloor t/\Delta \rfloor) \in [0,1]$.[66] Finally, the fifth line introduces a change of integration variables, $t' = \Delta\tau'$, and a continuous-time representation $\epsilon^{\Delta}(t)$, where $\epsilon^{\Delta}(t) \equiv \epsilon^{\Delta}_{\lfloor t/\Delta \rfloor}$.

Next, I characterize the convergence of each term on the right-hand side of equation (A-2), as $\Delta \to 0$. Regarding the first term, $\epsilon^{\Delta}(t) \to \epsilon(t)$ a.s. as $\Delta \to 0$ and the continuity of $\mu(\cdot)$ imply that $\mu\left(\epsilon^{\Delta}(t)\right) \to \mu\left(\epsilon(t)\right)$ a.s. as $\Delta \to 0$, by the continuous mapping theorem. Moreover, since $\mu(\cdot)$ is bounded, the bounded convergence theorem implies

$$\int_{\Delta}^{\lfloor t/\Delta \rfloor \Delta} \mu(\epsilon^{\Delta}(t'))dt' \to \int_{0}^{t} \mu(\epsilon(t'))dt' \text{ a.s. as } \Delta \to 0, \tag{A-4}$$

using the fact that $\lfloor t/\Delta \rfloor \Delta \to t$ as $\Delta \to 0$.

Regarding the limit of the second term on the right-hand side of (A-2), I start by noting two properties of B^{Δ}_{τ}. First, for all $s, \tau \in \{0, ..., T\}$ such that $s < \tau$, applying the law of iterated expectations yields

$$\begin{aligned} E\left(B^{\Delta}_{\tau}|\mathcal{F}^{\Delta}_{s}\right) &= \Delta^{1/2}\sum\nolimits_{\tau'=1}^{\tau} E\left(z_{\tau'} - \overline{z_{\tau'}}|\mathcal{F}^{\Delta}_{s}\right) \\ &= \Delta^{1/2}\sum\nolimits_{\tau'=1}^{s} \left(z_{\tau'} - \overline{z_{\tau'}}\right) + \Delta^{1/2}\sum\nolimits_{\tau'=s+1}^{\tau} \left[E\left(z_{\tau'}|\mathcal{F}^{\Delta}_{s}\right) - E\left(z_{\tau'}|\mathcal{F}^{\Delta}_{s}\right)\right] \\ &= B^{\Delta}_{s}. \end{aligned} \tag{A-5}$$

Second, for any τ, $E\left[\left(z_{\tau} - \overline{z_{\tau}}\right)^2|\mathcal{F}^{\Delta}_{\tau-1}\right] = 1 + \left(\overline{z_{\tau}}\right)^2 - \overline{z_{\tau}}E\left[z_{\tau}|\mathcal{F}^{\Delta}_{\tau-1}\right] = 1$. This implies

$$\sum_{\tau=1}^{T} E\left[\left(B^{\Delta}_{\tau} - B^{\Delta}_{\tau-1}\right)^2|\mathcal{F}^{\Delta}_{t-1}\right] = E\left[\sum_{\tau=1}^{T} E\left[\left(B^{\Delta}_{\tau} - B^{\Delta}_{\tau-1}\right)^2|\mathcal{F}^{\Delta}_{t-1}\right]\right] \text{ a.s.,} \quad \text{for all } \Delta. \tag{A-6}$$

Conditions (A-5) and (A-6) imply that, for given Δ, the process $\left\{B^{\Delta}_{\tau}\right\}_{\tau=0}^{T}$ belongs to the class of zero-mean, square-integrable martingales relative to $\left\{\mathcal{F}^{\Delta}_{\tau}\right\}_{\tau=0}^{T}$ studied in Brown (1971).[67] In particular, Theorem 3 in Brown (1971) implies that, as $\Delta \to 0$, the sequence of probability measures determined by the distribution of $\left\{B^{\Delta}(t); 0 \le t \le 1\right\}$ converges weakly to the Wiener measure in the space $C[0,1]$ with the uniform norm, provided that the Lindeberg condition holds, namely

$$E\left[\left(B^{\Delta}_{T}\right)^2\right]^{-2} \sum_{\tau=1}^{T} E\left[\Delta\left(z_{\tau} - \overline{z_{\tau}}\right)^2 I\left(\Delta^{1/2}\left|z_{\tau} - \overline{z_{\tau}}\right| \ge \delta E\left[\left(B^{\Delta}_{T}\right)^2\right]\right)\right] \to_p 0, \tag{A-7}$$

[66] Note that $b(\cdot)$ is a continuous function defined on the closed interval $[\epsilon_L, \epsilon_H]$, which ensures that $b(\cdot)$ is also bounded.

[67] Brown (1971) considers zero-mean, square-integrable martingales $\left\{\varepsilon^{\Delta}_{\tau}\right\}_{\tau=0}^{T}$ that satisfy

$$\frac{\sum_{\tau=1}^{T} E\left[\left(\varepsilon^{\Delta}_{\tau} - \varepsilon^{\Delta}_{\tau-1}\right)^2|\mathcal{F}^{\Delta}_{t-1}\right]}{E\left[\sum_{\tau=1}^{T} E\left[\left(\varepsilon^{\Delta}_{\tau} - \varepsilon^{\Delta}_{\tau-1}\right)^2|\mathcal{F}^{\Delta}_{t-1}\right]\right]} \to_p 1,$$

as $\Delta \to 0$, which is implied by (A-6).

as $\Delta \to 0$, for all $\delta > 0$, where $I(\cdot)$ is the indicator function.

To check (A-7), first note that $E\left[\left(B_T^\Delta\right)^2\right] = 1$ for all Δ, since martingale increments are uncorrelated[68] and thus

$$\begin{aligned}
E\left[\left(B_\tau^\Delta\right)^2\right] &= \Delta E\left[\left(\sum\nolimits_{\tau'=1}^{\tau}\left(z_{\tau'} - \overline{z_{\tau'}}\right)\right)^2\right], \\
&= \Delta E\sum\nolimits_{\tau'=1}^{\tau}\left(z_{\tau'} - \overline{z_{\tau'}}\right)^2, \\
&= \Delta E\sum_{\tau'=1}^{\tau} E\left[\left(z_{\tau'} - \overline{z_{\tau'}}\right)^2 | \mathcal{F}_{\tau'-1}^\Delta\right], \\
&= \Delta E\sum_{\tau'=1}^{\tau}\left[1 + \left(\overline{z_{\tau'}}\right)^2 - \overline{z_{\tau'}}E\left[z_{\tau'}|\mathcal{F}_{\tau'-1}^\Delta\right]\right], \\
&= \Delta\tau. \qquad \text{(A-8)}
\end{aligned}$$

Second, let $\mu_H = \sup_x |\mu(x)| < \infty$, since $\mu(\cdot)$ is bounded. Therefore, for any τ,

$$\Delta^{1/2}\left|z_\tau - \overline{z_\tau}\right| < \Delta^{1/2}\left(1 + \Delta^{1/2}\mu_H\right) \to 0, \quad \text{as } \Delta \to 0.$$

Letting K^Δ denote the left-hand side of (A-7) yields

$$\begin{aligned}
K^\Delta &= \sum_{\tau=1}^{T} E\left[\Delta\left(z_\tau - \overline{z_\tau}\right)^2 I\left(\Delta^{1/2}\left|z_\tau - \overline{z_\tau}\right| \geq \delta\right)\right], \\
&< I\left(\Delta^{1/2}\left(1 + \Delta^{1/2}\mu_H\right) \geq \delta\right)\sum_{\tau=1}^{T} E\left[\Delta\left(z_{i\tau} - \overline{z_{i\tau}}\right)^2\right], \\
&= I\left(\Delta^{1/2}\left(1 + \Delta^{1/2}\mu_H\right) \geq \delta\right).
\end{aligned}$$

Therefore, $K^\Delta \to 0$ a.s. as $\Delta \to 0$, which is sufficient to verify the Lindeberg condition (A-7). Therefore, by Theorem 3 in Brown (1971),

$$B^\Delta(t) \to^d B(t), \qquad \text{(A-9)}$$

in the space $C[0,1]$ with the uniform norm, where $B(t)$ is a Wiener process on $0 \leq t \leq 1$, such that $B(0) = 0$ a.s. and $E\left[B(1)^2\right] = 1$.

Applying the results (A-4) and (A-9) to equation (A-2), I conclude that, as $\Delta \to 0$, $Z^\Delta(t)$

[68] That is, for $\tau_1 > \tau_0 \geq 0$, let $s_{10} = E\left[\left(z_{\tau_0} - \overline{z_{\tau_0}}\right)\left(z_{\tau_1} - \overline{z_{\tau_1}}\right)\right]$. Then, applying the law of iterated expectations twice yields

$$\begin{aligned}
s_{10} &= E\left[\left(z_{\tau_0} - \overline{z_{\tau_0}}\right)E\left[\left(z_{\tau_1} - \overline{z_{\tau_1}}\right)|\mathcal{F}_{\tau_0}^\Delta\right]\right], \\
&= E\left[\left(z_{\tau_0} - \overline{z_{\tau_0}}\right)E\left[\left(z_{\tau_1} - z_{\tau_1}\right)|\mathcal{F}_{\tau_0}^\Delta\right]\right], \\
&= 0.
\end{aligned}$$

converges in distribution to a stochastic process $Z(t)$ in $C[0,1]$, such that

$$Z(t) = \int_0^t \mu\left(\epsilon\left(t'\right)\right) dt' + B(t),$$

which completes the proof.

A.2 Proof of Proposition 1

The proof proceeds in three steps:

Step 1. From the first-order conditions of worker i's problem, I show that if a contract $w_i\left(Z_i^1\right)$ implements the stochastic process $\epsilon_i \equiv \{\epsilon_i(t); t \in [0,1]\}$ with certain equivalent χ_i, then

$$\ln w_i(Z_i^1) = \ln \chi_i + \int_0^1 \frac{k'(\epsilon_i(t))}{\mu'(\epsilon_i(t))} \left[dZ_i(t) - \mu(\epsilon_i(t))dt\right] - \frac{1}{2}\int_0^1 \left[\frac{k'(\epsilon_i(t))}{\mu'(\epsilon_i(t))}\right]^2 dt. \tag{A-10}$$

I start by introducing a change of variables, letting $s_i(Z_i^1) \equiv \ln w_i\left(Z_i^1\right)$, to re-write the problem of worker i -constraint (iii) in Problem (8)- as

$$\max_{\boldsymbol{\epsilon}_i} E\left[\exp\left(s_i(Z_i^1) - \int_0^1 c(\epsilon_i(t'))dt'\right)\right] \tag{A-11}$$

$$s.t \quad Z_i(t) = \textstyle\int_0^t \mu\left(\epsilon_i\left(t'\right)\right) dt' + B_i(t) \ .$$

Formulated in this way, the worker's problem is similar to that in Holmström and Milgrom (1987), although they work with *negative* exponential utility -CARA- and set $\mu(x) = x$. I thus modify the proof of Theorem 6 in Holmström and Milgrom (1987) to allow for *positive* exponential utility and a general (differentiable) function $\mu(\cdot)$ to accommodate (A-11). To simplify notation, I suppress subscript i.

