Not for Publication:
Appendix to Performance Pay, Trade and Inequality

Germán P. Pupato
Ryerson University

September 2017

This Appendix contains the supplementary materials to *Performance Pay, Trade and Inequality*. Section A contains the proofs of all propositions and lemmas, and develops additional theoretical results. Sections B presents a detailed description of the data and additional empirical results. Section C provides an extended discussion of related research.

A Technical Appendix

A.1 A Microfoundation for the Production Process

I present a microfoundation for the continuous-time production process introduced in section 3.2. I show that this setting can be interpreted as the limit of a sequential production process with a discrete number of tasks, when task duration approaches zero. In particular, Lemma A-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, as in equation (1) of the paper.

Suppose that, in every firm, the production process requires each worker to perform a sequence of T symmetric tasks, indexed by $\iota = 1, ..., T$. Each task spans an interval of time of length $\Delta \equiv 1/T$. Worker i chooses a sequence $\left\{\epsilon_{i\iota}^{\Delta}\right\}_{\iota=1}^{T}$ of possibly history-dependent effort levels for each task ι, where $\epsilon_{i\iota}^{\Delta} \in [\epsilon_L, \epsilon_H] \subset \Re_{++}$.[48] This choice generates a stochastic sequence of worker-specific performance outcomes $\{z_{i\iota}\}_{\iota=1}^{T}$, where $z_{i\iota}$ is equal to 1 if worker i successfully completes task ι and equal to -1 in the event of a mistake, for $\iota = 1, ..., T$. For a fixed Δ, the probability of success in any task ι, denoted $\pi_{i\iota}^{\Delta}$, is given by

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

where $\mu(\cdot)$ is continuous and, since $\epsilon_{i\iota}^{\Delta} \in [\epsilon_L, \epsilon_H]$, bounded.[49] The expected performance of worker i in task ι is $E\left(z_{i\iota} | z_{i1}, ..., z_{i\iota-1}\right) = \mu(\epsilon_{i\iota}^{\Delta})\Delta^{1/2}$ and thus it is also natural to assume that $\mu(\cdot)$ is increasing. Note that $z_{i\iota}$ is independent of firm productivity and, conditional on effort, independent of $z_{i'\iota'}$ for any two tasks ι and ι' and any two workers i and i' (unless, of course, $\iota = \iota'$ and $i = i'$).

Let $Z_{i\iota}^{\Delta}$ denote the normalized cumulative performance of worker i up to task ι, $Z_{i\iota}^{\Delta} \equiv \Delta^{1/2}\sum_{\iota'=1}^{\iota} z_{i\iota'}$. Equivalently, $-Z_{i\iota}^{\Delta}$ is worker i's normalized number of mistakes in excess of successes up to task ι; 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 $\left\{Z_{i\iota}^{\Delta}\right\}_{\iota=1}^{T}$ in continuous time

[48] $\mu_{\min}$ 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.

[49] 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 A-1.

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

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

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 $\left\{\epsilon_{i\iota}^{\Delta}\right\}_{\iota=1}^{T}$. Endowing $C[0,1]$ with the uniform metric, I obtain the following result.

Lemma A-1 *In the sequential production process with task duration $\Delta = 1/T$, consider a sequence of effort $\left\{\epsilon_{i\iota}^{\Delta}\right\}_{\iota=1}^{T}$ and the corresponding process of cumulative performance $\left\{Z_{i\iota}^{\Delta}\right\}_{\iota=1}^{T}$ 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 $\left\{Z_{i\iota}^{\Delta}\right\}_{\iota=1}^{T}$ 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.[50] In particular, the proof of Lemma A-1 relies on a result due to Brown (1971). The assumptions of Bernoulli task outcomes and unidimensional effort are not essential.[51]

A.1.1 Proof of Lemma A-1

For any worker i and task duration $\Delta \equiv 1/T$, let $\left\{\mathcal{F}_{i\iota'}^{\Delta}\right\}_{\iota'=0}^{T}$ be a sequence of σ-fields on the underlying probability space, where $\mathcal{F}_{i0}^{\Delta}$ is the trivial σ-field and $\mathcal{F}_{i1}^{\Delta}$,...,$\mathcal{F}_{iT}^{\Delta}$ is the filtration generated by the random variables z_{i1},...,z_{iT}. Let $Z_{i\iota}^{\Delta}$ denote the

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

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

(normalized) cumulative performance of worker i up to task $\iota \in \{1, ..., T\}$ with initial condition $z_{i0} = 0$; i.e., $Z_{i\iota}^{\Delta} \equiv \Delta^{1/2} \sum_{\iota'=1}^{\iota} z_{i\iota'}$.[52] Let $\overline{z_{i\iota}} \equiv E\left(z_{i\iota} | \mathcal{F}_{i\iota-1}^{\Delta}\right) = \mu(\epsilon_{i\iota}^{\Delta})\Delta^{1/2}$ for $\iota \in \{1, ..., T\}$, where effort is allowed to be history-dependent; i.e., $\epsilon_{i\iota}^{\Delta} = \epsilon_{i\iota}^{\Delta}(z_{i1}, ..., z_{i\iota-1})$. Then, for $\iota \in \{0, ..., T\}$,

$$Z_{i\iota}^{\Delta} \equiv \Delta \sum_{\iota'=1}^{\iota} \mu(\epsilon_{i\iota'}^{\Delta}) + B_{i\iota}^{\Delta}, \tag{A-2}$$

where

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

denotes the cumulative deviation of worker i's performance from its expected value up to task ι. Equation (A-2) is an identity, by definitions of $B_{i\iota}^{\Delta}$ and $Z_{i\iota}^{\Delta}$. 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 (A-1). Then, using equation (A-2),

$$\begin{aligned} Z^{\Delta}(t) &= \left(1 - \frac{t}{\Delta} + \left\lfloor \frac{t}{\Delta} \right\rfloor\right) \left[\Delta \sum_{\iota'=1}^{\lfloor t/\Delta \rfloor} \mu(\epsilon_{\iota'}^{\Delta}) + B_{\lfloor t/\Delta \rfloor}^{\Delta}\right] + \\ &+ \left(\frac{t}{\Delta} - \left\lfloor \frac{t}{\Delta} \right\rfloor\right) \left[\Delta \sum_{\iota'=1}^{\lfloor t/\Delta \rfloor + 1} \mu(\epsilon_{\iota'}^{\Delta}) + B_{\lfloor t/\Delta \rfloor + 1}^{\Delta}\right], \\ &= \Delta \sum_{\iota'=1}^{\lfloor t/\Delta \rfloor} \mu(\epsilon_{\iota'}^{\Delta}) + B^{\Delta}(t) + \left(\frac{t}{\Delta} - \left\lfloor \frac{t}{\Delta} \right\rfloor\right) \Delta b(\epsilon_{\lfloor t/\Delta \rfloor + 1}^{\Delta}), \\ &= \Delta \sum_{\iota'=1}^{\lfloor t/\Delta \rfloor} \mu(\epsilon_{\iota'}^{\Delta}) + B^{\Delta}(t) + o(\Delta), \\ &= \int_{\Delta}^{\lfloor t/\Delta \rfloor \Delta} \mu(\epsilon^{\Delta}(t'))dt' + B^{\Delta}(t) + o(\Delta). \end{aligned} \tag{A-3}$$

In the third line, $B^{\Delta}(t)$ is the piecewise linear interpolation between the points $(0, 0)$, $\left(\Delta, B_1^{\Delta}\right)$, $\left(2\Delta, B_2^{\Delta}\right)$,...,$\left(1, B_T^{\Delta}\right)$. More specifically,[53]

$$B^{\Delta}(t) = \left(1 - \frac{t}{\Delta} + \left\lfloor \frac{t}{\Delta} \right\rfloor\right) B_{\lfloor t/\Delta \rfloor}^{\Delta} + \left(\frac{t}{\Delta} - \left\lfloor \frac{t}{\Delta} \right\rfloor\right) B_{\lfloor t/\Delta \rfloor + 1}^{\Delta}, \tag{A-4}$$

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

[53]This interpolation is equivalent to that employed in Brown (1971). Unlike the definition in equation (A-4), 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, $E\left[\left(\varepsilon_{\iota}^{\Delta}\right)^2\right] = \Delta\iota$ for all ι, by equation (A-9). It is then straightforward to show that the interpolation of $\left\{\varepsilon_{\iota}^{\Delta}\right\}_{\iota=0}^{T}$ following Brown's procedure is equivalent to equation (A-4).

