TECHNISCHE UNIVERSITÄT DRESDEN

CEPIE Working Paper No. 12/17

Center of Public and International Economics

THE EXPORTER WAGE PREMIUM WHEN FIRMS AND WORKERS ARE HETEROGENEOUS

July 2017

Hartmut Egger
Peter Egger
Udo Kreickemeier
Christoph Moser

Editors: Faculty of Business and Economics, Technische Universität Dresden.

ISSN 2510-1196

The Exporter Wage Premium When Firms and Workers are Heterogeneous*

Hartmut Egger†
University of Bayreuth
CESifo, GEP, and IfW

Peter Egger
ETH Zurich
CEPR, CESifo, GEP

Udo Kreickemeier
TU Dresden
CESifo, GEP

Christoph Moser
University of Salzburg
CESifo

July 18, 2017

Abstract

We set up a trade model with heterogeneous firms and a worker population that is heterogeneous in two dimensions: workers are either skilled or unskilled, and within each skill category there is a continuum of abilities. Workers with high abilities, both skilled and unskilled, are matched to firms with high productivities, and this leads to wage differentials within each skill category across firms. Self-selection of the most productive firms into exporting generates an exporter wage premium, and our framework with skilled and unskilled workers allows us to decompose this premium into its skill-specific components. We employ linked employer-employee data from Germany to structurally estimate the parameters of the model. Using these parameter estimates, we compute an average exporter wage premium of 5 percent. The decomposition by skill turns out to be quantitatively highly relevant, with exporting firms paying no wage premium at all to their unskilled workers, while the premium for skilled workers is 12 percent.

JEL classification: C31, F12, F15, J31

Keywords: Exporter wage premium; Heterogeneous firms; Ability differences of workers; Positive assortative matching; Trade and wage inequality

*We would like to thank Xenia Matschke, Peter Neary, and participants at the ETSG Annual Conference, the GEP China Conference in Ningbo, the CESifo Global Area Conference in Munich, the Dresden Conference on International Trade and Production, the Annual Conference of the German Economic Association's Research Group on International Economics, the Göttingen Workshop in International Economics as well as seminar participants at ETH Zürich, LMU Munich, University of Sussex and University of Tübingen for very helpful comments and suggestions.

†Corresponding author: University of Bayreuth, Department of Economics, Universitätsstr. 30, 95447 Bayreuth, Germany; Email: hartmut.egger@uni-bayreuth.de.

1 Introduction

The recent literature on international trade has pointed out that the asymmetric exposure of firms to exporting is an important channel through which trade affects the distribution of wage incomes in open economies. Based on strong empirical evidence that larger, more productive firms pay higher wages and have a higher probability of exporting (cf. Bernard and Jensen, 1995, 1999), existing theoretical studies argue that the selection of successful firms into exporting augments wage inequality through two channels: an increase in the number of jobs in exporting firms offering high wages; and an increase in the wages exporting firms pay relative to the wages paid by non-exporters, the so-called exporter wage premium. Different theoretical foundations for such a premium have been provided in recent trade models with firm heterogeneity (cf. Helpman et al., 2010; Amiti and Davis, 2012; Egger and Kreickemeier, 2012; Sampson, 2014).[1]

The aim of this paper is twofold. First, we develop a theoretical model that allows us to decompose the exporter wage premium into its skill-specific components. Second, we structurally estimate the model parameters using matched employer-employee data for Germany and use those estimated parameter values to compute the skill-specific exporter wage premia in a theory-consistent way. In our model, the population of workers is heterogeneous in two dimensions: formal skills and abilities. The productivity of firms is a composite of three components: the innate baseline productivity, and the average ability levels of the skilled and unskilled workers they hire, all of which differ across firms. Effective firm productivity is an increasing function of the respective firm's exogenous baseline productivity and of the abilities of its skilled and unskilled workforce. We assume that there is a firm-specific ability threshold below which firms cannot productively employ workers. The ability threshold is increasing in the baseline productivity of the firm, and as a consequence our model features positive assortative matching of high-ability skilled and unskilled workers to firms with a high baseline productivity. This generates wage dispersion between firms, which is explained by ability differences of the workforce. Whereas this outcome is similar to Sampson (2014), in our setting it is the consequence of an ability threshold and does not require strict log-supermodularity of effective firm productivity.[2] It is this feature that keeps our model analytically tractable and helps making it accessible to structural estimation.

There are two possible outcomes of the matching process in autarky. Wages per efficiency unit of labour can be equalised within a skill category, and firms hire workers with differing abilities within this category. In this case, we speak of a pooling equilibrium. Alternatively, if

[1]Evidence for the existence of an exporter wage premium is available for numerous countries, including the US (Bernard and Jensen, 1999), Mexico (Frias et al., 2009), and several European economies (Schank et al., 2007; Mayer and Ottaviano, 2008; Egger et al., 2013; Hauptmann and Schmerer, 2013; Irrazabal et al., 2013; Klein et al., 2013). Egger and Kreickemeier (2009) and Davis and Harrigan (2011) also propose theoretical models in which exporters pay higher wages than non-exporters. However, in their settings, wage differentials between firms are pinned down by technology parameters and are not affected by the exporting decision of firms. Hence, these papers do not feature an exporter wage premium.