Let $\{\mathcal{F}_t\}_{0\leq t\leq 1}$ denote the filtration generated by the path of observed performance $Z(\cdot)$. Suppose that, given a contract $s(Z^1)$, an $\mathcal{F}_t$-adapted process ϵ solves problem (A-11) with certain equivalent χ. Let

$$\begin{aligned} F(\tau;\boldsymbol{\epsilon}';\mathbf{m}) &\equiv E_m\left[\exp\left(s(Z^1) - \int_0^\tau c(\epsilon'(t))dt - \int_\tau^1 c(m(t))dt\right)|\mathcal{F}_\tau\right] \\ &= F(\tau;\boldsymbol{\epsilon};\mathbf{m})K\left(\tau;\boldsymbol{\epsilon}'\right), \end{aligned}$$

where

$$K\left(\tau;\boldsymbol{\epsilon}'\right) \equiv \exp\left(\int_0^\tau \left[k(\epsilon(t)) - k(\epsilon'(t))\right] dt\right).$$

F is the conditional expected utility at time τ if the worker has followed an effort sequence ϵ' for tasks $[0,\tau]$ and then switches to a sequence $\mathbf{m}$ for the remainder of the production process. Let $V\left(\tau;\boldsymbol{\epsilon}'\right)$ be the maximal value of the worker's problem given the information at time τ if the worker has followed an effort sequence ϵ' for tasks $[0,\tau]$. Then,

$$\begin{aligned} V(\tau;\boldsymbol{\epsilon}') &\equiv \max_m F(\tau;\boldsymbol{\epsilon}';\mathbf{m}) \\ &= \max_m F(\tau;\boldsymbol{\epsilon};\mathbf{m})C(\tau;\boldsymbol{\epsilon}') = V(\tau;\boldsymbol{\epsilon})\,C(\tau;\boldsymbol{\epsilon}'). \end{aligned} \tag{A-12}$$

Since $V(\tau;\boldsymbol{\epsilon}) = F(\tau;\epsilon;\epsilon)$, the law of iterated expectations implies $E_\epsilon\left[V(\tau';\boldsymbol{\epsilon})\,|\mathcal{F}_\tau\right] = V(\tau;\boldsymbol{\epsilon})$ for $\tau' > \tau \geq 0$. Therefore, $V(\tau;\boldsymbol{\epsilon})$ is a martingale relative to $\{\mathcal{F}_\tau\}_{0\leq t\leq 1}$. Since ϵ is $\mathcal{F}_\tau$-adapted, $V(\tau;\boldsymbol{\epsilon})$ is also a martingale relative to the filtration generated by the driftless Brownian motion $Z(\tau) - \int_0^\tau \mu(\epsilon(t))\,dt$. By the martingale representation theorem (e.g. Øksendal (2003), ch. 4), there exists a unique, square-integrable and $\mathcal{F}_\tau$-measurable stochastic process $\gamma \equiv \{\gamma(t); t\in[0,1]\}$ such that

$$dV(\tau;\boldsymbol{\epsilon}) = \gamma(\tau)d\left[Z(\tau) - \int_0^\tau \mu(\epsilon(t))\,dt\right]. \tag{A-13}$$

For any effort sequence ϵ', $dZ = \mu(\epsilon') + dB$; thus $dV(\tau;\boldsymbol{\epsilon}) = \gamma\left[\mu(\epsilon') - \mu(\epsilon)\right]dt + \gamma dB$. Together with (A-12), this implies

$$\begin{aligned} dV(\tau;\boldsymbol{\epsilon}') &= d\left[V(\tau;\boldsymbol{\epsilon})\,K(\tau;\boldsymbol{\epsilon}')\right] \\ &= \left\{\gamma\left[\mu(\epsilon') - \mu(\epsilon)\right] + \left[k(\epsilon) - k(\epsilon')\right]V(\tau;\epsilon)\right\}K(\tau;\epsilon')\,dt \\ &\quad +\gamma K(\tau;\epsilon')\,dB. \end{aligned}$$

By the Principle of Optimality, if ϵ' is an optimal effort sequence, then it maximizes the drift of $V(\tau;\boldsymbol{\epsilon}')$. By hypothesis, ϵ is optimal for the worker; thus it satisfies the following first-order necessary condition:

$$\gamma(t)\,\mu'(\epsilon(t)) = k'(\epsilon(t))\,V(t;\epsilon), \tag{A-14}$$

for all $t \in [0,1]$.

Let $\chi(t)$ denote the certain equivalent corresponding to $V(t;\boldsymbol{\epsilon})$. Therefore, $\chi(t)$ satisfies $E\left[\chi(t)\exp\left(-\int_0^1 c(\epsilon(t'))dt'\right)|\mathcal{F}_t\right] = V(t;\boldsymbol{\epsilon})$. Solving for $\chi(t)$ yields

$$\chi(t) = \frac{V(t;\boldsymbol{\epsilon})}{E\left[\exp\left(-\int_0^1 k(\epsilon(t'))dt'\right)|\mathcal{F}_t\right]}. \tag{A-15}$$

The derivative of the denominator in (A-15) with respect to t is zero. Thus,

$$\begin{aligned} \frac{d\chi(t)}{\chi(t)} &= \frac{dV(t;\boldsymbol{\epsilon})}{V(t;\boldsymbol{\epsilon})} \\ &= \frac{\gamma(t)}{V(t;\boldsymbol{\epsilon})}d\left[Z(t) - \int_0^t \mu(\epsilon(t'))dt'\right] \\ &= \frac{k'(\epsilon(t))}{\mu'(\epsilon(t))}d\left[Z(t) - \int_0^t \mu(\epsilon(t'))dt'\right], \end{aligned} \tag{A-16}$$

where the second and third lines follow from (A-13) and (A-14), respectively. Using Ito's Lemma for the function $\ln(\cdot)$ yields

$$\begin{aligned} d\ln\chi(t) &= \frac{d\chi(t)}{\chi(t)} - \frac{1}{2}\frac{1}{[\chi(t)]^2}[d\chi(t)]^2 \\ &= \frac{d\chi(t)}{\chi(t)} - \frac{1}{2}\frac{1}{[\chi(t)]^2}\left[\chi(t)\frac{k'(\epsilon(t))}{\mu'(\epsilon(t))}\right]^2 dt \\ &= \frac{d\chi(t)}{\chi(t)} - \frac{1}{2}\left[\frac{k'(\epsilon(t))}{\mu'(\epsilon(t))}\right]^2 dt. \end{aligned} \tag{A-17}$$

Using (A-16) to substitute for $d\chi(t)/\chi(t)$, integrating (A-17) and letting $\chi(0) = \chi$ yields

$$\ln\chi(1) = \ln\chi + \int_0^1 \frac{k'(\epsilon(t))}{\mu'(\epsilon(t))}[dZ(t) - \mu(\epsilon(t))dt] - \frac{1}{2}\int_0^1\left[\frac{k'(\epsilon(t))}{\mu'(\epsilon(t))}\right]^2 dt. \tag{A-18}$$

By the construction of $\chi(t)$, $\chi(1) = \exp[s(Z^1)]$; thus, $\chi(1) = w(Z^1)$. Substituting the latter in (A-18) delivers (A-10).

Step 2. Following the 'first-order' approach in the principal-agent literature (Schaettler and Sung (1993)), I formulate and solve the firm's *relaxed* optimal contracting problem, in which the incentive compatibility constraints in Problem (8) are replaced with the contract representations obtained in step 1. Importantly, the solution to this problem is not necessarily implementable by the contracts (A-10), since the latter were derived from only *necessary* conditions for optimality in the worker's problem. This issue is tacked in Step 3.

To obtain the firm's relaxed problem, insert (A-10) in the objective function of Problem (8) and drop the incentive compatibility constraints:

$$\begin{aligned} &\min_{\{\chi_i,\epsilon_i;i\in[0,h]\}} \int_0^h \chi_i E\left[\exp\left(\int_0^1 \frac{k'(\epsilon_i(t))}{\mu'(\epsilon_i(t))}[dZ_i(t) - \mu(\epsilon_i(t))dt] - \frac{1}{2}\int_0^1\left[\frac{k'(\epsilon_i(t))}{\mu'(\epsilon_i(t))}\right]^2 dt\right)\right] di \\ &s.t \quad (i)\ n_0 \geq -h^{-1}\textstyle\int_0^h E[Z_i(1)]\,di \\ &\qquad (ii)\ Z_i(t) = \textstyle\int_0^t \mu(\epsilon_i(t'))\,dt' + B_i(t), \text{ for } i\in[0,h] \\ &\qquad (iii)\ E[U(w_i,\epsilon_i)] \geq \overline{u}, \text{ for } i\in[0,h] \end{aligned} \tag{A-19}$$

The following steps simplify this problem. First, substitute $E[Z_i(1)] = \int_0^1 E[\mu(\epsilon_i(t))]\,dt$ in (i). Second, substitute (ii) in the objective function and use the fact that

$$E\left[\exp\left(\int_0^1 \frac{k'(\epsilon_i(t))}{\mu'(\epsilon_i(t))}dB_i(t) - \frac{1}{2}\int_0^1\left[\frac{k'(\epsilon_i(t))}{\mu'(\epsilon_i(t))}\right]^2 dt\right)\right] = 1,$$

for any $\boldsymbol{\epsilon}_i$.[69] Third, individual rationality constraints should bind at the optimum.[70] Expression (iii) can then be used to solve for χ_i as a function of $\boldsymbol{\epsilon}_i$,

$$\chi_i = \frac{\overline{u}}{E\left[\exp\left(-\int_0^1 k(\epsilon_i)dt\right)\right]}. \tag{A-20}$$

Problem (A-19) is then simplified to a problem of finding the optimal $\boldsymbol{\epsilon}_i$ that minimizes the certainty equivalent χ_i for $i \in [0,h]$, subject to a single performance constraint:

$$\min_{\{\boldsymbol{\epsilon}_i; i\in[0,h]\}} \int_0^h \frac{\overline{u}}{E\left[\exp\left(-\int_0^1 k(\epsilon_i)dt\right)\right]} di, \tag{A-21}$$

$$s.t \quad n_0 \geq -h^{-1} \int_0^h \int_0^1 E\left[\mu\left(\epsilon_i\left(t\right)\right)\right] dtdi.$$

Note that both the objective and constraint of problem (A-21) are independent of $B_i(t)$ for all $i \in [0,h]$ and $t \in [0,1]$. This implies that, without loss of generality, the domain for admissible effort sequences in (A-21) can be restricted to the set of deterministic (history-independent) sequences.

Dropping the expectations operator and disregarding the (positive) constant $\overline{u}$, (A-21) can be further simplified into

$$\min_{\{\boldsymbol{\epsilon}_i; i\in[0,h]\}} \int_0^h \exp\left[\int_0^1 k(\epsilon_i\left(t\right))dt\right] di \quad s.t \quad n_0 \geq -h^{-1} \int_0^h \int_0^1 \mu\left(\epsilon_i\left(t\right)\right) dtdi. \tag{A-22}$$

It is convenient to analyze this problem by introducing a set of auxiliary choice variables $\{a_i \in (-\mu\left(\epsilon_H\right), -\mu\left(\epsilon_L\right)]; i \in [0,h]\}$, satisfying $n_0 = h^{-1}\int_0^h a_i di$, interpreted as an allocation of $n_0 h$ mistakes across h workers. I compute the solution to (A-22) sequentially with the following two-step procedure:

(1) For given $\{a_i \in (-\mu\left(\epsilon_H\right), -\mu\left(\epsilon_L\right)]; i \in [0,h]\}$, determine the effort sequence $\boldsymbol{\epsilon}_i$ that solves, for each i,

$$\min_{\boldsymbol{\epsilon}_i} \int_0^1 k(\epsilon_i\left(t\right))dt \quad s.t \quad a_i \geq \int_0^1 -\mu\left(\epsilon_i\left(t\right)\right) dt. \tag{A-23}$$

Under the assumptions that $k(\cdot)$ and $\mu\left(\cdot\right)$ are convex and strictly concave, respectively, it is straightforward to verify that the solution to (A-23) is a unique constant effort for all $t \in [0,1]$, denoted $\epsilon(a_i)$, that satisfies $a_i = -\mu\left(\epsilon(a_i)\right)$. In addition, $a_i \in (-\mu\left(\epsilon_H\right), -\mu\left(\epsilon_L\right)]$ implies $\epsilon(a_i) \in (\epsilon_L, \epsilon_H]$.

[69] Under the assumption that $k(\cdot)$ and $\mu(\cdot)$ have continuous derivatives, then $k'(\cdot)/\mu'(\cdot)$ is bounded for all $\epsilon_i(t) \in [\epsilon_L, \epsilon_H]$. Then the process $\exp\left(\int_0^\tau \frac{k'(\epsilon_i(t))}{\mu'(\epsilon_i(t))} dB_i(t) - \frac{1}{2}\int_0^\tau \left[\frac{k'(\epsilon_i(t))}{\mu'(\epsilon_i(t))}\right]^2 dt\right)$ is an $\mathcal{F}_\tau$-martingale with expected value equal to one, for all τ and $\boldsymbol{\epsilon}$. See Karatzas and Shreve (1988), p.200.

[70] Suppose that IR constraints didn't bind for a positive measure of workers. Then it would be possible to decrease the certainty equivalent of these workers, shifting down their corresponding wage functions while holding the effort sequences constant.