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 $(t/\Delta - \lfloor t/\Delta \rfloor) \in [0,1]$.[54] Finally, the fifth line introduces a change of integration variables, $t' = \Delta\iota'$, 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-3), 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

$$\int_{\Delta}^{\lfloor t/\Delta \rfloor \Delta} \mu(\epsilon^{\Delta}(t'))dt' \to \int_0^t \mu(\epsilon(t'))dt' \text{ a.s. as } \Delta \to 0, \tag{A-5}$$

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-3), I start by noting two properties of B^{Δ}_{ι}. First, for all $s, \iota \in \{0, ..., T\}$ such that $s < \iota$, applying the law of iterated expectations yields

$$\begin{aligned} E\left(B^{\Delta}_{\iota} | \mathcal{F}^{\Delta}_s\right) &= \Delta^{1/2} \sum\nolimits_{\iota'=1}^{\iota} E\left(z_{\iota'} - \overline{z_{\iota'}} | \mathcal{F}^{\Delta}_s\right) \\ &= \Delta^{1/2} \sum\nolimits_{\iota'=1}^{s} \left(z_{\iota'} - \overline{z_{\iota'}}\right) + \Delta^{1/2} \sum\nolimits_{\iota'=s+1}^{\iota} \left[E\left(z_{\iota'} | \mathcal{F}^{\Delta}_s\right) - E\left(z_{\iota'} | \mathcal{F}^{\Delta}_s\right)\right] \\ &= B^{\Delta}_s. \end{aligned} \tag{A-6}$$

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

$$\sum_{\iota=1}^{T} E\left[\left(B^{\Delta}_{\iota} - B^{\Delta}_{\iota-1}\right)^2 | \mathcal{F}^{\Delta}_{t-1}\right] = E\left[\sum_{\iota=1}^{T} E\left[\left(B^{\Delta}_{\iota} - B^{\Delta}_{\iota-1}\right)^2 | \mathcal{F}^{\Delta}_{t-1}\right]\right] \text{ a.s., for all } \Delta. \tag{A-7}$$

Conditions (A-6) and (A-7) imply that, for given Δ, the process $\left\{B^{\Delta}_{\iota}\right\}_{\iota=0}^{T}$ belongs to the class of zero-mean, square-integrable martingales relative to $\left\{\mathcal{F}^{\Delta}_{\iota}\right\}_{\iota=0}^{T}$ studied in Brown (1971).[55] In particular, Theorem 3 in Brown (1971) implies that,

[54] 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.

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

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

as $\Delta \to 0$, which is implied by (A-7).

as $\Delta \to 0$, the sequence of probability measures determined by the distribution of $\left\{B^{\Delta}(t); 0 \leq t \leq 1\right\}$ converges weakly to the Wiener measure in the space $C[0,1]$ with the uniform norm, provided that the Lindeberg condition holds, namely

$$E\left[\left(B_T^{\Delta}\right)^2\right]^{-2} \sum_{\iota=1}^{T} E\left[\Delta\left(z_\iota - \overline{z_\iota}\right)^2 I\left(\Delta^{1/2}\left|z_\iota - \overline{z_\iota}\right| \geq \delta E\left[\left(B_T^{\Delta}\right)^2\right]\right)\right] \to_p 0, \qquad \text{(A-8)}$$

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

To check (A-8), first note that $E\left[\left(B_T^{\Delta}\right)^2\right] = 1$ for all Δ, since martingale increments are uncorrelated[56] and thus

$$\begin{aligned} E\left[\left(B_\iota^{\Delta}\right)^2\right] &= \Delta E\left[\left(\sum\nolimits_{\iota'=1}^{\iota}\left(z_{\iota'} - \overline{z_{\iota'}}\right)\right)^2\right], \\ &= \Delta E \sum\nolimits_{\iota'=1}^{\iota}\left(z_{\iota'} - \overline{z_{\iota'}}\right)^2, \\ &= \Delta E \sum_{\iota'=1}^{\iota} E\left[\left(z_{\iota'} - \overline{z_{\iota'}}\right)^2 | \mathcal{F}_{\iota'-1}^{\Delta}\right], \\ &= \Delta E \sum_{\iota'=1}^{\iota}\left[1 + \left(\overline{z_{\iota'}}\right)^2 - \overline{z_{\iota'}} E\left[z_{\iota'} | \mathcal{F}_{\iota'-1}^{\Delta}\right]\right], \\ &= \Delta \iota. \end{aligned} \qquad \text{(A-9)}$$

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

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

[56] That is, for $\iota_1 > \iota_0 \geq 0$, let $s_{10} = E\left[\left(z_{\iota_0} - \overline{z_{\iota_0}}\right)\left(z_{\iota_1} - \overline{z_{\iota_1}}\right)\right]$. Then, applying the law of iterated expectations twice yields

$$\begin{aligned} s_{10} &= E\left[\left(z_{\iota_0} - \overline{z_{\iota_0}}\right) E\left[\left(z_{\iota_1} - \overline{z_{\iota_1}}\right) | \mathcal{F}_{\iota_0}^{\Delta}\right]\right], \\ &= E\left[\left(z_{\iota_0} - \overline{z_{\iota_0}}\right) E\left[\left(z_{\iota_1} - z_{\iota_1}\right) | \mathcal{F}_{\iota_0}^{\Delta}\right]\right], \\ &= 0. \end{aligned}$$

Letting K^{Δ} denote the left-hand side of (A-8) yields

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

Therefore, $K^{\Delta} \to 0$ a.s. as $\Delta \to 0$, which is sufficient to verify the Lindeberg condition (A-8). Therefore, by Theorem 3 in Brown (1971),

$$B^{\Delta}(t) \to^{d} B(t), \tag{A-10}$$

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-5) and (A-10) to equation (A-3), I conclude that, as $\Delta \to 0$, $Z^{\Delta}(t)$ 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

$$\log w_i(Z_i^1) = \log \chi_i + \int_0^1 \frac{k'(\epsilon_i(t))}{\mu'(\epsilon_i(t))}\left[dZ_i(t) - \mu(\epsilon_i(t))dt\right] - \frac{1}{2}\int_0^1 \left[\frac{k'(\epsilon_i(t))}{\mu'(\epsilon_i(t))}\right]^2 dt. \tag{A-11}$$

I start by introducing a change of variables, letting $s_i(Z_i^1) \equiv \log w_i\left(Z_i^1\right)$, to re-write the problem of worker i -constraint (iii) in Problem (6)- as

$$\max_{\epsilon_i} E\left[\exp\left(s_i(Z_i^1) - \int_0^1 k(\epsilon_i(t'))dt'\right)\right] \tag{A-12}$$

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

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-12). 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-12) with certain equivalent χ. Let

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

where

$$K(\iota;\epsilon') \equiv \exp\left(\int_0^\iota [k(\epsilon(t)) - k(\epsilon'(t))]\,dt\right).$$

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

$$\begin{aligned} V(\iota;\epsilon') &\equiv \max_m F(\iota;\epsilon';\mathbf{m}) \\ &= \max_m F(\iota;\epsilon;\mathbf{m})K(\iota;\epsilon') = V(\iota;\epsilon)\,K(\iota;\epsilon'). \end{aligned} \tag{A-13}$$

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

$$dV(\iota;\epsilon) = \gamma(\iota)d\left[Z(\iota) - \int_0^\iota \mu(\epsilon(t))\,dt\right]. \tag{A-14}$$

For any effort sequence ϵ', $dZ = \mu(\epsilon') + dB$; thus $dV(\iota;\epsilon) = \gamma[\mu(\epsilon') - \mu(\epsilon)]\,dt + \gamma dB$. Together with (A-13), this implies

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

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

$$\gamma(t)\,\mu'(\epsilon(t)) = k'(\epsilon(t))\,V(t;\epsilon), \tag{A-15}$$

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

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

$$\chi(t) = \frac{V(t;\epsilon)}{E\left[\exp\left(-\int_0^1 k(\epsilon(t'))dt'\right)|\mathcal{F}_t\right]}. \tag{A-16}$$

The derivative of the denominator in (A-16) with respect to t is zero. Thus,

$$\begin{aligned}\frac{d\chi(t)}{\chi(t)} &= \frac{dV(t;\epsilon)}{V(t;\epsilon)} \\ &= \frac{\gamma(t)}{V(t;\epsilon)}d\left[Z(t) - \int_0^t \mu(\epsilon(t'))dt'\right] \\ &= \frac{k'(\epsilon(t))}{\mu'(\epsilon(t))}d\left[Z(t) - \int_0^t \mu(\epsilon(t'))dt'\right],\end{aligned} \tag{A-17}$$

where the second and third lines follow from (A-14) and (A-15), respectively. Using Ito's Lemma for the function $\log(\cdot)$ yields

$$\begin{aligned}d\log\chi(t) &= \frac{d\chi(t)}{\chi(t)} - \frac{1}{2}\frac{1}{[\chi(t)]^2}[d\chi(t)]^2 \\ &= \frac{d\chi(t)}{\chi(t)} - \frac{1}{2}\frac{1}{[\chi(t)]^2}\left[\chi(t)\frac{k'(\epsilon(t))}{\mu'(\epsilon(t))}\right]^2 dt \\ &= \frac{d\chi(t)}{\chi(t)} - \frac{1}{2}\left[\frac{k'(\epsilon(t))}{\mu'(\epsilon(t))}\right]^2 dt.\end{aligned} \tag{A-18}$$

Using (A-17) to substitute for $d\chi(t)/\chi(t)$, integrating (A-18) and letting $\chi(0) = \chi$ yields

$$\log\chi(1) = \log\chi + \int_0^1 \frac{k'(\epsilon(t))}{\mu'(\epsilon(t))}[dZ(t) - \mu(\epsilon(t))dt] - \frac{1}{2}\int_0^1\left[\frac{k'(\epsilon(t))}{\mu'(\epsilon(t))}\right]^2 dt. \tag{A-19}$$

By the construction of $\chi(t)$, $\chi(1) = \exp[s(Z^1)]$; thus, $\chi(1) = w(Z^1)$. Substituting the latter in (A-19) delivers (A-11).

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 (6) are replaced with the contract representations obtained in step 1. Importantly, the solution to this problem is not necessarily implementable by the contracts (A-11), 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-11) in the objective function of Problem (6) and drop the incentive compatibility constraints:

$$\min_{\{\chi_i,\epsilon_i;i\in[0,h]\}} \int_0^h \chi_i E\left[\exp\left(\int_0^1 \frac{k'(\epsilon_i(t))}{\mu'(\epsilon_i(t))}\left[dZ_i(t)-\mu(\epsilon_i(t))dt\right] - \frac{1}{2}\int_0^1 \left[\frac{k'(\epsilon_i(t))}{\mu'(\epsilon_i(t))}\right]^2 dt\right)\right] di$$

$$\begin{aligned} s.t\quad & (i)\ n_0 \geq -h^{-1}\textstyle\int_0^h E\left[Z_i(1)\right] di \\ & (ii)\ Z_i(t) = \textstyle\int_0^t \mu\left(\epsilon_i\left(t'\right)\right) dt' + B_i(t), \text{ for } i\in[0,h] \\ & (iii)\ E\left[U(w_i,\epsilon_i)\right] \geq \overline{u}, \text{ for } i\in[0,h] \end{aligned} \qquad \text{(A-20)}$$

The following steps simplify this problem. First, substitute $E\left[Z_i(1)\right] = \int_0^1 E\left[\mu\left(\epsilon_i\left(t\right)\right)\right] dt$ 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 ϵ_i.[57] Third, individual rationality constraints should bind at the optimum.[58] Expression (iii) can then be used to solve for χ_i as a function of ϵ_i,

$$\chi_i = \frac{\overline{u}}{E\left[\exp\left(-\int_0^1 k(\epsilon_i)dt\right)\right]}. \qquad \text{(A-21)}$$

[57] 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 $\exp\left(\int_0^\iota \frac{k'(\epsilon_i(t))}{\mu'(\epsilon_i(t))} dB_i(t) - \frac{1}{2}\int_0^\iota \left[\frac{k'(\epsilon_i(t))}{\mu'(\epsilon_i(t))}\right]^2 dt\right)$ is an $\mathcal{F}_\iota$-martingale with expected value equal to one, for all ι and ϵ. See Karatzas and Shreve (1988), p.200.

[58] 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.

Problem (A-20) is then simplified to a problem of finding the optimal ϵ_i that minimizes the certainty equivalent χ_i for $i \in [0, h]$, subject to a single performance constraint:

$$\min_{\{\epsilon_i; i \in [0,h]\}} \int_0^h \frac{\overline{u}}{E\left[\exp\left(-\int_0^1 k(\epsilon_i)dt\right)\right]} di, \tag{A-22}$$

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

Note that both the objective and constraint of problem (A-22) 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-22) can be restricted to the set of deterministic (history-independent) sequences.

Dropping the expectations operator and disregarding the (positive) constant $\overline{u}$, (A-22) can be further simplified into

$$\min_{\{\epsilon_i; i \in [0,h]\}} \int_0^h \exp\left[\int_0^1 k(\epsilon_i(t))dt\right] di \quad s.t \quad n_0 \geq -h^{-1} \int_0^h \int_0^1 \mu\left(\epsilon_i\left(t\right)\right) dtdi. \tag{A-23}$$

It is convenient to analyze this problem by introducing a set of auxiliary choice variables $\{a_i \in (-\mu(\epsilon_H), -\mu(\epsilon_L)]; i \in [0, h]\}$, 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-23) sequentially with the following two-step procedure:

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

$$\min_{\epsilon_i} \int_0^1 k(\epsilon_i(t))dt \quad s.t \quad a_i \geq \int_0^1 -\mu\left(\epsilon_i\left(t\right)\right) dt. \tag{A-24}$$

Under the assumptions that $k(\cdot)$ and $\mu(\cdot)$ are convex and strictly concave, respectively, it is straightforward to verify that the solution to (A-24) is a unique constant effort for all $t \in [0, 1]$, denoted $\epsilon(a_i)$, that satisfies $a_i = -\mu(\epsilon(a_i))$. In addition, $a_i \in (-\mu(\epsilon_H), -\mu(\epsilon_L)]$ implies $\epsilon(a_i) \in (\epsilon_L, \epsilon_H]$.

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

$$\min_{\substack{\{a_i \in [-\mu(\epsilon_H), -\mu(\epsilon_L)];\\ i \in [0,h]\}}} \int_0^h \exp\left[k(\epsilon(a_i))\right] di \quad s.t \quad n_0 = h^{-1} \int_0^h a_i di. \tag{A-25}$$

Since $\exp[k(\cdot)]$ is strictly convex, the solution to (A-25) 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-20), 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-20) is implementable. That is, I show that if worker i is assigned contract (A-11) 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.[59]

Evaluating the wage representation (A-11) at a constant effort ϵ^* and using the expression for the certainty equivalent (A-21) at $t = 1$, yields[60]

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

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

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

The worker's problem becomes,

$$\max_{\epsilon} E \left[\exp\left(\alpha^* + \beta^* Z_i(1)\right) - \int_0^1 k(\epsilon(t)) dt \right] \quad s.t \;\; Z_i(1) = \int_0^1 \mu(\epsilon(t)) dt + B_i(1), \tag{A-26}$$

where ϵ is an adapted process. Substituting the constraint in the objective function

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

[60]From (A-21), the certainty equivalent for a constant effort ϵ^* is $\chi_i = \overline{u}/E\left[\exp\left(-\int_0^1 k(\epsilon_i) dt\right)\right] = \overline{u}/\exp(-k(\epsilon^*))$. Therefore, $\log \chi_i = \log \overline{u} + k(\epsilon^*)$.

and rearranging yields,

$$E\left[\exp\left(\alpha^* + \beta^* Z_i(1)\right) - \int_0^1 k(\epsilon(t))dt\right]$$
$$= \exp(\alpha^*)E\left[\exp\left(\int_0^1 \left[\beta^*\mu(\epsilon(t)) - k(\epsilon(t))\right]dt + \beta^* B_i(1)\right)\right].$$

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-26). 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 $\log w_i = \alpha^* + \beta^* Z_i(1)$.

A.3 Proof of Corollary 1

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

$$w_i^* = \overline{u}e^{k(\epsilon)+\beta(\epsilon)\left[B_i(1)-\frac{1}{2}\beta(\epsilon)\right]}, \tag{A-27}$$

for all $i \in [0,h]$, where $\beta(\epsilon) = k'(\epsilon)/\mu'(\epsilon)$ and $\epsilon_L < \epsilon \leq \epsilon_H$.[61] 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.

[61] 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 $\epsilon_L < \epsilon \leq \epsilon_H$.

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 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-27) and part (a) of Corollary 1. Normalized wages are log-normally distributed; in particular,

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

Observe that: (a) $E\left[\log\left(\widehat{w}_{ij}(x)\right)|\epsilon_j\right]$ is strictly decreasing in ϵ_j; (b) $Var\left[\log\left(\widehat{w}_{ij}(x)\right)|\epsilon_j\right]$ is strictly increasing in ϵ_j; (c) $E\left[\log\left(\widehat{w}_{ij}(x)\right)|\epsilon_j\right] = -Var\left[\log\left(\widehat{w}_{ij}(x)\right)|\epsilon_j\right]/2$, 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 Alternative Moral Hazard Settings.

The considerable complexity of the contracting problem studied in this section warrants a brief discussion of alternative modelling strategies. The dynamic formulation adopted here delivers two properties that will play a key role in the analysis: (i) tractable optimal contracts that can be embedded in general equilibrium; (ii) a positive correlation between the first and second moments of firm-level wage distributions -Corollary 1- that is consistent with the empirical pattern documented in section 7 and Figure B-1. Next, I argue that standard (static) formulations of the moral hazard problem fail to satisfy at least one of these two criteria.[62]

[62] I restrict the discussion to static moral hazard settings because alternative dynamic formulations are not necessarily simpler than the formulation adopted here (see chapter 10 in Bolton and Dewatripont (2005)).

Consider a class of static moral hazard settings in which workers have additively separable utility in income and effort. For each employee, the firm observes an outcome that is a stochastic function of a single individual effort choice. Conditional on effort, the distribution of individual outcomes is firm-specific and i.i.d. across employees.[63] The intractability of optimal contracts in higher-dimensional formulations (multiple outcomes and effort levels) of this bilateral contracting environment is a well-known problem.[64] Sufficient tractability is attained when the observable outcome takes only two values (for example, section 5.1.2 in Laffont and Martimort (2002)). This special case, however, delivers Bernoulli firm-level wage distributions that fail to generate a positive correlation between first and second moments.[65] Yet another popular approach is to assume normally distributed outcomes, exponential utility, monetary cost of effort and linear contracts. This setting generates tractable contracts in which both the expected wage and variance of wages increase in effort. However, linear contracts are suboptimal; indeed, in this specific environment, the first-best effort can be implemented by contracts that generate negligible wage dispersion (Bolton and Dewatripont (2005), page 139).

[63]In particular, the dispersion of outcomes is allowed to vary arbitrarily across firms. The outcome can be interpreted as either an intermediate input in the quality production function or as the quality of the worker's output.

[64]See, for example, Holmström and Milgrom (1987) and sections 4.4 and 4.5 in Bolton and Dewatripont (2005)). The monotonicity of optimal compensation with respect to performance can be guaranteed by imposing a monotone likelihood ratio property on the distribution of performance that is conditional on effort (Bolton and Dewatripont (2005), page 147). Therefore, it is possible to conclude that high-effort workers will receive, on average, higher wages. However, there are no general results characterizing the dispersion of wages induced by the optimal contract.