(2) Given $\epsilon(a_i)$, determine the optimal allocation of mistakes across workers that solves

$$\min_{\substack{\{a_i\in[-\mu(\epsilon_H),-\mu(\epsilon_L)];\\ i\in[0,h]\}}} \int_0^h \exp\left[k(\epsilon(a_i))\right] di \quad s.t \quad n_0 = h^{-1}\int_0^h a_i di. \tag{A-24}$$

Since $\exp[k(\cdot)]$ is strictly convex, the solution to (A-24) is a unique constant allocation of mistakes across workers satisfying the constraint of the problem; that is, $a_i = n_0$ for all $i \in [0, h]$.

In light of these results, I conclude that there exists a unique solution to problem (A-19), in which every worker exerts an identical constant effort throughout the production process. This solution, denoted ϵ^*, satisfies $n_0 = -\mu(\epsilon^*)$ for all $i \in [0, h]$ and $t \in [0, 1]$. In addition, $n_0 \in (-\mu(\epsilon_H), -\mu(\epsilon_L)]$ implies $\epsilon^* \in (\epsilon_L, \epsilon_H]$.

Step 3. I check the validity of the first-order approach by verifying that the solution to Problem (A-19) is implementable. That is, I show that if worker i is assigned contract (A-10) evaluated at effort ϵ^*, then a constant effort ϵ^* is the a.s. unique maximizer of worker i's expected utility. This step is needed because the wage representations in step 1 were derived from only *necessary* conditions for optimality in the worker's problem.[71]

Evaluating the wage representation (A-10) at a constant effort ϵ^* and using the expression for the certainty equivalent (A-20) at $t = 1$, yields[72]

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

Define constants α^* and β^* such that,

$$\begin{aligned} \alpha^* &\equiv \ln \overline{u} + k(\epsilon^*) - \frac{k'(\epsilon^*)}{\mu'(\epsilon^*)}\mu(\epsilon^*) - \frac{1}{2}\left[\frac{k'(\epsilon^*)}{\mu'(\epsilon^*)}\right]^2, \\ \beta^* &\equiv \frac{k'(\epsilon^*)}{\mu'(\epsilon^*)}. \end{aligned}$$

The worker's problem becomes,

$$\max_{\boldsymbol{\epsilon}} E\left[\exp\left(\alpha^* + \beta^* Z_i(1)\right) - \int_0^1 k(\epsilon(t))dt\right] \quad s.t \;\; Z_i(1) = \int_0^1 \mu(\epsilon(t))dt + B_i(1), \tag{A-25}$$

where $\boldsymbol{\epsilon}$ is an adapted process. Substituting the constraint in the objective function and

[71] Schaettler and Sung (1993) provide an in-depth analysis of the first-order approach to the moral hazard problem in a continuous-time environment.

[72] From (A-20), the certainty equivalent for a constant effort ϵ^* is $\chi_i = \overline{u}/E\left[\exp\left(-\int_0^1 k(\epsilon_i)dt\right)\right] = \overline{u}/\exp\left(-k(\epsilon^*)\right)$. Therefore, $\ln \chi_i = \ln \overline{u} + k(\epsilon^*)$.

rearranging yields,

$$
\begin{aligned}
& E\left[\exp\left(\alpha^* + \beta^* Z_i(1)\right) - \int_0^1 k(\epsilon(t))dt\right] \\
= \quad & \exp(\alpha^*)E\left[\exp\left(\int_0^1 \left[\beta^*\mu(\epsilon(t)) - k(\epsilon(t))\right]dt + \beta^* B_i(1)\right)\right].
\end{aligned}
$$

The distribution of $B_i(1)$ is independent of $\epsilon(t)$, for any $t \in [0,1]$. This implies that, for any realization of $B_i(1)$, utility is maximized if and only if ϵ maximizes

$$
J(\epsilon) \equiv \int_0^1 \left[\beta^*\mu(\epsilon(t)) - k(\epsilon(t))\right]dt.
$$

Any effort strategy that mandates the worker to deviate from maximizing J will reduce expected utility. Therefore, optimal effort is deterministic. Clearly, J is maximized when effort in task t, $\epsilon(t)$, maximizes $\beta^*\mu(\epsilon(t)) - k(\epsilon(t))$ for all $t \in [0,1]$. The convexity and strict concavity of $k(\cdot)$ and $\mu(\cdot)$, respectively, imply that there is an a.s. unique effort, denoted $\widehat{\epsilon}$, which is constant for all $t \in [0,1]$ and solves the worker's problem (A-25). In particular, $\widehat{\epsilon}$ satisfies

$$
\beta^* = \frac{k'(\widehat{\epsilon})}{\mu'(\widehat{\epsilon})} = \frac{k'(\epsilon^*)}{\mu'(\epsilon^*)}.
$$

where the second equality follows by the definition of β^*.

Under Assumption (1), $k'(\cdot)/\mu'(\cdot)$ is a strictly increasing function. It follows that $\widehat{\epsilon} = \epsilon^*$, and therefore ϵ^* is a.s.uniquely implemented by contract $\ln w_i = \alpha^* + \beta^* Z_i(1)$.

A.3 Proof of Corollary 1

From Proposition 1, the (stochastic) wage of worker i is

$$
w_i^* = \overline{u}e^{k(\epsilon)+\beta(\epsilon)\left[B_i(1)-\frac{1}{2}\beta(\epsilon)\right]}, \tag{A-26}
$$

for all $i \in [0,h]$, where $\beta(\epsilon) = k'(\epsilon)/\mu'(\epsilon)$ and $\epsilon_L < \epsilon \leq \epsilon_H$.[73] Because $B_i(1)$ is normally distributed with mean zero and unit variance, it follows that $E\left[\exp\left(\beta(\epsilon)B_i(1)\right)|\epsilon\right] = \exp\left[\beta(\epsilon)^2/2\right]$. Therefore, $E\left[w_i^*|\epsilon\right] = \overline{u}\exp\left[k(\epsilon)\right]$, as stated in part (a) of Corollary 1.

For part (b), consider two effort levels ϵ_1 and ϵ_2 such that $\epsilon_L < \epsilon_1 < \epsilon_2 \leq \epsilon_H$. From Proposition 1,

$$
Var\left[\log\left(w_i^*\right)|\epsilon\right] = \beta(\epsilon)^2,
$$

since $Z_i(1)$ is normally distributed with mean $\mu(\epsilon)$ and unit variance. Under Assumption (1), $0 < \beta(\epsilon_1) < \beta(\epsilon_2)$, which establishes (i).

To prove statement (ii) of part (b), first rescale firm-level wages by their mean to obtain normalized wages. If the firm implements effort ϵ_j, $j = \{1,2\}$, the normalized wages can be

[73] If $\epsilon = \epsilon_L$, Corollary 1 holds trivially. In this case, $\beta(\epsilon_L) = 0$ and thus $w_i^* = \overline{u}\exp\left[k(\epsilon_L)\right]$ for all i. Note also that $\beta(\epsilon) > 0$ and thus inequality is strictly positive whenever $\epsilon_L < \epsilon \leq \epsilon_H$. The rest of the proof focuses on the case $\epsilon_L < \epsilon \leq \epsilon_H$.

written as

$$\widehat{w}_{ij}(x) \equiv \frac{w_i^*}{E\left[w_i^* | \epsilon_j\right]} = e^{\beta(\epsilon_j)\left[B_i(1) - \frac{1}{2}\beta(\epsilon_j)\right]},$$

where the second equality follows from (A-26) and part (a) of Corollary 1. Normalized wages are log-normally distributed; in particular,

$$\ln\left(\widehat{w}_{ij}(x)\right) \overset{d}{\sim} N\left(-\frac{1}{2}\beta(\epsilon_j)^2, \beta(\epsilon_j)^2\right).$$

Observe that: (a) $E\left[\ln\left(\widehat{w}_{ij}(x)\right)|\epsilon_j\right]$ is strictly decreasing in ϵ_j; (b) $Var\left[\ln\left(\widehat{w}_{ij}(x)\right)|\epsilon_j\right]$ is strictly increasing in ϵ_j; (c) $E\left[\ln\left(\widehat{w}_{ij}(x)\right)|\epsilon_j\right] = -Var\left[\ln\left(\widehat{w}_{ij}(x)\right)|\epsilon_j\right]/2$, for $j = \{1, 2\}$. Conditions (a), (b) and (c) are sufficient to conclude that the firm-level distribution of normalized wages when the firm implements effort ϵ_1 second-order stochastically dominates the firm-level distribution of normalized wages when the firm implements effort ϵ_2 (see Levy (1973), Theorem 5). This completes the proof of (ii).

A.4 Profit Maximization

A.4.1 Optimality of $q(\theta)$

In this section, I show that the expression for firm quality (19) is optimal in problem (10); i.e., $q(\theta) = q_d(\theta) = q_x(\theta)$, under the functional form assumptions (16)-(18). By equation (12), this is equivalent to showing that $q(\theta)$ is the unique global minimizer of the average cost of quality $c^\theta(q)/q$ for $q \in [q_L, q_H]$.

Recall that $c^\theta(q) \equiv \omega(\epsilon(\theta, q))/\theta^s$, where $\epsilon(\theta, q)$ is implicitly defined by $q = q(\theta, -\mu(\epsilon))$. Under (16)-(18),

$$\begin{aligned} c^\theta(q) &= \overline{u}\exp\left[k\epsilon(\theta, q)\right]\theta^{-s}, \\ &= \overline{u}\exp\left[k\left(\frac{\log q}{(\gamma\log\theta)^z}\right)^{1/(1-z)}\right]\theta^{-s}, \\ &= \overline{u}\exp\left[\Lambda\left(\log q\right)^{1/(1-z)}\right]\theta^{-s}, \end{aligned} \tag{A-27}$$

where $\Lambda \equiv k/\left(\gamma\log\theta\right)^{z/(1-z)}$ is independent of q. For any $q > 0$,

$$\frac{d\left(c^\theta(q)/q\right)}{dq} = \overline{u}\theta^{-s}\exp\left[\Lambda\left(\log q\right)^{1/(1-z)}\right]q^{-2}\left(\frac{\Lambda}{1-z}\left(\log q\right)^{z/(1-z)} - 1\right). \tag{A-28}$$

Moreover, $\overline{u}\theta^{-s}\exp\left[\Lambda\left(\log q\right)^{1/(1-z)}\right]q^{-2} > 0$ for all $\theta \in [\theta_L, \theta_H]$ and $q \in [q_L, q_H]$.

From (19),

$$\left[\log\left(q(\theta)\right)\right]^{z/(1-z)} = \left[\frac{1-z}{k}\right]\left[\gamma\log(\theta)\right]^{z/(1-z)} = \frac{1-z}{\Lambda}. \tag{A-29}$$

For any $q > 0$, (A-28) and (A-29) imply

$$\frac{d\left(c^{\theta}(q)/q\right)}{dq} = \begin{cases} <0 & \text{if} \quad q < q(\theta), \\ =0 & \text{if} \quad q = q(\theta), \\ >0 & \text{if} \quad q > q(\theta). \end{cases}$$

If $q(\theta) \in [q_L, q_H]$, then $q(\theta)$ is the unique global minimizer of the average cost of quality $c^{\theta}(q)/q$ and therefore optimal in problem (10). As shown in the text, $q(\theta) \in [q_L, q_H]$ provided that effort is defined over a sufficiently large interval.

A.4.2 Equilibrium Unit Costs and Prices Across Firms

This section analyses the variation in equilibrium unit costs and output prices across firms. Let $c_*^{\theta} \equiv c^{\theta}(q(\theta))$ denote the equilibrium unit cost in firm θ. Then,

$$\begin{aligned} c_*^{\theta} &= \overline{u}\exp\left[k\left(\frac{\log q(\theta)}{(\gamma\log\theta)^z}\right)^{1/(1-z)}\right]\theta^{-s}, \\ &= \overline{u}\exp\left[\gamma k\left(\frac{1-z}{k}\right)^{1/z}\log(\theta)\right]\theta^{-s}, \\ &= \overline{u}\theta^{\gamma\left[(1-z)^{1/z}k^{(1-z)/z}\right]-s}, \end{aligned}$$

where the first and second lines follow from (A-27) and (19), respectively. Equilibrium unit costs increase across firms if and only if $\gamma\left[(1-z)^{1/z}k^{(1-z)/z}\right] > s$. This pattern reflects two countervailing forces. First, optimal quality increases in θ. Higher γ and k imply a higher elasticity of quality with respect to productivity and a higher cost of effort, respectively. Second, if $s > 0$ then labor productivity increases in θ. Higher s implies that the firm requires fewer workers to produce one unit of output.