[65]To see this, suppose that the probability of the superior outcome increases in effort. The expected wage is then monotonically increasing in effort while the variance of wages decreases when effort is sufficiently high. This result holds for any two points in the support of a Bernoulli distribution, so the support of the wage distribution can be allowed to vary arbitrarily across firms. Moreover, consider the standard case in which the worker is protected by a non-negative limited liability constraint (section 5.1.2 in Laffont and Martimort (2002)). The optimal contracts sets the wage equal to zero when low performance is realized. It is straightforward to verify that the coefficient of variation of firm-level wages is monotonically *decreasing* in effort. In this case, the two-outcome formulation implies that high-wage firms are low-inequality firms, which contradicts the evidence in section 7.

A.5 Profit Maximization

A.5.1 Proof of Lemma 1

From expression (7) in the main text, the optimal quality $q(\theta)$ minimizes the average cost of quality (per unit of output); that is,

$$\min_{q\in[q_L,q_H]} \frac{c^{\theta}\left(q\right)}{q} \equiv \frac{\omega(\epsilon(\theta,q))}{q\theta^s},$$

where $q_\ell = q(\theta, -\mu(\epsilon_\ell))$, $\ell = \{L, H\}$, $\omega(\epsilon) \equiv \overline{u}e^{k(\epsilon)}$ from Corollary 1 and $\epsilon(\theta, q)$ is implicitly defined by $q = q(\theta, -\mu(\epsilon))$. If it exists, the solution to this problem is independent of the variable trade cost. Moreover, if the solution is unique then firm θ offers products of identical quality in the domestic and foreign markets and hence $q(\theta)$ is also independent of the export status of the firm.

Next, I establish the existence and uniqueness of $q(\theta)$. Under Assumption 1, the function $q(\theta, -\mu(\epsilon))$ provides a one-to-one mapping between effort and quality, hence I can focus on solving:

$$\max_{\epsilon\in[\epsilon_L,\epsilon_H]} \frac{q(\theta,-\mu(\epsilon))\theta^s}{\omega(\epsilon)}.$$

The first-order condition yields a critical point ϵ^* such that $q_\epsilon(\theta, -\mu(\epsilon^*))/q(\theta, -\mu(\epsilon^*)) = k'(\epsilon^*)$, where q_ϵ denotes the partial derivative of $q(\theta, -\mu(\epsilon))$ with respect to ϵ. It is easy to check from the second-order condition that concavity of $q(\theta, -\mu(\epsilon))$ ensures that the objective function is strictly concave. Hence ϵ^* is the unique global maximizer of $f(\epsilon)$. Concavity of $q(\theta, -\mu(\epsilon))$ in ϵ and Assumption 1 imply that q_ϵ/q and k' are decreasing and increasing, respectively, in ϵ. Since $q_\epsilon/q \to \infty$ as $\epsilon \to 0$, q_ϵ/q and k' intersect exactly once at point $\epsilon^* > 0$. Moreover, $\epsilon^* \in [\epsilon_L, \epsilon_H]$ provided that ϵ_L is sufficiently small and ϵ_H is sufficiently large. This establishes the existence of a unique optimal quality $q(\theta) \equiv q(\theta, -\mu(\epsilon^*))$ that satisfies $q(\theta) \in [q_L, q_H]$.

A.5.2 Proof of Proposition 2

Consider a reformulation of the optimization problem in the previous section:

$$\max_{q\in[q(\theta,-\mu(\epsilon_L)),q(\theta,-\mu(\epsilon_H))]} f(\theta,q) \equiv \log\left(\frac{q}{c^{\theta}\left(q\right)}\right) = \log\left(\frac{q\theta^s}{\overline{u}}\right) - k(\epsilon(\theta,q)), \qquad \text{(A-28)}$$

where $\epsilon(\theta, q)$ is implicitly defined by $q = q(\theta, -\mu(\epsilon))$. It is now easy to verify the following hypotheses of the Monotone Maximum Theorem (Carter (2001), page 205): (i) by Lemma 1, the solution to problem (A-28) exists for all $\theta \in [\theta_L, \theta_H]$; (ii) $[\theta_L, \theta_H]$

and $[q_L, q_H]$ are totally ordered sets; (iii) $f(\theta, q)$ is trivially supermodular in q; (iv) the constraint correspondence $G(\theta) : [\theta_L, \theta_H] \rightrightarrows [q(\theta, -\mu(\epsilon_L)), q(\theta, -\mu(\epsilon_H))]$ is increasing in θ.

To meet the remaining requirement of the theorem, I show that $f(\theta, q)$ is supermodular in θ and q.[66] The latter is equivalent to showing that

$$\frac{\partial^2 f(\theta, q)}{\partial \theta \partial q} = -k''(\epsilon(\theta, q)) \frac{\partial \epsilon(\theta, q)}{\partial q} \frac{\partial \epsilon(\theta, q)}{\partial \theta} - k'(\epsilon(\theta, q)) \frac{\partial^2 \epsilon(\theta, q)}{\partial \theta \partial q} \geq 0.$$

Recall that $k'' > 0$, $k' > 0$ and $\mu'' < 0 < \mu'$ by Assumption 1. Applying the Implicit Function Theorem to $T(\theta, q, \epsilon) \equiv q(\theta, -\mu(\epsilon)) - q = 0$ yields

$$\begin{aligned} \frac{\partial \epsilon(\theta, q)}{\partial \theta} &= \left. \frac{q_\theta(\theta, -\mu(\epsilon))}{q_n(\theta, -\mu(\epsilon))\mu'(\epsilon)} \right|_{\epsilon=\epsilon(\theta,q)} < 0, \\ \frac{\partial \epsilon(\theta, q)}{\partial q} &= -\left. \frac{1}{q_n(\theta, -\mu(\epsilon))\mu'(\epsilon)} \right|_{\epsilon=\epsilon(\theta,q)} > 0, \end{aligned}$$

since $q_n < 0 < q_\theta$. Hence it suffices to show that $\partial^2 \epsilon(\theta, q)/\partial \theta \partial q \leq 0$. Indeed,

$$\frac{\partial^2 \epsilon(\theta, q)}{\partial \theta \partial q} = \underbrace{\left(\frac{\partial \epsilon}{\partial q}\right)^{-2}}_{(>0)} \left\{ \underbrace{\frac{\partial \epsilon}{\partial \theta}}_{(<0)} \underbrace{\left[q_n \mu'' - q_{nn} (\mu')^2\right]}_{(\geq 0)} + \underbrace{q_{n\theta} \mu'}_{(\leq 0)} \right\} \leq 0,$$

because $q_{n\theta} \leq 0$ (q is submodular in θ and n) and concavity of $q(\theta, -\mu(\epsilon))$ in ϵ implies $q_{\epsilon\epsilon} = q_{nn} (\mu')^2 - q_n \mu'' \leq 0$.

By the Monotone Maximum Theorem, I conclude that the (unique) solution to problem (A-28), $q(\theta)$, is increasing in θ. Since higher quality requires higher effort, optimal effort, $\epsilon(\theta)$, is also increasing in θ.

Provided that effort is defined over a sufficiently large interval, $q(\theta) \in (q_L, q_H)$. In this case, the Envelope Theorem implies

$$\frac{d}{d\theta} \log \left(\frac{q(\theta)}{c^\theta (q(\theta))} \right) = \left. \frac{\partial f(\theta, q)}{\partial \theta} \right|_{q=q(\theta)} = -k'(\epsilon(\theta)) \frac{\partial \epsilon(\theta, q(\theta))}{\partial \theta} > 0,$$

hence the average cost of quality per unit of output is decreasing in firm productivity. From the first-order condition (8), the change in optimal output, $y(\theta)$, associated to

[66] Note that $f(\theta, q)$ is supermodular in θ and q.if and only if $f(\theta, q)$ has increasing differences in θ and q, since θ and q are elements of totally ordered sets.

a change in firm productivity can be written as

$$d\log(y(\theta)) = \frac{1}{1-\rho} d\log\left(\frac{q(\theta)}{c^{\theta}\left(q(\theta)\right)}\right) - d\log(q(\theta)).$$

The two terms on the right-hand side operate in opposite directions. When ρ is sufficiently close to 1, however, the first term overwhelms the second and $y(\theta)$ increases in productivity. In turn, from (3), employment $h(\theta)$ increases in productivity. Firm revenue in each market served is monotonic in output and quality and hence increases in productivity. This implies that there is an exporting cutoff θ_x such that $I_x(\theta) = 0$ if and only if $\theta < \theta_x$ and thus firm market access $\Upsilon(\theta)$ increases in productivity. Finally, by (10) and (11), total revenue $r(\theta)$ and profits $\Pi(\theta)$ increase in productivity.

A.5.3 Optimality of the Closed-form Solution for $q(\theta)$

In this section, I verify that the expression for firm quality (15) is optimal in problem (7); i.e., $q(\theta) = q_d(\theta) = q_x(\theta)$, under the functional form assumptions (12)-(14). By equation (9), 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 (12)-(14),

$$\begin{aligned} c^{\theta}(q) &= \overline{u}\exp\left[k\epsilon(\theta,q)\right]\theta^{-s}, \\ &= \overline{u}\exp\left[k\left(\frac{\log q}{(\gamma\log\theta)^{z}}\right)^{1/(1-z)}\right]\theta^{-s}, \\ &= \overline{u}\exp\left[\Lambda\left(\log q\right)^{1/(1-z)}\right]\theta^{-s}, \end{aligned} \tag{A-29}$$

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

$$\frac{d\left(c^{\theta}(q)/q\right)}{dq} = \overline{u}\theta^{-s}\exp\left[\Lambda\left(\log q\right)^{1/(1-z)}\right]q^{-2}\left(\frac{\Lambda}{1-z}\left(\log q\right)^{z/(1-z)} - 1\right). \tag{A-30}$$

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 (15),

$$\left[\log\left(q(\theta)\right)\right]^{z/(1-z)} = \left[\frac{1-z}{k}\right]\left[\gamma\log(\theta)\right]^{z/(1-z)} = \frac{1-z}{\Lambda}. \tag{A-31}$$

For any $q > 0$, (A-30) and (A-31) 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 (7). As shown in the text, $q(\theta) \in [q_L, q_H]$ provided that effort is defined over a sufficiently large interval.

A.5.4 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-29) and (15), respectively. Equilibrium unit costs increase across firms if and only if $\gamma\left[(1-z)^{1/z}k^{(1-z)/z}\right] > s$. This condition 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.6 Proof of Proposition 3

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 $G_\theta(\theta)$, (25) and (26) 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 (23) and (27); 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 (28) and solving for $\overline{u}$ yields

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

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.7 Proof of Lemma 2

Preliminaries. From equation (29), re-write the employment distribution across firms in equilibrium $j \in \{0, 1\}$ as

$$G_{h,j}(\theta) = \begin{cases} 0, & \theta_L \le \theta \le \theta_{d,j}, \\ D_j^{-1} \int_{\theta_{d,j}}^{\theta} (\theta')^{\Gamma - k\kappa_\epsilon} dG_\theta(\theta'), & \theta_{d,j} \le \theta \le \theta_{x,j}, \\ D_j^{-1} \int_{\theta_d}^{\theta} \Upsilon_j(\theta') (\theta')^{\Gamma - k\kappa_\epsilon} dG_\theta(\theta'), & \theta_{x,j} \le \theta \le \theta_H, \end{cases} \tag{A-32}$$

where

$$D_j \equiv \int_{\theta_{d,j}}^{\theta_H} \Upsilon_j(\theta') \left(\theta'\right)^{\Gamma - k\kappa_\epsilon} dG_\theta(\theta'),$$

$$\Upsilon_j(\theta') = \begin{cases} \Upsilon_{x,j} & \text{if} \quad \theta' \geq \theta_{x,j}, \\ 1 & \text{if} \quad \theta' < \theta_{x,j}, \end{cases}, \tag{A-33}$$

$$\Upsilon_{x,j} = 1 + (f_x/f_d)\left(\theta_{d,j}/\theta_{x,j}\right)^{\Gamma}.$$

Letting $\theta_{x,0} \to \theta_H$ or $\theta_{x,1} \to \theta_{d,1}$ in (A-32) 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 $\tau_0 > \tau_1$ by hypothesis, then $\theta_{d,1}/\theta_{x,1} > \theta_{d,0}/\theta_{x,0}$ by equation (25) and thus $\Upsilon_{x,1} > \Upsilon_{x,0} > 1$. In addition, the free entry condition (26) 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 $\tau_0 > \tau_1$, $\tau_1 \in [\underline{\tau}, \overline{\tau})$, 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, $\tau_0 \geq \overline{\tau}$ 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 $\tau_0 < \overline{\tau}$ or (ii) $\theta_L < \theta_{d,0} < \theta_{d,1} \leq \theta_{x,1} < \theta_H < \theta_{x,0}$ if $\tau_0 \geq \overline{\tau}$.

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, $\tau_0 < \overline{\tau}$, 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 $\tau_0 \geq \overline{\tau}$, 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-2 *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*

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

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-32). 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:

$$G'_{h,0}(\theta) = \begin{cases} \Lambda G'_{h,1}(\theta), & \text{if } \theta \in R_C, \\ \frac{\Lambda}{\Upsilon_{x,1}} G'_{h,1}(\theta), & \text{if } \theta \in R_D, \\ \Lambda \frac{\Upsilon_{x,0}}{\Upsilon_{x,1}} G'_{h,1}(\theta), & \text{if } \theta \in R_E. \end{cases} \tag{A-35}$$

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-35).

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-34) yields

$$G_{h,1}(\theta) - G_{h,0}(\theta) = (1 - s_i) \int_{\widetilde{\theta}}^{\theta} G'_{h,1}(v) dv. \tag{A-36}$$

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 $\widetilde{\theta}$.[67] A similar argument establishes the single-crossing property from above if and only if $s_i > 1$. Finally, setting $s_i = 1$ in (A-36) 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-36) imply $s_i = 1$. ■

[67]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.

Part (a) of Lemma 2. If $\tau_0 \geq \overline{\tau}$, 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,

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

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.8 Proof of Proposition 4

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 $\tau_0 < \overline{\tau}$. From the employment distribution (29),

$$1 - G_{h,j}(\widehat{\theta}) = \frac{\int_{\widehat{\theta}}^{\theta_H} \theta^{\Gamma - k\kappa_\epsilon} dG_\theta(\theta)}{(\Upsilon_{x,j})^{-1} \int_{\theta_{d,j}}^{\theta_{x,j}} \theta^{\Gamma - k\kappa_\epsilon} dG_\theta(\theta) + \int_{\theta_{x,j}}^{\theta_H} \theta^{\Gamma - k\kappa_\epsilon} dG_\theta(\theta)}, \tag{A-37}$$

for $j \in \{0, 1\}$, where $(\Upsilon_{x,j})^{-1} = 1/(1 + (\tau_j)^{-\rho/(1-\rho)}) \in (0, 1)$. Therefore, $G_{h,1}(\widehat{\theta}) \leq G_{h,0}(\widehat{\theta})$ if and only if the denominator of (A-37) is increasing in the variable trade cost. Without loss of generality, I focus on infinitesimal changes in τ and show that $D'(\tau) > 0$, where

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

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

$$
\begin{aligned}
D'(\tau) &= \frac{\rho}{1-\rho}\left(\Upsilon_x\right)^{-2}\tau^{-1/(1-\rho)}\int_{\theta_d}^{\theta_x}\theta^{\Gamma-k\kappa_\epsilon}dG_\theta(\theta)+ \\
&+\left[\left(\Upsilon_x\right)^{-1}-1\right]\left(\theta_x\right)^{\Gamma-k\kappa_\epsilon}\theta_x'(\tau)g_\theta(\theta_x)-\left(\Upsilon_x\right)^{-1}\left(\theta_d\right)^{\Gamma-k\kappa_\epsilon}\theta_d'(\tau)g_\theta(\theta_d), \\
&>\left(\Upsilon_x\right)^{-1}\left(\theta_d\right)^{\Gamma-k\kappa_\epsilon}\left\{\left[\frac{\left(\Upsilon_x\right)^{-1}-1}{\left(\Upsilon_x\right)^{-1}}\right]\left(\frac{\theta_x}{\theta_d}\right)^{\Gamma-k\kappa_\epsilon}\theta_x'(\tau)g_\theta(\theta_x)-\theta_d'(\tau)g_\theta(\theta_d)\right\}, \\
&>\left(\Upsilon_x\right)^{-1}\left(\theta_d\right)^{\Gamma-k\kappa_\epsilon}\left\{-\frac{f_x}{f_d}\theta_x'(\tau)g_\theta(\theta_x)-\theta_d'(\tau)g_\theta(\theta_d)\right\},
\end{aligned}
\tag{A-38}
$$

where $\theta_m'(\tau)\equiv d\theta_m/d\tau$ for $m\in\{d,x\}$. The last line follows from $\left(\Upsilon_x\right)^{-1}-1=-\left(\Upsilon_x\right)^{-1}(f_x/f_d)\left(\theta_d/\theta_x\right)^{\Gamma}$ (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 (25), the free entry condition (26) and the monotonicity of $J(\cdot)$ imply $\theta_d'(\tau)<0$ and $\theta_x'(\tau)>0$. Therefore, from (A-38), $D'(\tau)>0$ if

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

where the equality follows by applying the Implicit Function Theorem on the free entry condition (26). To conclude, $D'(\tau)>0$ 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.9 Productivity Distributions that Satisfy Assumption 2

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

$$
J'(\theta)=-\Gamma\theta^{-(\Gamma+1)}\int_\theta^{\theta_H}v^{\Gamma}dG_\theta(v),
\tag{A-39}
$$

and

$$
J''(\theta)=-\left(\Gamma+1\right)\frac{J'(\theta)}{\theta}+\Gamma\frac{g_\theta(\theta)}{\theta}.
\tag{A-40}
$$

Densities with Elasticity Greater Than Or Equal To $-(\Gamma+1)$. From (A-40), it follows that

$$\begin{aligned}\frac{d}{d\theta}\left(\frac{J'(\theta)}{g_\theta(\theta)}\right) &\geq 0\\ &\Leftrightarrow J''(\theta)g_\theta(\theta) \geq J'(\theta)g_\theta'(\theta),\\ &\Leftrightarrow -J'(\theta)\left[(\Gamma+1)\frac{1}{\theta}+\frac{g_\theta'(\theta)}{g_\theta(\theta)}\right] \geq -\Gamma\frac{g_\theta(\theta)}{\theta}. \qquad \text{(A-41)}\end{aligned}$$

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-41) 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-39),

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

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.10 Characterization of Assumption 2 in Terms of Revenue

The additional structure imposed by Assumption 2 on the productivity distribution has implications for the distribution of revenue. This section derives a characterization of Assumption 2 in terms of the hazard function of a normalized distribution of revenue in the domestic market. Because domestic revenue is typically easier to observe than productivity, this characterization provides a testable implication of Assumption 2, a point I return to at the end of this section.