With CES demand, output prices are constant mark-ups over marginal costs. Therefore, for sufficiently high γ and k or small s, the model delivers positive correlations between output prices, average wages, employment and revenue across firms, which is consistent with the empirical evidence documented in Kugler and Verhoogen (2012).

A.5 Proof of Proposition 2

For $\theta_m \in [\theta_L, \theta_H]$, $0 < \theta_L$, $J(\theta_m)$ is positive and finite provided that the distribution of firm productivity has a finite Γ-th uncentered moment. Therefore, given f_x, f_d, f_e, Γ and $G_{\theta}(\theta)$, (26) and (27) yield positive and finite equilibrium cutoffs θ_d and θ_x. Since $J(\theta_m)$ is monotonically, the cutoffs are uniquely determined. Moreover, since $\lim_{\theta_m \to 0} J(\theta_m) = \infty$, there exists a sufficiently small θ_L such that $0 < \theta_L < \theta_d$.

Given the productivity cutoffs, I obtain E and M_e as one-to-one functions of $\overline{u}$ from (24) and (28); that is,

$$E = \frac{f_d(\overline{u})^{\rho/(1-\rho)}}{\kappa_r(1-\rho)\theta_d^{\Gamma}},$$

and

$$M_e = \frac{(1-\rho) f_d (\overline{u})^{\rho/(1-\rho)}}{[f_e + f_d [1 - G_\theta(\theta_d)] + f_x [1 - G_\theta(\theta_x)]] \kappa_r (1-\rho) \theta_d^\Gamma}.$$

Note that, under the assumed parameter restrictions, E and M_e are positive if and only if $\overline{u}$ is positive.

Inserting these two expressions in the labor market clearing condition (29) and solving for $\overline{u}$ yields

$$(\overline{u})^{(1-2\rho)/(1-\rho)} = \frac{\kappa_y (f_d)^2 \int_{\theta_d}^{\theta_H} \Upsilon(\theta) \theta^{\Gamma - k\kappa_\epsilon} dG_\theta(\theta)}{L(1-\rho) \left[\kappa_r \theta_d^\Gamma\right]^2 [f_e + f_d [1 - G_\theta(\theta_d)] + f_x [1 - G_\theta(\theta_x)]]}.$$

Since G_θ has a finite Γ-th uncentered moment, the right-hand side of this expression is a finite positive number. This ensures the existence of a unique and positive equilibrium reservation utility $\overline{u}$ and, consequently, the existence of unique and positive equilibrium values for E and M_e.

A.6 Proof of Lemma 2

Preliminaries. From equation (30), re-write the employment distribution across firms in equilibrium $j \in \{0, 1\}$ as

$$G_{h,j}(\theta) = \begin{cases} 0, & \theta_L \leq \theta \leq \theta_{d,j}, \\ D_j^{-1} \int_{\theta_{d,j}}^{\theta} (\theta')^{\Gamma - k\kappa_\epsilon} dG_\theta(\theta'), & \theta_{d,j} \leq \theta \leq \theta_{x,j}, \\ D_j^{-1} \int_{\theta_d}^{\theta} \Upsilon_j(\theta') (\theta')^{\Gamma - k\kappa_\epsilon} dG_\theta(\theta'), & \theta_{x,j} \leq \theta \leq \theta_H, \end{cases} \tag{A-30}$$

where

$$\begin{aligned} D_j &\equiv \int_{\theta_{d,j}}^{\theta_H} \Upsilon_j(\theta') (\theta')^{\Gamma - k\kappa_\epsilon} dG_\theta(\theta'), \\ \Upsilon_j(\theta') &= \begin{cases} \Upsilon_{x,j} & \text{if } \theta' \geq \theta_{x,j}, \\ 1 & \text{if } \theta' < \theta_{x,j}, \end{cases}, \\ \Upsilon_{x,j} &= 1 + (f_x/f_d) (\theta_{d,j}/\theta_{x,j})^\Gamma. \end{aligned} \tag{A-31}$$

Letting $\theta_{x,0} \to \theta_H$ or $\theta_{x,1} \to \theta_{d,1}$ in (A-30) yields the employment distribution corresponding to autarky or to the equilibrium in which all firms export following trade liberalization, respectively.

To determine how cutoffs change in response to trade liberalization, first note that since $\iota_0 > \iota_1$ by hypothesis, then $\theta_{d,1}/\theta_{x,1} > \theta_{d,0}/\theta_{x,0}$ by equation (26) and thus $\Upsilon_{x,1} > \Upsilon_{x,0} > 1$. In addition, the free entry condition (27) implies that $\theta_{d,0} < \theta_{d,1}$ if and only if $\theta_{x,0} > \theta_{x,1}$; i.e., cutoffs change in opposite directions as a result of trade liberalization. Given $\iota_0 > \iota_1$, $\iota_1 \in [\underline{\iota}, \overline{\iota})$, and assuming that productivity is defined over a sufficiently large interval (see section 5), then $\theta_L < \theta_{d,0} < \theta_{d,1} \leq \theta_{x,1} < \theta_{x,0}$, where $\theta_{d,1} = \theta_{x,1}$ if all firms export following trade liberalization. Moreover, $\iota_0 \geq \bar{\iota}$ if and only if $\theta_{x,0} \geq \theta_H$. As a result, the productivity cutoffs in equilibria before and after trade liberalization satisfy either: (i) $\theta_L < \theta_{d,0} < \theta_{d,1} \leq \theta_{x,1} < \theta_{x,0} < \theta_H$ if $\iota_0 < \bar{\iota}$ or (ii) $\theta_L < \theta_{d,0} < \theta_{d,1} \leq \theta_{x,1} < \theta_H < \theta_{x,0}$ if $\iota_0 \geq \bar{\iota}$.

For the analysis below, it is convenient to partition the domain of θ into regions denoted by R_i, $i = \{A, B, C, D, E\}$. When the initial equilibrium is not autarkic, $\iota_0 < \bar{\iota}$, let $R_A = [\theta_L, \theta_{d,0}]$, $R_B = [\theta_{d,0}, \theta_{d,1}]$, $R_C = [\theta_{d,1}, \theta_{x,1}]$, $R_D = [\theta_{x,1}, \theta_{x,0}]$, $R_E = [\theta_{x,0}, \theta_H]$. Otherwise, if $\iota_0 \geq \bar{\iota}$, let $R_A = [\theta_L, \theta_{d,0}]$, $R_B = [\theta_{d,0}, \theta_{d,1}]$, $R_C = [\theta_{d,1}, \theta_{x,1}]$, $R_D = [\theta_{x,1}, \theta_H]$. In either case, if all firms export in $j = 1$, then $\theta_{x,1} = \theta_{d,1}$ and $R_C = \{\theta_{d,1}\}$.

The following preliminary result introduces three properties of $G_{h,j}$, denoted P1, P2 and P3, that are used repeatedly in the proof of Lemma 2.

Lemma A-1 *Let $(s_C, s_D, s_E) = (\Lambda, \Lambda/\Upsilon_{x,1}, \Lambda\Upsilon_{x,0}/\Upsilon_{x,1})$, where $\Lambda \equiv D_1/D_0$ is positive and independent of θ.*

***P1**. For $j \in \{0, 1\}$, $G_{h,j}(\theta)$ is non-decreasing, continuous and piecewise differentiable in $[\theta_L, \theta_H]$. Moreover, for $i = \{C, D, E\}$, $G'_{h,j}(\theta) > 0$ for all θ in the interior of R_i.*

***P2**. If $\theta^* \in R_i$ and $\theta^{**} \in R_i$, $i = \{C, D, E\}$, then*

$$G_{h,1}(\theta^*) - G_{h,0}(\theta^*) = G_{h,1}(\theta^{**}) - G_{h,0}(\theta^{**}) + (1 - s_i) \int_{\theta^{**}}^{\theta^*} G'_{h,1}(v)dv. \tag{A-32}$$

***P3**. Suppose that $G_{h,0}$ and $G_{h,1}$ intersect at a point $\widetilde{\theta}$ in R_i, for $i = \{C, D, E\}$. If R_i has a non-empty interior, then:*

(i) $G_{h,1}$ intersects $G_{h,0}$ once in R_i, from below, if and only if $s_i < 1$.

(ii) $G_{h,1}$ intersects $G_{h,0}$ once in R_i, from above, if and only if $s_i > 1$.

(iii) $G_{h,0}(\theta) = G_{h,1}(\theta)$ for all θ in R_i if and only if $s_i = 1$.

Proof. P1 is immediately verified from (A-30). To establish P2, first note that the *slopes* of $G_{h,0}$ and $G_{h,1}$ satisfy the following 'proportionality property' in the interior of regions C, D and E:

$$G'_{h,0}(\theta) = \begin{cases} \Lambda G'_{h,1}(\theta), & \text{if } \theta \in R_C, \\ \frac{\Lambda}{\Upsilon_{x,1}} G'_{h,1}(\theta), & \text{if } \theta \in R_D, \\ \Lambda \frac{\Upsilon_{x,0}}{\Upsilon_{x,1}} G'_{h,1}(\theta), & \text{if } \theta \in R_E. \end{cases} \tag{A-33}$$

Next, fix a region R_i, $i = \{C, D, E\}$, and consider $\theta^* \in R_i$ and $\theta^{**} \in R_i$. Then,

$$\begin{aligned} G_{h,1}(\theta^*) - G_{h,0}(\theta^*) &= G_{h,1}(\theta^{**}) + [G_{h,1}(\theta^*) - G_{h,1}(\theta^{**})] \\ &\quad - G_{h,0}(\theta^{**}) - [G_{h,0}(\theta^*) - G_{h,0}(\theta^{**})], \\ &= G_{h,1}(\theta^{**}) - G_{h,0}(\theta^{**}) + \int_{\theta^{**}}^{\theta^*} \left[G'_{h,1}(v) - G'_{h,0}(v)\right] dv, \\ &= G_{h,1}(\theta^{**}) - G_{h,0}(\theta^{**}) + (1 - s_i) \int_{\theta^{**}}^{\theta^*} G'_{h,1}(v)dv, \end{aligned}$$

where the last line uses (A-33).

P3 is a corollary of P2. Fix a region i, $i = \{C, D, E\}$. For any $\theta \in R_i$, $\theta \neq \widetilde{\theta}$, setting $\theta = \theta^*$ and $\widetilde{\theta} = \theta^{**}$ in expression (A-32) yields

$$G_{h,1}(\theta) - G_{h,0}(\theta) = (1 - s_i) \int_{\widetilde{\theta}}^{\theta} G'_{h,1}(v)dv. \tag{A-34}$$

By P1, $G'_{h,1} > 0$ in the interior of R_i. Therefore, $s_i < 1$ if and only if $G_{h,1}(\theta) < G_{h,0}(\theta)$ for $\theta < \widetilde{\theta}$ and $G_{h,1}(\theta) > G_{h,0}(\theta)$ for $\theta > \widetilde{\theta}$. In this case, $G_{h,1}(\theta)$ intersects $G_{h,0}(\theta)$ once, from below at point $\widetilde{\theta}$.[74] A similar argument establishes the single-crossing property from above if and only if $s_i > 1$. Finally, setting $s_i = 1$ in (A-34) yields $G_{h,1}(\theta) = G_{h,0}(\theta)$ for all $\theta \in R_i$. For the converse, if $G_{h,1}(\theta) = G_{h,0}(\theta)$ for some $\theta \in R_i$, then P1 and (A-34) imply $s_i = 1$. ■

Part (a) of Lemma 2. If $\iota_0 \geq \iota$, the initial equilibrium is autarkic. I show that $G_{h,1}(\theta) \leq G_{h,0}(\theta)$ in each region of the domain of θ (i.e. R_i, $i = \{A, B, C, D\}$), with strict inequality for some $\theta \in [\theta_L, \theta_H]$.