Let $r_N(\theta)$ denote the *normalized* domestic revenue of firm θ, defined as the ratio of domestic revenue of firm θ to domestic revenue of a benchmark firm, say θ_b. Setting $\Upsilon(\theta) = \Upsilon(\theta_b) = 0$ in the revenue function (17),

$$r_N(\theta) = (\theta/\theta_b)^{\Gamma} . \tag{A-42}$$

$r_N(\theta)$ has two important properties. First, the normalization implies that $r_N(\theta)$ is invariant across any two equilibria of the model (i.e. independent of A and $\overline{u}$), as long as the benchmark firm θ_b is held constant. In this case, the distribution of normalized domestic revenue is a structural object that depends only on the underlying distribution of firm productivity. Second, $r_N(\cdot)$ provides a strictly increasing and differentiable mapping between domestic revenue and productivity.[68]

These properties allow me to implement a change of variables to re-write the key ratio in Assumption 2, $J'(\theta)/g_\theta(\theta)$, in terms of normalized domestic revenue. Fix $\theta \in [\theta_L, \theta_H]$. For the denominator $g_\theta(\theta)$, note that the distribution of normalized domestic revenue at point $r = (\theta/\theta_b)^{\Gamma}$, denoted $G_N(r)$, satisfies $G_N(r) = G_\theta(r_N^{-1}(r))$. Hence its density, denoted $g_N(r)$, is linked to $g_\theta(\theta)$ by

$$g_N(r) = g_\theta(\theta) r^{1/\Gamma - 1} \Gamma^{-1} \theta_b .$$

Turning to the numerator $J'(\theta)$, use (A-39) and (A-42) to obtain

$$J'(\theta) = -r^{-(\Gamma+1)/\Gamma} \Gamma \left(\theta_b\right)^{-1} \int_r^{r_H} v dG_r(v),$$

where $r_H = (\theta_H/\theta_b)^{\Gamma}$. Therefore,

$$\begin{aligned} \frac{J'(\theta)}{g_\theta(\theta)} &= \frac{-r^{-2}}{g_N(r)} \int_r^{r_H} v dG_r(v), \\ &= -r^{-2} \frac{1 - G_N(r)}{g_N(r)} E\left[v \mid v \geq r\right]. \end{aligned}$$

Note that $-r^{-2}$ and $E\left[v \mid v \geq r\right]$ are non-decreasing in r, for any $G_N(r)$. Hence Assumption 2 holds if the hazard function of the distribution of normalized domestic revenue, $g_N(r)/\left[1 - G_N(r)\right]$, is non-increasing.

Importantly, all Pareto distributions have decreasing hazard functions. Di Giovanni et al. (2011) show that the distribution of domestic sales of French firms is very well approximated by a power law (Pareto), with an R-squared in excess of 99 percent.

[68] In contrast, note from (17) that total revenue is discontinuous in productivity.

A.11 Proof of Proposition 5

From part (a) of Corollary 1 and the expression for optimal effort (16), $\omega(\theta) = \overline{u}_j \theta^{k\kappa_\epsilon}$. From expression (30), 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-2. 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-35). Since the exact definition of Λ is immaterial in the proof of Lemma A-2, 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 $\tau_0 \geq \overline{\tau}$, then $G_{w,1}$ first-order stochastically dominates $G_{w,0}$.

(b) If $\tau_0 < \overline{\tau}$, 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 5, it thus suffices to show that, if $\tau_0 < \overline{\tau}$, 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 $\tau_0 < \overline{\tau}$ and fix $\widehat{\theta} \in [\theta_{x,0}, \theta_H)$. Using the cross-firm wage distribution (30) 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-33), yields

$$1 - G_{w,j}(\widehat{\theta}) = \frac{\int_{\widehat{\theta}}^{\theta_H} \theta^\Gamma dG_{\theta,j}(\theta)}{(\Upsilon_{x,j})^{-1} \int_{\theta_{d,j}}^{\theta_{x,j}} \theta^\Gamma dG_{\theta,j}(\theta) + \int_{\theta_{x,j}}^{\theta_H} \theta^\Gamma dG_{\theta,j}(\theta)}, \quad \text{(A-43)}$$

for $j \in \{0, 1\}$, where $(\Upsilon_{x,j})^{-1} = 1/(1 + (\tau_j)^{-\rho/(1-\rho)}) \in (0, 1)$. Therefore, $G_{w,1}(\widehat{\theta}) \leq G_{w,0}(\widehat{\theta})$ if and only if the denominator of (A-43) is increasing in the variable trade cost. I proceed by adjusting the proof of Proposition 4 to show that $D'(\tau) > 0$, where

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

after dropping index j to simplify notation. Following the steps leading to equation (A-38) yields

$$D'(\tau) > (\Upsilon_x)^{-1} (\theta_d)^\Gamma \left\{ -\frac{f_x}{f_d} \theta_x'(\tau) g_\theta(\theta_x) - \theta_d'(\tau) g_\theta(\theta_d) \right\}. \tag{A-44}$$

As shown in the proof of Proposition 4, Assumption 2 guarantees that the right-hand side of (A-44) 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.12 Proof of Proposition 6

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, τ, 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 $r_x = A^* q_x^\rho [y_x/\tau]^\rho$, 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 (7) and therefore the firm's solutions from section 4.2 apply here as well. In particular, note that the expression for firm employment (19) and thus the distribution of employment across firms (29) are still valid.

In asymmetric equilibria, the cutoff conditions for Home are straightforward extensions of equations (23) and (24). In equilibrium $j \in \{0, 1\}$,

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

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

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

$$\tau_j^{-\frac{\rho}{1-\rho}} \left(\frac{A_j^*}{A_j}\right)^{\frac{1}{1-\rho}} \left(\frac{\theta_{x,j}}{\theta_{d,j}}\right)^\Gamma = \frac{f_x}{f_d}, \tag{A-45}$$

in Home. Similarly, for Foreign,

$$\left(\tau_j^*\right)^{-\frac{\rho}{1-\rho}} \left(\frac{A_j}{A_j^*}\right)^{\frac{1}{1-\rho}} \left(\frac{\theta_{x,j}^*}{\theta_{d,j}^*}\right)^{\Gamma^*} = \frac{f_x^*}{f_d^*}. \tag{A-46}$$

Using (A-45) yields $\Upsilon_{x,j} = 1 + (f_x/f_d)\,(\theta_{d,j}/\theta_{x,j})^{\Gamma}$, which ensures that the employment distribution across firms in Home (29) can be expressed solely as a function of the productivity cutoffs in Home in any asymmetric equilibrium.

Consider a unilateral trade liberalization in Home, $\tau_1^* < \tau_0^*$. If Home is a small open economy, then $\theta_{x,j}^*$, $\theta_{x,j}^*$ and A_j^* are independent of τ_j^*. From (A-46), $A_1 < A_0$. This implies $\theta_{x,0}/\theta_{d,0} < \theta_{x,1}/\theta_{d,1}$ from (A-45). The free entry condition in Home is still written as (26) 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 (4) and (5) continue to hold in asymmetric equilibria, under the conditions stated in Proposition 6.

A.13 More on Lorenz-consistent Inequality Measures

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

$$\begin{aligned} MLD_j &\equiv E_j\left[\log\left(\frac{w_j^*}{w_i}\right)\right], \\ &= \int_{\theta_L}^{\theta_H} \log\left(\frac{w_j^*}{\omega(\theta)}\right) dG_{h,j}(\theta) + \int_{\theta_L}^{\theta_H} MLD\left(w_i|\theta\right) dG_{h,j}(\theta), \end{aligned} \tag{A-47}$$

and

$$\begin{aligned} T_j &\equiv E_j\left[\log\left(\frac{w_i}{w_j^*}\right)\frac{w_i}{w_j^*}\right], \\ &= \int_{\theta_L}^{\theta_H} \log\left(\frac{\omega(\theta)}{w_j^*}\right)\frac{\omega(\theta)}{w_j^*} dG_{h,j}(\theta) + \int_{\theta_L}^{\theta_H} T\left(w_i|\theta\right) dG_{w,j}(\theta), \end{aligned} \tag{A-48}$$

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 (A-47) and (A-48) 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).

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).[69] 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.[70]

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

[69] 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.

[70] 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.

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 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.

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 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.

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 firms 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 7, I use a finite population correction to construct unbiased estimates of firm-level variances, for both unadjusted log wages ($Var[\log w|\theta]$) and residual log wages ($\psi_{\theta,2}$). Here 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 $\lambda_{FPC} \sum_{i=1}^{n_\theta} (\log w_i - \widetilde{w}_\theta)^2 / n_\theta$ is an unbiased estimator of $Var[\log w|\theta]$, where $\lambda_{FPC} \equiv [1 - (1 - f_\theta)/n_\theta]^{-1}$ 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.[71]

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 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 ($Var[\log w|\theta]$) and residual log wages ($\psi_{\theta,2}$) reported in section 7 are adjusted using the finite population correction described in this section.

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

B.3 Additional Results

B.3.1 Prevalence of Performance Pay

Table B-III documents key facts about performance pay in Canada. As reported in the main text, 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.

Table B-III - Prevalence of Performance Pay

		PP Jobs		All jobs
Row	Sample	Per cent	Mean $\frac{pp_i}{wage_i}$	$\frac{std\ dev\ (pp_i)}{std\ dev\ (fixed_i)}$
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.

Moreover, while performance pay accounts for a relatively modest fraction of total compensation, it makes a substantial contribution to overall wage dispersion. 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.[72]

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

B.3.2 Between-firm and Within-firm Inequality

To gauge the contributions of between-firm and within-firm inequality across Canadian workers, I compute the ANOVA decomposition in equation (31). In row 1 of Table B-IV, within-firm inequality accounts for 35% of the variance of log weekly wages in the full sample, as reported in the main text. In the manufacturing sector, this share increases to 42% (row 2).[73] Table B-IV 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.[74]

Table B-IV - 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 B-IV 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.

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

[74] 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 (A-47) and (A-48), respectively.

The main takeaway from rows 4 to 6 is that the relative magnitude of within-firm inequality is particularly large when analyzing residual wage dispersion. This is important because the theory focuses precisely on within firm-inequality across homogeneous workers in a single-industry model; i.e. it is natural to interpret inequality in the model as residual wage dispersion across workers with identical observable characteristics, occupation and industry affiliation.

B.3.3 Within-firm Inequality in Other Countries

Is within-firm inequality in other countries comparable to Canada's? For comparability with recent studies in the literature, I compute the following decomposition of the variance of log wages within sectors and occupations:

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

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:

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

where e_i is the vector of observable worker characteristics, ψ_θ is a firm fixed effect and $v_{i\theta}$ is the error term.

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 B-IV, the decomposition (B-1) accounts for the covariance between worker observables and firm fixed effects when computing the relative magnitudes of between-firm and within-firm inequality.

Table B-V 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. (2017) 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-V across countries.

Table B-V - 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. (2017); Sweden, Akerman et al. (2013); Canada, this paper; France, Tito (2015). Figures may not sum to 100 due to rounding error.

Table B-V 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 B-IV. 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.

Akerman et al. (2013), Helpman et al. (2017) and Tito (2015) also report significant changes over time in within-firm inequality. While large increases in within-firm inequality are observed in Sweden and France, the opposite is true in Brazil. Card et al. (2013) and Bloom et al. (2016) report moderate increases in the U.S. and Germany, respectively. 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.[75] 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. It is, quite generally, not possible to infer causality from a simple before-after comparison. 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

[75]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.

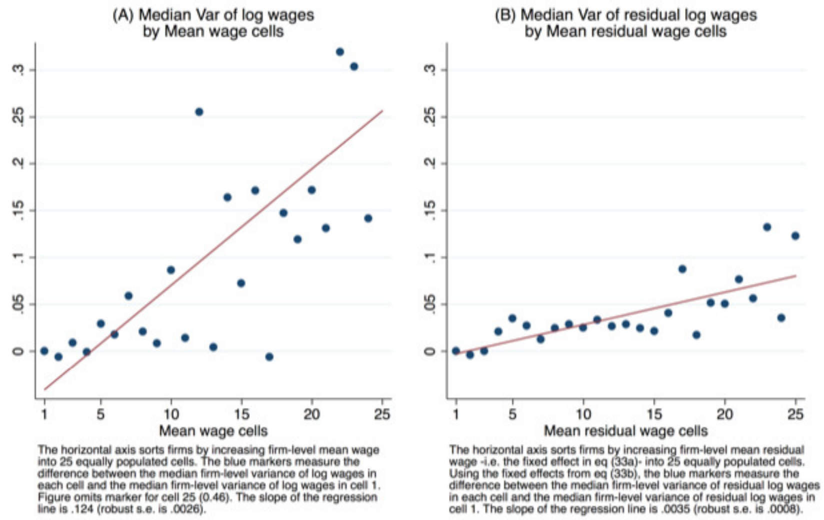

Figure B-1: Heterogeneity in Firm-level Wage Distributions

from the WES and by citing cross-sectional rather than time series evidence in the literature.

B.3.4 High Wage Firms are High Inequality Firms

Figure B-1 illustrates the finding that high wage firms are typically high inequality firms. 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.

In turn, Figure B-2 shows that the empirical patterns found in Figure B-1 hold under two alternative inequality measures: the Theil Index (Panels A and B) and the mean log deviation (Panels C and D).

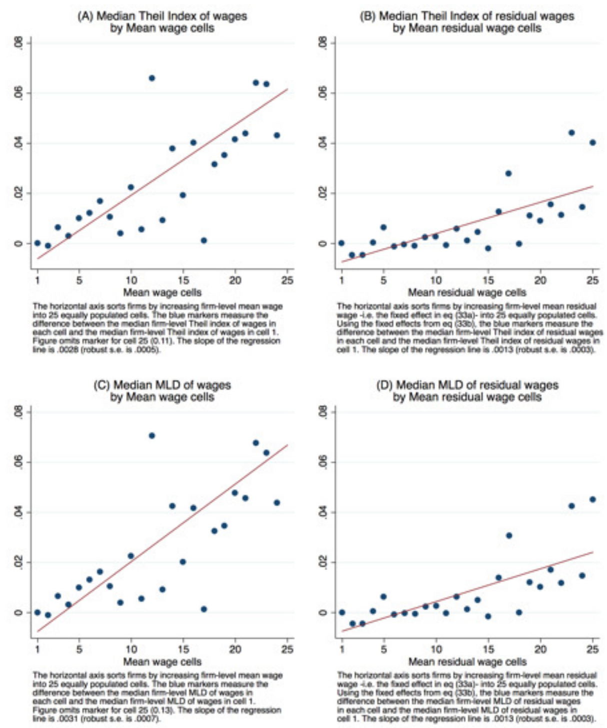

Figure B-2: Heterogeneity in Firm-Level Wage Distributions - Alternative Measures

B.3.5 Performance Pay and Firm Size

Table B-VI shows that the findings reported in Table I hold when an alternative proxy for firm size, total employment, is used.

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

Outcome	$P[pp_{i\theta} > 0 \mid \cdot]$			$E[pp_{i\theta} \mid \cdot]$			$Var[\log pp_{i\theta} \mid \cdot]$		
Dep. Var.	$I_{[pp_{i\theta}>0]}$			$pp_{i\theta}$			See table notes		
	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	18.477	17.748	27.433	0.200	0.208	0.192
	$(0.012)^c$	$(0.013)^b$	$(0.011)^a$	(12.100)	$(10.325)^c$	$(10.584)^a$	$(0.040)^a$	$(0.043)^a$	$(0.039)^a$
Exporter		-0.024	-0.000		6.229	25.587		-0.069	-0.071
		(0.038)	(0.036)		(41.339)	(48.846)		(0.117)	(0.118)
Controls	No	No	Yes	No	No	Yes	No	No	Yes
R-sq	0.00	0.00	0.12	0.00	0.00	0.17	0.03	0.03	0.10
Obs	14,265	14,265	14,265	6,775	6,775	6,775	6,775	6,775	6,775

Notes: this table reports OLS estimates of the right-hand side parameters of (34a) in columns 1 to 3, (34b) in columns 4 to 6 and (34c) in columns 7 to 9. '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. In columns 7 to 9, the dependent variable is the squared residual obtained from a regression of $\log pp_{i\theta}$ on firm fixed effects and Controls.

Standard errors (in parentheses) are clustered at the firm level. ^{a}, ^{b} and ^{c} denote statistical significance at the 1%, 5% and 10% levels, respectively.

B.3.6 Wages, Performance-independent Compensation and Firm Size

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

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

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

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 analyzed in Table I, $P[Y_{i\theta} > 0 \mid Size_\theta, Ex_\theta, e_i]$.

The results in Table B-VII indicate that firm size is positively correlated with all outcomes studied. The correlation between firm size and the dispersion of performance-independent compensation is weaker and significant only at the 10% level.[76] As in Table I, conditional on firm size, export status is uncorrelated with all outcomes.

Table B-VII - Firm Size and Wages

Outcome	$E[Y_{i\theta}\vert\cdot]$			$Var[Y_{i\theta}\vert\cdot]$		
Dep. Var.	$Y_{i\theta}$			See table notes		
	Basic	Add Exp Status	Add Controls	Basic	Add Exp Status	Add Controls
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A - $Y_{i\theta} = w_{i\theta}$ (wage)						
Firm size	84.465	84.414	69.177	0.008	0.008	0.010
	$(9.413)^a$	$(9.705)^a$	$(7.523)^a$	$(0.002)^a$	$(0.002)^a$	$(0.002)^a$
Exporter		0.594	-18.902		-0.008	-0.002
		(38.349)	(35.987)		(0.006)	(0.005)
Controls	No	No	Yes	No	No	Yes
R-sq	0.05	0.05	0.40	0.01	0.01	0.07
Obs	14,265	14,265	14,265	14,265	14,265	14,265
PANEL B - $Y_{i\theta} = fixed_{i\theta}$ (performance-independent compensation)						
Firm size	65.298	64.373	46.317	0.064	0.079	0.072
	$(7.694)^a$	$(8.165)^a$	$(6.792)^a$	$(0.039)^c$	$(0.048)^c$	$(0.041)^c$
Exporter		10.685	-25.630		-0.171	-0.160
		(30.481)	(25.941)		(0.108)	(0.107)
Controls	No	No	Yes	No	No	Yes
R-sq	0.04	0.04	0.41	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 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. In columns 4 to 6, the dependent variable is the squared residual obtained from a regression of $\log Y_{i\theta}$ on firm fixed effects and (in column 6) Controls.
Standard errors (in parentheses) are clustered at the firm level. ^{a} and ^{c} denote statistical significance at the 1% and 10% levels, respectively.