For region D, note that $G_{h,1}$ intersects $G_{h,0}$ at point θ_H, where $\theta_H \in R_D$. Suppose that $G_{h,1}$ intersects $G_{h,0}$ from above at point θ_H. Then $\Upsilon_{x,1} > 1$ and (P3) imply $s_C > s_D > 1$. In addition, by (P3) the intersection is unique in R_D, thus $G_{h,1}(\theta_{x,1}) > G_{h,0}(\theta_{x,1})$ by continuity (P1). Since $\theta_{d,1} \in R_C$ and $\theta_{x,1} \in R_C$, let $\theta^* = \theta_{d,1}$ and $\theta^{**} = \theta_{x,1}$ in P2, which implies $G_{h,1}(\theta_{d,1}) > 0$, which is false. Then, $G_{h,1}$ must intersect $G_{h,0}$ from below at point θ_H. By (P3) the intersection is unique in R_D, thus $G_{h,1}(\theta) \leq G_{h,0}(\theta)$ for all $\theta \in R_D$, with equality if and only if $\theta = \theta_H$. For region C, $\theta_{x,1} \in R_D \cap R_C$ implies $G_{h,1}(\theta_{x,1}) < G_{h,0}(\theta_{x,1})$. Since $G_{h,1}(\theta_{d,1}) = 0 < G_{h,0}(\theta_{d,1})$, continuity (P1) ensures that $G_{h,1}$ and $G_{h,0}$ do not intersect in the interior of region C. Therefore, $G_{h,1}(\theta) < G_{h,0}(\theta)$ for all $\theta \in R_C$. For $\theta \in R_B$, $G_{h,0}(\theta) \geq 0 = G_{h,1}(\theta)$, with strict inequality if $\theta > \theta_{d,0}$. Finally, for $\theta \in R_A$, $G_{h,0}(\theta) = 0 = G_{h,1}(\theta)$.

Part (b) of Lemma 2.

Case: $G_{h,1}(\widehat{\theta}) \leq G_{h,0}(\widehat{\theta})$ **for some** $\widehat{\theta} \in [\theta_{x,0}, \theta_H)$**.** I show that $G_{h,1}(\theta) \leq G_{h,0}(\theta)$ in each region of the domain of θ (i.e. R_i, $i = \{A, B, C, D, E\}$), with strict inequality for some $\theta \in [\theta_L, \theta_H]$.

First, since $\widehat{\theta} \in R_E$, then $G_{h,1}(\widehat{\theta}) \leq G_{h,0}(\widehat{\theta})$ implies that $G_{h,1}$ intersects $G_{h,0}$ from below at point θ_H. By P3, $s_E \leq 1$ and $G_{h,1}(\theta) \leq G_{h,0}(\theta)$ for all $\theta \in R_E$. For region D, $s_E \leq 1$ and $\Upsilon_{x,0} > 1$ imply $s_D < 1$. Note that $\theta_{x,0} \in R_D \cap R_E$ and $\theta_{x,0} \geq \theta$ for all $\theta \in R_D$. Therefore, letting $\theta^{**} = \theta_{x,0}$ in P2 yields $G_{h,1}(\theta) < G_{h,0}(\theta)$ for $\theta < \theta_{x,0}$ for all $\theta \in R_D$. For region C, suppose that $G_{h,1}$ intersects $G_{h,0}$ at point θ_I in R_C. Since $\theta_{x,1} \in R_C \cap R_D$ and $G_{h,1}(\theta_{x,1}) < G_{h,0}(\theta_{x,1})$, P3 implies that the intersection is from above and unique. Moreover, $s_C > 1$. Letting $\theta^* = \theta_{d,1}$ in P2,

$$G_{h,1}(\theta_{d,1}) - G_{h,0}(\theta_{d,1}) = G_{h,1}(\theta_I) - G_{h,0}(\theta_I) + (1 - s_C) \int_{\theta_I}^{\theta_{d,1}} G'_{h,1}(v)dv \geq 0,$$

[74] Technically, this argument applies to any θ in the interior of R_i. However, the continuity of $G_{h,j}$ ensures that the conclusion can be extended to the boundary of R_i.

with strict inequality if $\theta_{d,1} < \theta_I$. But then $G_{h,0}(\theta_{d,1}) > 0$ implies $G_{h,1}(\theta_{d,1}) > 0$, which is false. Therefore, $G_{h,1}$ does not intersect $G_{h,0}$ in R_C. Since both employment functions are continuous by P1, $G_{h,1}(\theta_{d,1}) = 0 < G_{h,0}(\theta_{d,1})$ and $G_{h,1}(\theta_{x,1}) < G_{h,0}(\theta_{x,1})$ imply $G_{h,1}(\theta) < G_{h,0}(\theta)$ for all $\theta \in R_C$. For $\theta \in R_B$, $G_{h,0}(\theta) \geq 0 = G_{h,1}(\theta)$, with strict inequality if $\theta > \theta_{d,0}$. Finally, for $\theta \in R_A$, $G_{h,0}(\theta) = 0 = G_{h,1}(\theta)$.

Case: $G_{h,1}(\widehat{\theta}) > G_{h,0}(\widehat{\theta})$ for some $\widehat{\theta} \in [\theta_{x,0}, \theta_H)$. I show that $G_{h,1}$ intersects $G_{h,0}$ once, from below, in regions C, D and interior of E.

Since $\widehat{\theta} \in R_E$, then $G_{h,1}$ intersects $G_{h,0}$ from above at point θ_H. By P3, $s_E > 1$ and $G_{h,1}(\theta) > G_{h,0}(\theta)$ for all θ in the interior of region E. Next, suppose that $G_{h,1}$ does not intersect $G_{h,0}$ in region D. By P1, both employment distributions are continuous and thus $G_{h,1}(\theta_{x,0}) > G_{h,0}(\theta_{x,0})$ implies $G_{h,1}(\theta_{x,1}) > G_{h,0}(\theta_{x,1})$. Letting $\theta^* = \theta_{d,1}$ in P2,

$$G_{h,1}(\theta_{d,1}) - G_{h,0}(\theta_{d,1}) = G_{h,1}(\theta_{x,1}) - G_{h,0}(\theta_{x,1}) + (1 - s_D) \int_{\theta_{x,1}}^{\theta_{d,1}} G'_{h,1}(v)dv > 0,$$

since $s_D > s_E > 1$. But then $G_{h,1}(\theta_{d,1}) > 0$, which is false. Therefore $G_{h,1}$ intersects $G_{h,0}$ in region D. Moreover, by P3 the intersection is unique and, because $G_{h,1}(\theta_{x,0}) > G_{h,0}(\theta_{x,0})$, from below. This implies that, in region C, $G_{h,1}(\theta_{x,1}) \leq G_{h,0}(\theta_{x,1})$, with equality if and only if $G_{h,1}$ intersects $G_{h,0}$ at point $\theta_{x,1}$. Since $G_{h,1}(\theta_{d,1}) = 0 < G_{h,0}(\theta_{d,1})$, continuity (P1) ensures that $G_{h,1}$ and $G_{h,0}$ do not intersect in the interior of region C.

A.7 Proof of Proposition 3

In light of Lemma 2, it is sufficient to show that Assumption (2) implies $G_{h,1}(\widehat{\theta}) \leq G_{h,0}(\widehat{\theta})$ for some $\widehat{\theta} \in [\theta_{x,0}, \theta_H)$, when $\iota_0 < \bar{\iota}$. From the employment distribution (30),

$$1 - G_{h,j}(\widehat{\theta}) = \frac{\int_{\widehat{\theta}}^{\theta_H} \theta^{\Gamma - k\kappa_\epsilon} dG_\theta(\theta)}{(\Upsilon_{x,j})^{-1} \int_{\theta_{d,j}}^{\theta_{x,j}} \theta^{\Gamma - k\kappa_\epsilon} dG_\theta(\theta) + \int_{\theta_{x,j}}^{\theta_H} \theta^{\Gamma - k\kappa_\epsilon} dG_\theta(\theta)}, \tag{A-35}$$

for $j \in \{0, 1\}$, where $(\Upsilon_{x,j})^{-1} = 1/(1 + (\iota_j)^{-\rho/(1-\rho)}) \in (0, 1)$. Therefore, $G_{h,1}(\widehat{\theta}) \leq G_{h,0}(\widehat{\theta})$ if and only if the denominator of (A-35) is increasing in the variable trade cost. Without loss of generality, I focus on infinitesimal changes in ι and show that $D'(\iota) > 0$, where

$$D(\iota) \equiv (\Upsilon_x)^{-1} \int_{\theta_d}^{\theta_x} \theta^{\Gamma - k\kappa_\epsilon} dG_\theta(\theta) + \int_{\theta_x}^{\theta_H} \theta^{\Gamma - k\kappa_\epsilon} dG_\theta(\theta),$$

after dropping index j to simplify notation. Note that Υ_x, θ_d and θ_x are functions of ι. Therefore,

$$
\begin{aligned}
D'(\iota) &= \frac{\rho}{1-\rho}\left(\Upsilon_x\right)^{-2}\iota^{-1/(1-\rho)}\int_{\theta_d}^{\theta_x}\theta^{\Gamma-k\kappa_\epsilon}dG_\theta(\theta)+ \\
&\quad +\left[\left(\Upsilon_x\right)^{-1}-1\right]\left(\theta_x\right)^{\Gamma-k\kappa_\epsilon}\theta_x'(\iota)g_\theta(\theta_x)-\left(\Upsilon_x\right)^{-1}\left(\theta_d\right)^{\Gamma-k\kappa_\epsilon}\theta_d'(\iota)g_\theta(\theta_d), \\
&> \left(\Upsilon_x\right)^{-1}\left(\theta_d\right)^{\Gamma-k\kappa_\epsilon}\left\{\left[\frac{\left(\Upsilon_x\right)^{-1}-1}{\left(\Upsilon_x\right)^{-1}}\right]\left(\frac{\theta_x}{\theta_d}\right)^{\Gamma-k\kappa_\epsilon}\theta_x'(\iota)g_\theta(\theta_x)-\theta_d'(\iota)g_\theta(\theta_d)\right\}, \\
&> \left(\Upsilon_x\right)^{-1}\left(\theta_d\right)^{\Gamma-k\kappa_\epsilon}\left\{-\frac{f_x}{f_d}\theta_x'(\iota)g_\theta(\theta_x)-\theta_d'(\iota)g_\theta(\theta_d)\right\}, \qquad \text{(A-36)}
\end{aligned}
$$

where $\theta_m'(\iota)\equiv d\theta_m/d\iota$ for $m\in\{d,x\}$. The last line follows from $\left(\Upsilon_x\right)^{-1}-1=-\left(\Upsilon_x\right)^{-1}\left(f_x/f_d\right)\left(\theta_d/\theta_x\right)^{\Gamma}$ (by definition of Υ_x), $\theta_x>\theta_d>0$ and $\Gamma>k\kappa_\epsilon>0$.

Recall that, in a symmetric equilibrium, the expression for relative cutoffs (26), the free entry condition (27) and the monotonicity of $J(\cdot)$ imply $\theta_d'(\iota)<0$ and $\theta_x'(\iota)>0$. Therefore, from (A-36), $D'(\iota)>0$ if

$$\frac{f_x}{f_d}\frac{g_\theta(\theta_x)}{g_\theta(\theta_d)}<-\frac{\theta_d'(\iota)}{\theta_x'(\iota)}=\frac{f_x}{f_d}\frac{J'(\theta_x)}{J'(\theta_d)},$$

where the equality follows by applying the Implicit Function Theorem on the free entry condition (27). To conclude, $D'(\iota)>0$ if

$$\frac{J'(\theta_d)}{g_\theta(\theta_d)}<\frac{J'(\theta_x)}{g_\theta(\theta_x)},$$

which, given $\theta_x>\theta_d>0$, is guaranteed by Assumption (2).