[76] Note that the model has nothing to say about this interesting fact. Since each firm offers a single contract to homogeneous employees, performance-independent compensation does not vary across co-workers. This empirical fact, however, provides a testable restriction for a futire extension of the theory that incorporates ex-ante worker heterogeneity.

C Extended Literature Review

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.[77]The latter is the main goal of this paper.

The theoretical results in this paper do not rely on the existence of trade-induced effects on firm-level wage distributions. This is an important property of the model because these effects receive only mixed support in empirical studies that control for worker heterogeneity.[78] In particular, $E(w|\theta)$ and $Var(\log w|\theta)$ do not change in

[77]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. As mentioned, within-firm inequality can 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.

[78]For example, controlling for worker-firm match productivity in Brazilian data, Krishna et al. (2014) find an insignificant exporter wage premium and insignificant differential effects of trade liberalization on the wages of workers at exporting firms relative to workers at non-exporting firms. Frıas et al. (2009), however, report significant differential effects following the 1994 Mexican devaluation. Schank et al. (2007) find small, yet significant, export wage premia conditional on workforce composition in German firms. Using Danish data, Munch and Skaksen (2008) find exporter wage premia only in firms where the skill intensity is sufficiently high. Frías et al. (2012) is the only paper I am aware of that studies firm-level wage dispersion. They find that an exogenous increase in the incentive to export, triggered by the Mexican peso devaluation in 1994, resulted in higher

response to trade liberalization because there is no quality upgrading or downgrading associated to exporting in the model. Heterogeneity in performance-pay contracts across firms, however, ensures that reductions in variable trade costs will still increase within-firm inequality through labor reallocations. Evidently, this mechanism will, in turn, be amplified by increases in firm-level variances driven by quality upgrading.[79]

There is a class of theories in which firm-level wage dispersion is driven by workforce composition, including Garicano and Rossi-Hansberg (2004, 2006), Antràs et al. (2006), Bustos (2011), Monte (2011), Burstein and Vogel (2012), Caliendo and Rossi-Hansberg (2012), Caliendo et al. (2015) and Harrigan and Reshef (2015). 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. Any two workers therefore earn different wages if and only if they have different levels of human capital. Conditional on worker skills, these models generate neither between-firm nor within-firm inequality. 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. The precise implications of this approach have not yet been formulated.[80] Still, while any potential explanation of the link between trade and residual within-firm inequality should feature wage variation among co-workers, i.e. $Var(\log w|\theta) > 0$ for

firm-level wage dispersion. However, Frías et al. (2012) do not control for workforce composition.

[79] A footnote on page 14 discusses how quality upgrading in response to trade liberalization can be introduced in the model, along the lines of Verhoogen (2008). With this feature, exporting 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)).

[80] 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.

some θ, it is essential to realize that the latter is a necessary yet insufficient ingredient for the task at hand. A comparative statics analysis of within-firm inequality -the second term on the right-hand side of (31)- inevitably requires a joint characterization of firm-level wage distributions *and* equilibrium changes in the distribution of workers across firms. In fact, both of these elements are typically absent in this literature, since it has focused on studying wage premia and firm-level outcomes. This paper is the first to tackle the comparative statics exercise. In addition, it focuses on a distinct and empirically relevant channel, performance pay, that is not easily understood purely in terms of unobserved skill heterogeneity priced in competitive labor markets.

Two recent papers combine cross-firm variation in workforce composition and, unlike the papers discussed in the previous two paragraphs, frictions in the labor market. Chen (2015) extends Caliendo and Rossi-Hansberg (2012) by introducing efficiency wage considerations to study the pro-competitive effects of an improvement in monitoring technology. Bombardini et al. (2015) analyze how international trade affects sorting patterns between heterogeneous firms and workers. Each firm hires workers of different observable skills but wages are bargained due to the existence of search frictions in the labor market. In both of these papers, however, identical co-workers earn identical wages. Interestingly, Bombardini et al. (2015) show that the dispersion of worker *types* decreases in firm size, conditional on export status, and provide evidence of this using matched employer-employee data for France. These results suggest that unobserved variation in workforce composition might, if anything, generate a bias against finding that high wage firms are also high inequality firms, a prediction of the model that is confirmed in my empirical analysis.[81]

Wage dispersion among co-workers can also arise in dynamic models of firm heterogeneity with directed search, such as Felbermayr et al. (2015) and Ritter (2017). In the presence of 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.[82] 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.

[81] Due to the cross-sectional nature of the WES, I do not control for cross-firm variation in time-invariant characteristics of workers.

[82] This observation also applies to implicit contract models, such as Harris and Holmstrom (1982) and Beaudry and DiNardo (1991), in which the history of labor market conditions experienced since the start of an employment spell affects current wages.

This paper is related to a growing empirical literature studying managerial practices at the firm level, surveyed in Bloom and Van Reenen (2011). The evidence indicates that performance pay covers about 40% to 50% of U.S. workers in the 2000s. Individual incentives typically increase worker productivity and were almost three times more prevalent that group incentives in large U.S. firms in 1999. Moreover, in line with the theory in this paper, Bloom and Van Reenen (2007) show that the intensity of performance pay increases in firm size. They report positive correlations between the extent to which firms reward performance and total revenue in the U.S., France, Germany, and the United Kingdom.

A large empirical literature in labor economics and international trade studies the contribution of sources of wage variation using matched employer-employee data. In particular, Akerman et al. (2013), Card et al. (2013), Helpman et al. (2017), Bloom et al. (2016) and Tito (2015) decompose wage (or earnings) inequality into between and within components in Sweden, Germany, Brazil, the U.S. and France, respectively.[83] 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.[84]

Finally, the paper is 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)). 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.

[83] These papers also report significant changes in within-firm inequality over time. I do not emphasize these findings for reasons discussed in section B.3.3.

[84] 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 on page 26 provides the intuition.

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

Antràs, P., Garicano, L., Rossi-Hansberg, E., 2006. Offshoring in a knowledge economy. Quarterly Journal of Economics 121 (1).

Beaudry, P., DiNardo, J., 1991. The effect of implicit contracts on the movement of wages over the business cycle: Evidence from micro data. Journal of Political Economy, 665–688.

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

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

Bloom, N., Van Reenen, J., 2011. Human resource management and productivity. Handbook of labor economics 4, 1697–1767.

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., Monte, F., Rossi-Hansberg, E., 2015. The anatomy of french production hierarchies. Journal of Political Economy 123 (4), 809–852.

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.

Carter, M., 2001. Foundations of mathematical economics. MIT Press.

Chen, C., 2015. Management quality and firm organization in a global economy. University of Hong Kong, Mimeo.

Cuñat, V., Guadalupe, M., 2005. How does product market competition shape incentive contracts? Journal of the European Economic Association 3 (5), 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.

Davis, S. J., Haltiwanger, J., 1995. Employer size and the wage structure in u.s. manufacturing. NBER Working Paper 5393.

Di Giovanni, J., Levchenko, A. A., Ranciere, R., 2011. Power laws in firm size and openness to trade: Measurement and implications. Journal of International Economics 85 (1), 42–52.

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.

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.

Frıas, J. A., Kaplan, D. S., Verhoogen, E. A., 2009. Exports and wage premia: Evidence from mexican employer-employee data. Unpub. paper, Columbia University.

Garicano, L., Rossi-Hansberg, E., 2004. Inequality and the organization of knowledge. The American Economic Review: Papers and Proceedings 94 (2), 197–202.

Garicano, L., Rossi-Hansberg, E., 2006. Organization and inequality in a knowledge economy. Quarterly Journal of Economics 121 (4), 1383–1435.

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

Harrigan, J., Reshef, A., 2015. Skill-biased heterogeneous firms, trade liberalization and the skill premium. Canadian Journal of Economics 48 (3), 1024–1066.

Harris, M., Holmstrom, B., 1982. A theory of wage dynamics. The Review of Economic Studies 49 (3), 315–333.

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., 2017. Trade and inequality: From theory to estimation. The Review of Economic Studies 84 (1), 357–405.

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.

Krishna, P., Poole, J. P., Senses, M. Z., 2014. Wage effects of trade reform with endogenous worker mobility. Journal of International Economics 93 (2), 239–252.

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

Laffont, J.-J., Martimort, D., 2002. The Theory of Incentives: The Principal-Agent Model. Princeton University Press.

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

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.

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.

Munch, J. R., Skaksen, J. R., 2008. Human capital and wages in exporting firms. Journal of International Economics 75 (2), 363–372.

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

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

Ritter, M., 2017. Inequality and international trade: The role of skill-biased technology and search frictions. Macroeconomic Dynamics 21 (3), 624–643.

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.

Schank, T., Schnabel, C., Wagner, J., 2007. Do exporters really pay higher wages? first evidence from german linked employer–employee data. Journal of International Economics 72 (1), 52–74.

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.

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.

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