A.8 Productivity Distributions that Satisfy Assumption (2)

Recall that $J(\theta)\equiv\int_\theta^{\theta_H}\left[(\upsilon/\theta)^\Gamma-1\right]dG(\upsilon)$. Therefore,

$$J'(\theta)=-\Gamma\theta^{-(\Gamma+1)}\int_\theta^{\theta_H}\upsilon^\Gamma dG(\upsilon), \qquad \text{(A-37)}$$

and

$$J''(\theta)=-\left(\Gamma+1\right)\frac{J'(\theta)}{\theta}+\Gamma\frac{g_\theta(\theta)}{\theta}. \qquad \text{(A-38)}$$

Densities with Elasticity Greater Than Or Equal To $-(\Gamma+1)$**.** From (A-38), it follows that

$$\begin{aligned} \frac{d}{d\theta}\left(\frac{J'(\theta)}{g_\theta(\theta)}\right) &\geq 0 \\ &\Leftrightarrow J''(\theta)g_\theta(\theta) \geq J'(\theta)g'_\theta(\theta), \\ &\Leftrightarrow -J'(\theta)\left[(\Gamma+1)\frac{1}{\theta}+\frac{g'_\theta(\theta)}{g_\theta(\theta)}\right] \geq -\Gamma\frac{g_\theta(\theta)}{\theta}. \end{aligned} \tag{A-39}$$

Next, let $\varepsilon(\theta) \equiv \theta g'_\theta(\theta)/g_\theta(\theta)$ denote the elasticity of $g_\theta(\theta)$ at point $\theta \in [\theta_L, \theta_H]$. Since $J'(\theta) < 0$, equation (A-39) implies

$$\varepsilon(\theta) \geq -(\Gamma+1) \Rightarrow \frac{d}{d\theta}\left(\frac{J'(\theta)}{g_\theta(\theta)}\right) \geq 0.$$

Therefore, Assumption (2) is satisfied by a class of productivity densities satisfying $\varepsilon(\theta) \geq -(\Gamma+1)$ for all $\theta \in [\theta_L, \theta_H]$. Distributions with non-decreasing densities satisfy $\varepsilon(\theta) \geq 0$ for all $\theta \in [\theta_L, \theta_H]$, thus they are included in this class.

Truncated Pareto Distribution. Consider $g_\theta(\theta) = z(\theta_L)^z\,\theta^{-z-1}/\left[1-(\theta_L/\theta_H)^z\right]$, $z > 0$, for $\theta \in [\theta_L, \theta_H]$. From equation (A-37),

$$\frac{J'(\theta)}{g_\theta(\theta)} = \frac{\Gamma}{z-\Gamma}\left[\theta^{z-\Gamma}\theta_H^{\Gamma-z}-1\right].$$

Therefore,

$$\frac{d}{d\theta}\left(\frac{J'(\theta)}{g_\theta(\theta)}\right) = \Gamma\theta_H^{\Gamma-z}\theta^{z-\Gamma-1} > 0,$$

for all $z > 0$. For a truncated Pareto distribution, $\varepsilon(\theta) = -(z+1)$ for all $\theta \in [\theta_L, \theta_H]$. Therefore, not every truncated Pareto belongs to the class of productivity densities satisfying $\varepsilon(\theta) \geq -(\Gamma+1)$ for all $\theta \in [\theta_L, \theta_H]$. Still, all of them satisfy Assumption (2).

A.9 Proof of Proposition 4

From part (a) of Corollary (1) and the expression for optimal effort (20), $\omega(\theta) = \overline{u}_j\theta^{k\kappa_\epsilon}$. From expression (31), this implies

$$dG_{w,j}(\theta) = \frac{\theta^{k\kappa_\epsilon}}{\int_{\theta_{d,j}}^{\theta_H} v^{k\kappa_\epsilon}dG_{h,j}(v)}dG_{h,j}(\theta),$$

for all $\theta \in [\theta_{d,j}, \theta_H]$. Next, let $\widetilde{\Lambda} \equiv \Lambda \widetilde{D}_1/\widetilde{D}_0 > 0$, where $\widetilde{D}_j \equiv \int_{\theta_{d,j}}^{\theta_H} v^{k\kappa_\epsilon} dG_{h,j}(v)$ for $j \in \{0,1\}$ and Λ is defined as in Lemma (A-1). Therefore,

$$G'_{w,0}(\theta) = \begin{cases} \widetilde{\Lambda} G'_{w,1}(\theta), & \text{if } \theta \in R_C, \\ \frac{\widetilde{\Lambda}}{\Upsilon_{x,1}} G'_{w,1}(\theta), & \text{if } \theta \in R_D, \\ \widetilde{\Lambda} \frac{\Upsilon_{x,0}}{\Upsilon_{x,1}} G'_{w,1}(\theta), & \text{if } \theta \in R_E. \end{cases}$$

where R_i is defined as in the proof of Lemma (2), for $i \in \{C, D, E\}$.

The slopes of $G_{w,0}$ and $G_{w,1}$ thus satisfy a proportionality property which is identical to the proportionality property for employment distributions, after a redefinition of the positive constant Λ, in expression (A-33). Since the exact definition of Λ is immaterial in the proof of Lemma (A-1), then $G_{w,0}$ and $G_{w,1}$ satisfy P1, P2 and P3, after replacing Λ with $\widetilde{\Lambda}$. It is then trivial to adjust the proof of Lemma (2) to show:

(a) If $\iota_0 \geq \bar{\iota}$, then $G_{w,1}$ first-order stochastically dominates $G_{w,0}$.

(b) If $\iota_0 < \bar{\iota}$, consider $\widehat{\theta} \in [\theta_{x,0}, \theta_H)$:

If $G_{w,1}(\widehat{\theta}) \leq G_{w,0}(\widehat{\theta})$, then $G_{w,1}$ first-order stochastically dominates $G_{w,0}$.

If $G_{w,1}(\widehat{\theta}) > G_{w,0}(\widehat{\theta})$, then $G_{w,1}$ intersects $G_{w,0}$ once, from below, in $[\theta_{d,1}, \theta_H)$.

To establish Proposition (4), it thus suffices to show that, if $\iota_0 < \bar{\iota}$, then Assumption (2) implies $G_{w,1}(\widehat{\theta}) \leq G_{w,0}(\widehat{\theta})$, for any $\widehat{\theta} \in [\theta_{x,0}, \theta_H)$.

Suppose that $\iota_0 < \bar{\iota}$ and fix $\widehat{\theta} \in [\theta_{x,0}, \theta_H)$. Using the cross-firm wage distribution (31) together with (i) $\omega(\theta) = \overline{u}_j \theta^{k\kappa_\epsilon}$, (ii) $dG_{h,j}(\theta) = M_j L^{-1} h_j(\theta) dG_{\theta,j}(\theta)$, (iii) $h_j(\theta) = \kappa_y \Upsilon_j(\theta) \left(A_j \overline{u}_j^{-1}\right)^{1/(1-\rho)} \theta^{\Gamma - k\kappa_\epsilon}$ and (iv) definition of $\Upsilon_j(\theta)$ in equation (A-31), yields

$$1 - G_{w,j}(\widehat{\theta}) = \frac{\int_{\widehat{\theta}}^{\theta_H} \theta^\Gamma dG_{\theta,j}(\theta)}{(\Upsilon_{x,j})^{-1} \int_{\theta_{d,j}}^{\theta_{x,j}} \theta^\Gamma dG_{\theta,j}(\theta) + \int_{\theta_{x,j}}^{\theta_H} \theta^\Gamma dG_{\theta,j}(\theta)}, \tag{A-40}$$

for $j \in \{0,1\}$, where $(\Upsilon_{x,j})^{-1} = 1/(1 + (\iota_j)^{-\rho/(1-\rho)}) \in (0,1)$. Therefore, $G_{w,1}(\widehat{\theta}) \leq G_{w,0}(\widehat{\theta})$ if and only if the denominator of (A-40) is increasing in the variable trade cost. I proceed by adjusting the proof of Proposition 3 to show that $D'(\iota) > 0$, where

$$D(\iota) \equiv (\Upsilon_x)^{-1} \int_{\theta_d}^{\theta_x} \theta^\Gamma dG_\theta(\theta) + \int_{\theta_x}^{\theta_H} \theta^\Gamma dG_\theta(\theta),$$

after dropping index j to simplify notation. Following the steps leading to equation (A-36) yields

$$D'(\iota) > (\Upsilon_x)^{-1} (\theta_d)^\Gamma \left\{ -\frac{f_x}{f_d} \theta'_x(\iota) g_\theta(\theta_x) - \theta'_d(\iota) g_\theta(\theta_d) \right\}. \tag{A-41}$$

As shown in the proof of Proposition 3, Assumption (2) guarantees that the right-hand side of (A-41) is positive. Therefore, Assumption (2) implies $G_{w,1}(\widehat{\theta}) \leq G_{w,0}(\widehat{\theta})$ for any $\widehat{\theta} \in [\theta_{x,0}, \theta_H)$, which completes the proof.

A.10 Proof of Proposition 5

I consider two countries with asymmetric parameterizations of labor endowments, trade costs, effort costs, technology and firm productivity distributions. More specifically, I allow parameters L, ι, f_d, f_x, s, k, γ, z and function $G_\theta(\theta)$ to differ across countries. I maintain, however, a constant and symmetric elasticity of substitution (constant ρ), thus allowing only a partial asymmetry in preferences (different k). The analysis focuses on Home. I use asterisks to indicate parameters and variables of Foreign.

With a common ρ, revenues from domestic and foreign sales are still written $r_d = Aq_d^\rho y_d^\rho$ and $r_x = A^* q_x^\rho [y_x/\iota]^\rho$, respectively, where $A \equiv E^{1-\rho}$ and $A^* \equiv (P^*)^\rho (E^*)^{1-\rho}$. I maintain the choice of numeraire ($P = 1$). This implies that the firm's profit maximization problem is still written as in (10) and therefore the firm's solutions from section 4.2 apply here as well. In particular, note that the expression for firm employment (23) and thus the distribution of employment across firms (30) are still valid.

In asymmetric equilibria, the cutoff conditions for Home are straightforward extensions of equations (24) and (25). In equilibrium $j \in \{0, 1\}$,

$$\kappa_r (1-\rho) \left(A_j \overline{u}_j\right)^{-\rho/(1-\rho)} \theta_{d,j}^\Gamma = f_d,$$

$$\kappa_r (1-\rho) \left[\Upsilon_{x,j} - 1\right] \left(A_j \overline{u}_j\right)^{-\rho/(1-\rho)} \theta_{x,j}^\Gamma = f_x,$$

where $\Upsilon_{x,j} \equiv 1 + \iota_j^{-\frac{\rho}{1-\rho}} \left(A_j^*/A_j\right)^{\frac{1}{1-\rho}}$. Dividing these conditions yields

$$\iota_j^{-\frac{\rho}{1-\rho}} \left(\frac{A_j^*}{A_j}\right)^{\frac{1}{1-\rho}} \left(\frac{\theta_{x,j}}{\theta_{d,j}}\right)^\Gamma = \frac{f_x}{f_d}, \tag{A-42}$$

in Home. Similarly, for Foreign,

$$\left(\iota_j^*\right)^{-\frac{\rho}{1-\rho}} \left(\frac{A_j}{A_j^*}\right)^{\frac{1}{1-\rho}} \left(\frac{\theta_{x,j}^*}{\theta_{d,j}^*}\right)^{\Gamma^*} = \frac{f_x^*}{f_d^*}. \tag{A-43}$$

Using (A-42) yields $\Upsilon_{x,j} = 1 + (f_x/f_d)\left(\theta_{d,j}/\theta_{x,j}\right)^\Gamma$, which ensures that the employment distribution across firms in Home (30) can be expressed solely as a function of the productivity cutoffs in Home in any asymmetric equilibrium.

Consider a unilateral trade liberalization in Home, $\iota_1^* < \iota_0^*$. If Home is a small open economy, then $\theta_{x,j}^*$, $\theta_{x,j}^*$ and A_j^* are independent of ι_j^*. From (A-43), $A_1 < A_0$. This implies $\theta_{x,0}/\theta_{d,0} < \theta_{x,1}/\theta_{d,1}$ from (A-42). The free entry condition in Home is still written as (27) in any asymmetric equilibrium. Therefore, $\theta_{d,0} < \theta_{d,1}$ and $\theta_{x,1} < \theta_{x,0}$. If Home is not a small open economy, suppose that $\theta_{d,0} < \theta_{d,1}$. Free entry again implies $\theta_{x,1} < \theta_{x,0}$. Restricting attention to equilibria in which the most productive firms export yields $\theta_{d,0} < \theta_{d,1} \leq \theta_{x,1} < \theta_{x,0}$ in both cases.

Given this configuration of the productivity cutoffs, it is then straightforward to verify that the proofs of Lemma (2) and Propositions (3) and (4) continue to hold in asymmetric equilibria, under the conditions stated in Proposition 5.

B Data Appendix

B.1 Data Description

The Workplace and Employee Survey is a matched employer-employee survey conducted by Statistics Canada between 1999 and 2005. The empirical analysis is restricted to cross-sectional data from the 2003 survey. There are two reasons for this. First, exploiting the panel dimension of the survey (e.g. to include worker fixed-effects in wage decompositions) would raise concerns of selection bias due to non-random attrition of workers. The WES followed employees for two years only, due to the difficulty of integrating new employers into the location sample as workers change companies. Fresh samples of employees were only drawn on every second survey occasion (i.e. in 1999, 2001, 2003 and 2005).[75] Second, while sample sizes in the WES declined over time, 2003 is the first year with an updated questionnaire that substantially improved the quality of performance pay data.[76]

In 2003, the target population for the employer component of the survey is defined as all business locations operating in Canada that have paid employees in March (Statistics Canada (2003)). The survey thus collects information at the level of the establishment rather than the firm. With slight abuse of language, however, I refer to establishments as firms in the empirical section of the paper. The target population for the employee component is all employees working or on paid leave in March in the selected workplaces who receive a Customs Canada and Revenue Agency T-4 Supplementary form (excludes self-employed workers). The WES draws its employer sample from the Business Register maintained by the Business Register Division of Statistics Canada, and from lists of employees provided by the surveyed employers. The response rates in 2003 were 83.1% and 82.7% of sampled workplaces and employees, respectively. The empirical work for this paper was carried out at the Toronto Research Data Centre (RDC) in the University of Toronto. Survey weights were used to produce all descriptive statistics, regression output and figures reported in the paper.

The baseline sample for the empirical analysis is composed of firms with at least two matched full-time employees 16 to 64 years of age. Full-time employees report an average of at least 30 paid hours per week in the current job, excluding overtime. I exclude non-profit firms and workers with missing values in the vector of individual characteristics (details below). The sample includes 14,265 workers and 3,540 firms. Tables B-I and B-II report

[75] Alternatively, the empirical analysis could have been carried out by pooling two or more years of data but not exploiting the panel structure. Relative to the empirical analysis in the paper, this approach would result in more efficient estimates only by imposing restrictions on the variation of estimated parameters over time. For example, assuming time-invariant returns to education, firm fixed effects or, more generally, firm-level wage-generating processes.

[76] In both 2001 and 2003, the employee questionnaires first ask for pre-tax wage or salary information and subsequently intend to measure earnings due to overtime and different forms of performance pay. The 2003 questionnaire explicitly asks, for *each* entry of overtime and performance pay, whether the reported amount was included in the wage or salary initially reported. The 2001 questionnaire, on the other hand, does not verify whether individual entries for overtime and performance pay were already included in the initial salary reported. Therefore, in 2003 it is possible to obtain a cleaner decomposition of total compensation into performance-pay and performance-independent components, net of any overtime payments.

descriptive statistics for workers and firms, respectively. Due to confidentiality constraints, the number of unweighted observations reported in descriptive statistics and regression output is rounded to the nearest multiple of 5, for all variables in the WES.

The individual compensation measure used in this section is the average weekly wage before taxes and deductions and net of overtime payments in the current job, over the twelve months prior to March 2003 (or period of time since start of job, if less than 12 months). This measure is constructed by adjusting the units in which gross total compensation is reported (hourly, daily, monthly, yearly, etc.) to its weekly equivalent using the appropriate conversion (e.g. usual paid hours per week) and substracting weekly-equivalent overtime payments. Performance pay is computed as weekly-equivalent tips, commissions, piecework payments and bonuses received by the worker. The performance-independent component is, in turn, computed as the gross weekly wage net of performance and overtime payments.

Table B-I - Descriptive Statistics: Employees

Variable	Non-PP Jobs		PP Jobs		Mean Test
	Mean	SD	Mean	SD	P-value
Weekly wage (C$)	722	455	903	664	0.000
Union membership (dummy)	0.11	0.32	0.04	0.08	0.000
Language mismatch (dummy)	0.89	0.31	0.91	0.29	0.311
Foreign born (dummy)	0.17	0.37	0.18	0.38	0.630
Female (dummy)	0.46	0.50	0.47	0.50	0.553
Tenure (years)	6.94	7.90	7.77	7.90	0.053
Education categories					
HS dropout	0.19	0.39	0.14	0.35	0.043
HS completed	0.20	0.40	0.21	0.41	0.616
Some college	0.25	0.43	0.28	0.45	0.232
College completed	0.24	0.43	0.23	0.42	0.535
University	0.13	0.34	0.15	0.35	0.294
Experience categories (years)					
0-10	0.35	0.48	0.28	0.45	0.002
11-20	0.29	0.45	0.31	0.46	0.351
21-30	0.22	0.42	0.28	0.45	0.019
31+	0.13	0.34	0.14	0.34	0.929
Manager (dummy)	0.15	0.36	0.22	0.41	0.002
Manufacturing worker (dummy)	0.16	0.37	0.12	0.14	0.000
Unweighted observations	7,490		6,775		

Notes: This table reports descriptive statistics for employees in the WES (2003) sample used in the paper. Workers in PP jobs received strictly positive performance pay in 2003. Detailed variable definitions are provided in the main text. Mean Test reports a t-test on the equality of means (columns 2 and 4).

The vector of observable worker characteristics contains 14 industry dummies (NAICS 2002), 47 occupation dummies (SOC 1991), tenure at the current job (years), a full set of interactions between 5 education dummies and 4 experience dummies, and indicators for the following binary variables: union membership (member of a union or covered by a collective bargaining agreement), gender, language mismatch between home and work (language most often used at work is not the language most often used at home), foreign-born worker.

The experience categories are a function of the employee's years of full-time working experience in all jobs held until 2003: 0-10, 11-20, 21-30 and 31 or more years of experience.

The educational categories are: (i) high school dropout; (ii) high school graduate; (iii) some college (trade or vocational diploma or certificate; some college, CEGEP, institute of technology or nursing school; some university; industry certified training or certification courses); (iv) college (completed college, CEGEP, institute of technology or nursing school; teachers' college; university certificate or diploma above bachelor level); (v) university (bachelor or undergraduate degree or teachers' college; university certificate or diploma above bachelor level; master's degree; doctorate).

The analysis also employs a number of firm characteristics. These include total annual revenue, total employment of full-time workers and export status (dummy equal to one if the firm exports in 2003).

Table B-II - Descriptive Statistics: Firms

Variable	Non-PP firm		PP Firms		Mean Test
	Mean	SD	Mean	SD	P-value
Revenue (CA$, in millions)	3.10	16.60	5.35	28.30	0.000
Employment (full time)	12.50	38.03	19.31	63.34	0.000
Exporter (dummy)	0.18	0.39	0.19	0.39	0.872
Manufacturing (dummy)	0.15	0.36	0.11	0.31	0.074
Unweighted observations	940		2,600		

Notes: This table reports descriptive statistics for employees in the WES (2003) sample used in the paper. PP Firms have at least one matched worker with strictly positive performance in 2003. Detailed variable definitions are provided in the main text. Mean Test reports a t-test on the equality of means (columns 2 and 4).

B.2 The Finite Population Correction

The WES features a relatively low number of matched workers per firm. Moreover, the latter typically increases with firm size. In light of this, a potential concern is whether the precision of estimated firm-level variances is correlated with firm size. In section 2.2, I use a finite population correction to construct unbiased estimates of firm-level variances, for both unadjusted log wages ($Var\left[\widetilde{w}|\theta\right]$) and residual log wages ($\psi_{\theta,2}$). I describe the procedure for unadjusted wages, without loss of generality.

The correction acknowledges that the sample of workers in each firm is drawn from a finite population (total employment in the firm) without replacement. Under this sampling scheme, it can be shown that $\lambda_{FPC}\sum_{i=1}^{n_\theta}\left(\widetilde{w}_i-\widetilde{w}_\theta\right)^2/n_\theta$ is an unbiased estimator of $Var\left[\widetilde{w}|\theta\right]$, where $\lambda_{FPC}\equiv\left[1-(1-f_\theta)/n_\theta\right]^{-1}$ is the finite population correction factor. n_θ, f_θ and $\widetilde{w}_\theta$ are the number of matched employees, the fraction of matched employees in firm employment and the mean log wage in firm θ, respectively.[77]

The finite population correction results in firm-level log wage variances that are, on average, 29% larger than the unadjusted sample variance. More importantly, for the purposes of this paper, the correction has only minor differential effects across firms. To see this, note that λ_{FPC} decreases in n_θ and f_θ, and converges to the usual sample variance estimator when firm employment is arbitrarily large; that is, when f_θ tends to zero. In the WES sample, the

[77] I thank Min Seong Kim for providing a proof of this result. Details are available upon request.

correlation between n_θ and f_θ across firms is -0.35. Moreover, the firm size (total revenue) is positively correlated with n_θ (0.26) and negatively correlated with f_θ (-0.18). Therefore, while smaller firms have fewer sampled workers, each of those workers represents a higher fraction of the firm's total employment. These two forces operate in opposite directions on the adjustment factor. Overall, λ_{FPC} is positively correlated with revenue, although the correlation is small (0.01). The correlation between λ_{FPC} and total employment is slightly larger (0.05).

To conclude, although the number of matched workers is systematically smaller for small firms in the WES, this does not result in differential biases in firm-level variance estimation across firms. Still, firm-level variances of unadjusted log wages ($Var\left[\widetilde{w}|\theta\right]$) and residual log wages ($\psi_{\theta,2}$) reported in section 2.2 are adjusted using the finite population correction described in this section.

B.3 Within-firm Inequality in Other Countries

Is within-firm inequality in other countries comparable to Canada's? For comparability with several studies in the literature, I compute the following extension of the decomposition (1):

$$Var(r\widetilde{w}_{i\theta}) = Var\left[\widehat{\psi}_\theta\right] + E\left[Var(r\widetilde{w}_{i\theta}|\theta, e_i)\right] + Var\left[\widehat{\phi}e_i\right] + 2Cov\left[\widehat{\phi}e_i, \widehat{\psi}_\theta\right]. \qquad \text{(B-1)}$$

For worker i employed in firm θ, $r\widetilde{w}_{i\theta}$ is the residual obtained from an OLS regression of log weekly wages on sector and occupation dummies. Parameters $\widehat{\phi}$ and $\widehat{\psi}_\theta$ are OLS estimates obtained from:

$$r\widetilde{w}_{i\theta} = \phi e_i + \psi_\theta + v_{i\theta}.$$

The four terms on the right-hand side of equation (B-1) decompose the variance of log wages within sectors and occupations into: between-firm inequality, within-firm inequality, inequality in worker observable characteristics and the covariance between worker observables and firm fixed effects. Relative to row 6 of Table I, the decomposition (B-1) accounts for the covariance between worker observables and firm fixed effects when computing the relative magnitudes of between- and within-firm inequality.

Table B-III collects results of decomposition (B-1) for Canada using the WES data and three other countries studied in the recent literature: Brazil, Sweden and France. See Akerman et al. (2013), Helpman et al. (2014) and Tito (2015), respectively. The empirical analyses in all of these papers use matched-employer data and report inequality decompositions based on (B-1). Naturally, there remain non-trivial differences in sample sizes, variable definitions and measurements across these data sets, among other important caveats, that demand caution when comparing the results in Table B-III across countries.

Table B-III shows that within-firm wage inequality is a major component of wage inequality within sectors and occupations in these four countries. In Canada, within-firm inequality as a share of within- plus between-firm inequality is 0.42 ($= 39/(39 + 53)$), similar to that reported in row 6 of Table I. Note that this share is lower in Canada than in other countries, a fact that might be partly attributable the relatively low number of matched employees per firm in the WES (see discussion in footnote of p. 10).

Table B-III - Decomposition of Log Wage Inequality Within Sectors and Occupations (%)

	Brazil (1994)	Sweden (2001)	Canada (2003)	France (2007)
Between-firm inequality	39	19	53	14
Within-firm inequality	37	65	39	63
Worker observables	13	16	7	27
Cov observables and firm fixed effects	11	1	1	2

Notes: This table collects results of decomposition (B-1) from several studies.

Sources: Brazil, Helpman et al. (2014); Sweden, Akerman et al. (2013); Canada, this paper; France, Tito (2015). Figures may not sum to 100 due to rounding error.

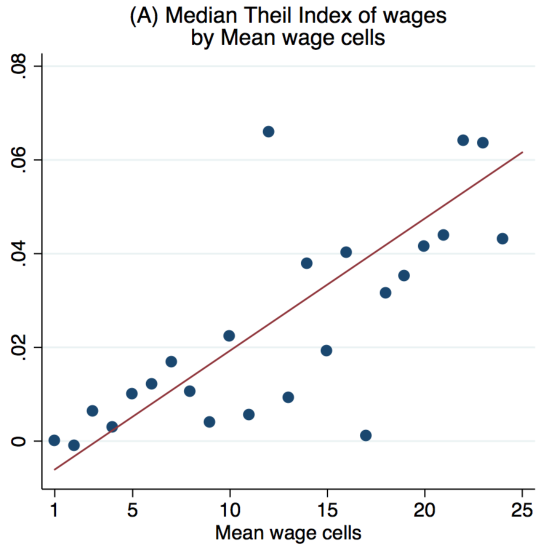
Akerman et al. (2013), Helpman et al. (2014) and Tito (2015) also report significant changes over time in within-firm inequality. While sharp increases in within-firm inequality are observed in Sweden and France, the opposite is true in Brazil. I do not emphasize these findings because, unfortunately, they do not provide evidence either in favor or against the main result of this paper. The theory predicts that within-firm inequality should, ceteris paribus, increase in response to reductions in variable trade costs.[78] The 'all else constant' condition, however, cannot be expected to hold in either France, Sweden or Brazil during the period analyzed in the corresponding study. On the other hand, in any given equilibrium (i.e. at any point in time), the theory has clear-cut predictions regarding the cross-sectional variation in firm-level wage distributions. These predictions play a crucial role in the mechanism that links trade liberalization and inequality in this paper. For these reasons, I motivate the theory by reporting cross-sectional findings from the WES and by citing cross-sectional rather than time series evidence in the literature.

B.4 Heterogeneity in Firm-level Wage Distributions

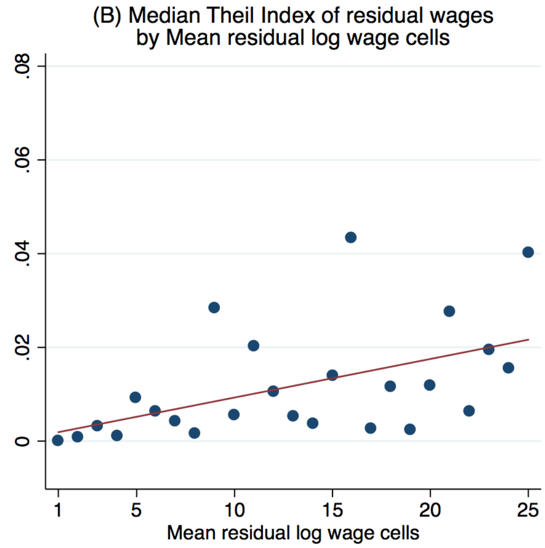
B.4.1 Alternative Inequality Measures

The following figure shows that the empirical patterns found in Figure 1 hold under two alternative inequality measures: the Theil Index (Panels A and B) and the mean log deviation (Panels C and D).

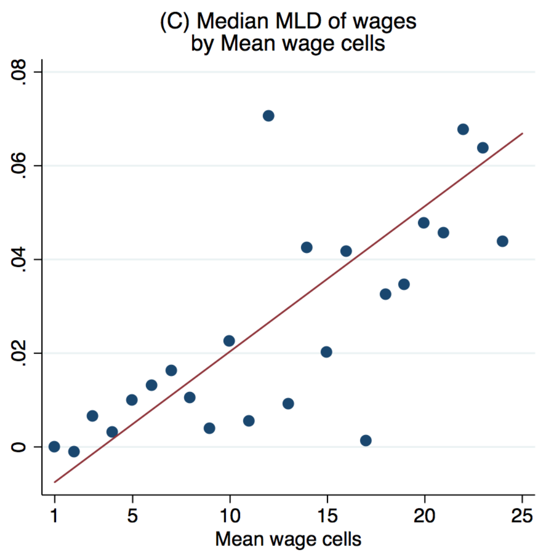
[78] If, in the model, trade liberalization triggered labor reallocations towards low-productivity, low-inequality firms, then within-firm inequality would decrease. Interestingly, studying trade liberalization in Brazil during 1986-2001, Menezes-Filho and Muendler (2011) find that exporters hire relatively fewer workers than the average employer. This suggests that a modification of the baseline model that enabled this alternative pattern of labor reallocations would rationalize a negative causal effect of trade liberalization on within-firm wage inequality.



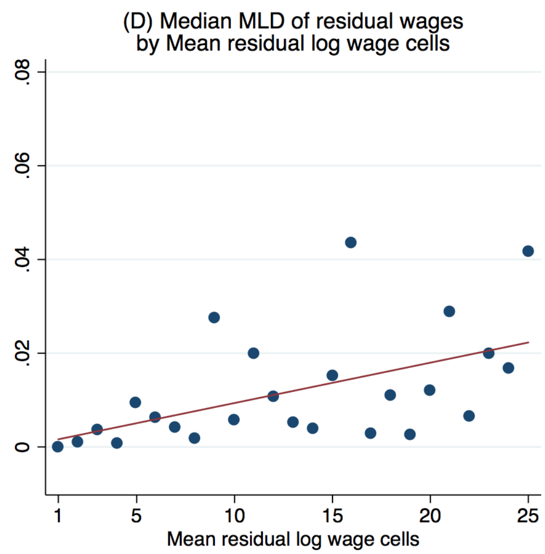
The horizontal axis sorts firms by increasing firm-level mean wage into 25 equally populated cells. The blue markers measure the difference between the median firm-level Theil index of wages in each cell and the median firm-level Theil index of wages in cell 1. Figure omits marker for cell 25 (0.11). The slope of the regression line is .002 (robust s.e. is .0005).



The horizontal axis sorts firms by increasing firm-level mean residual log wage -i.e. the fixed effect in eq (2a)- into 25 equally populated cells. Using the fixed effects from eq (2b), the blue markers measure the difference between the median firm-level Theil index of residual wages in each cell and the median firm-level Theil index of residual wages in cell 1. The slope of the regression line is .0008 (robust s.e. is .0003).



The horizontal axis sorts firms by increasing firm-level mean wage into 25 equally populated cells. The blue markers measure the difference between the median firm-level MLD of wages in each cell and the median firm-level MLD of wages in cell 1. Figure omits marker for cell 25 (0.13). The slope of the regression line is .003 (robust s.e. is .0007).



The horizontal axis sorts firms by increasing firm-level mean residual log wage -i.e. the fixed effect in eq (2a)- into 25 equally populated cells. Using the fixed effects from eq (2b), the blue markers measure the difference between the median firm-level MLD of residual wages in each cell and the median firm-level MLD of residual wages in cell 1. The slope of the regression line is .0009 (robust s.e. is .0003).

Figure B-1: Heterogeneity in Firm-Level Wage Distributions - Alternative Measures

B.5 Additional Results

B.5.1 Performance Pay and Firm Size

The following table shows that the findings reported in Table III hold when an alternative proxy for firm size, total employment, is used.

Table B-IV - Performance Pay Across Firms (Firm Size Proxy: Employment)

Outcome	$P[pp_{i\theta} > 0 \vert \cdot]$			$E[\ln pp_{i\theta} \vert \cdot]$			$Var[\ln pp_{i\theta} \vert \cdot]$		
Dep. Var.	$I_{[pp_{i\theta}>0]}$			$\ln pp_{i\theta}$			Sq. residuals from (4) to (6)		
	Basic	Add Exp Status	Add Controls	Basic	Add Exp Status	Add Controls	Basic	Add Exp Status	Add Controls
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Firm size	0.023	0.026	0.051	0.099	0.107	0.121	0.200	0.208	0.192
	(0.012)[c]	(0.013)[b]	(0.011)[a]	(0.071)	(0.072)	(0.041)[a]	(0.040)[a]	(0.043)[a]	(0.039)[a]
Exporter		-0.024	-0.000		-0.071	-0.041		-0.069	-0.071
		(0.038)	(0.036)		(0.147)	(0.133)		(0.117)	(0.118)
Controls	No	No	Yes	No	No	Yes	No	No	Yes
R-sq	0.00	0.00	0.12	0.01	0.00	0.35	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 employment. 'Exporter' is a dummy equal 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] and [b] denote statistical significance at the 1% and 5% levels, respectively.

B.5.2 Wages, Performance-independent Compensation and Firm Size

The following table repeats the analysis in Table III for two additional outcome variables: raw, unadjusted wages ($w_{i\theta}$) in Panel A and performance-independent compensation ($fixed_{i\theta}$) in Panel B. Letting $Y_{i\theta} \in \{\ln w_{i\theta}, \ln fixed_{i\theta}\}$, for worker i employed in firm θ, Table B-IV reports estimates of linear approximations to the mean and the variance of the conditional distribution of $Y_{i\theta}$ for workers with identical observable skills:

$$E[\ln Y_{i\theta} | Size_\theta, Ex_\theta, e_i] \approx \delta_B Size_\theta + \zeta_B Ex_\theta + \phi_B e_i, \tag{B-2a}$$

$$Var[\ln Y_{i\theta} | Size_\theta, Ex_\theta, e_i] \approx \delta_C Size_\theta + \zeta_C Ex_\theta + \phi_C e_i, \tag{B-2b}$$

where $Size_\theta$ 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 defined in section B-1. Since all workers in the sample report strictly positive wages and only very few report zero performance-independent compensation, the analysis disregards one of the three outcomes in Table III, $P[Y_{i\theta} > 0 | Size_\theta, Ex_\theta, e_i]$.

The results in Table B-IV indicate that firm size is positively correlated with all outcomes studied. The correlation between firm size and the dispersion of performance-independent compensation, however, is weaker and significant only at the 10% level. As in Table III, conditional on firm size, export status is uncorrelated with all outcomes.

Table B-V - Firm Size and Wages

Outcome	$E[Y_{i\theta}\|\cdot]$			$Var[Y_{i\theta}\|\cdot]$		
Dep. Var.	$Y_{i\theta}$			Sq. residuals from (1) to (3)		
	Basic	Add Exp Status	Add Controls	Basic	Add Exp Status	Add Controls
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A - $Y_{i\theta} = \ln w_{i\theta}$ (log wage)						
Firm size	0.100	0.099	0.073	0.008	0.008	0.010
	$(0.011)^a$	$(0.010)^a$	$(0.007)^a$	$(0.002)^a$	$(0.002)^a$	$(0.002)^a$
Exporter		0.015	-0.021		-0.008	-0.002
		(0.038)	(0.027)		(0.006)	(0.005)
Controls	No	No	Yes	No	No	Yes
R-sq	0.08	0.08	0.54	0.01	0.01	0.07
Obs	14,265	14,265	14,265	14,265	14,265	14,265
PANEL B - $Y_{i\theta} = \ln fixed_{i\theta}$ (log performance-indepependent compensation)						
Firm size	0.087	0.083	0.051	0.064	0.079	0.072
	$(0.011)^a$	$(0.011)^a$	$(0.007)^a$	$(0.039)^c$	$(0.048)^c$	$(0.041)^c$
Exporter		0.045	-0.005		-0.171	-0.160
		(0.038)	(0.026)		(0.108)	(0.107)
Controls	No	No	Yes	No	No	Yes
R-sq	0.05	0.05	0.40	0.00	0.00	0.01
Obs	14,185	14,185	14,185	14,185	14,185	14,185

Notes: this table reports OLS estimates of the right-hand side parameters of (B-2a) in columns 1 to 3 and (B-2b) in columns 4 to 6. 'Firm size' is the natural log of firm total employment. '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} and ^{c} denote statistical significance at the 1% and 10% levels, respectively.