[2]With wage differences rooted in ability differences, our paper is complementary to the approaches of Helpman et al. (2010), Amiti and Davis (2012), and Egger and Kreickemeier (2012), in which the wage dispersion between firms is a consequence of rent sharing in the presence of labour market imperfections.

there is excess demand for high-ability workers along a wage profile with equalised wages per efficiency unit of labour, relative wages of high-ability workers are driven up, thereby leading to a separating equilibrium in which firms employ only workers of the lowest ability compatible with the firm-specific ability threshold. In the separating equilibrium, high-productivity firms end up paying higher wages per efficiency unit of labour than their low-productivity competitors, thereby reducing the cost advantage of more productive firms. The equilibrium matching outcome crucially depends on the scarcity of high abilities of the two skill types in the production process, which is unobservable. However, we observe in the data that wages are more dispersed for unskilled than for skilled labour, which suggests that abilities of unskilled workers are more dispersed than abilities of skilled workers. To capture the empirical pattern of wage dispersion, we therefore impose a parameter constraint ensuring a more equal ability distribution for skilled than for unskilled workers. This parameter constraint implies that in autarky we have a pooling equilibrium for unskilled labour and a separating equilibrium for skilled labour.

In the open economy, our model features three groups of firms. All highly productive firms self-select into exporting, while none of the low-productivity firms export. Of particular interest are firms with an intermediate productivity level, since we can show that in this group exporters and non-exporters co-exist. Hence, our model features an overlap in the productivity distributions of exporting and non-exporting firms, which is typically observed in the data but not accounted for in most international trade models featuring firm heterogeneity.[3] Export market entry of the most productive firms leads to excess demand for skilled workers of high ability along the autarky wage profile, leading to an exporter wage premium for skilled workers, defined in a theory-consistent way as the average increase in the wages of exporting firms relative to the counterfactual benchmark of autarky. In addition to the premium for skilled workers, there may or may not be an exporter wage premium for unskilled workers, depending on whether the additional demand for high-ability workers in the unskilled category can be met with equalised wages for efficiency units of unskilled labour across firms. In both cases, the effect differs between the two skill groups, and, hence, our model features skill-specific exporter wage premia, and thus skill-specific effects of trade on the dispersion of wages.

In a second part of the paper, we use our theoretical model as guidance for an empirical analysis. We employ information from the linked employer-employee dataset (LIAB) of the Institute for Employment Research (IAB), which provides detailed information on German firms and workers over the period 1996-2008, to structurally estimate key parameters of our model. Based on these parameter estimates we then contrast observed measures of wage inequality with those computed from the model. For doing so, we rely on the Theil index as a measure of inequality that allows us to decompose total wage inequality into its within and between skill-group components. The model does a fairly good job in explaining overall wage inequality in the data and correctly predicts wage inequality within the group of unskilled workers. However,

[3]An exception in this respect is Armenter and Koren (2015), who show that presuming a sharp selection of the best firms into exporting exaggerates the role played by the extensive margin for explaining the effect of changes in trade costs on export sales.

the model underestimates wage inequality within the group of skilled workers as well as wage inequality between the two skill groups. Overall, accounting for scarcity of skilled workers with high abilities helps to improve the fit of our model with various aspects of wage inequality observed in the data.

Evaluating our model at the estimated parameter values, we find that exporters pay a wage premium only to their skilled workforce. This premium is sizable and amounts to 12 percent. Weighted with the share of skilled workers employed by exporters, we compute an exporter wage premium of almost 5 percent on average for all workers. This magnitude is well in line with estimates from previous empirical research (cf. Mayer and Ottaviano, 2008; Egger et al., 2013; Hauptmann and Schmerer, 2013; Klein et al., 2013, among others). However, existing empirical work on the exporter wage premium does usually not distinguish between skills, and, hence, the insights from the present model allow to draw a more nuanced picture of how exporting affects the wage differential between firms.[4]

We conduct two counterfactual experiments based on the estimated model in order to shed light on how the existence of an exporter wage premium affects the economy-wide distribution of labour income. In the first experiment, we introduce prohibitive trade costs and enforce a closed economy in the counterfactual equilibrium. This eliminates the exporter wage premium for skilled workers and has sizable distributional effects relative to the benchmark scenario. Relying on the Theil index, a movement to autarky reduces wage inequality within the group of skilled workers by 36 percent, whereas economy-wide inequality falls by 7 percent. In a second experiment, we assume that Germany moves from its observed degree of openness to free trade, and show that the elimination of iceberg trade costs has only small distributional effects and leads to a more equal distribution of wages, despite an increase in the exporter wage premium. This illustrates that the impact of trade on income inequality is non-monotonic, because the increase in the exporter wage premium triggered by a fall in variable trade cost is counteracted by a compositional effect as more (skilled) workers find employment in exporting firms and can therefore benefit from the exporter wage premium.

The remainder of the paper is organised as follows. In Section 2 we present our theoretical framework and derive the closed economy equilibrium. In Sections 3 and 4 we analyse the open-economy equilibrium, introduce the Theil index as measure of inequality, and discuss the distributional effects of trade. In Section 5 we introduce the dataset, structurally estimate key model parameters, and analyse the fit of our model with the observed distribution of wages. Section 6 uses the parameter estimates to quantify the exporter wage premium for skilled and unskilled workers, respectively. There, we also study in two counterfactual experiments how changes in the exporter wage premia affect economy-wide measures of inequality. In Section

[4]Two exceptions are the papers by Schank et al. (2007) and Klein et al. (2013), who derive estimates for skill-specific exporter wage premia from reduced-form regressions. By conditioning on firm size and using plant and/or spell (worker-firm) fixed effects, they control for most of the underlying firm heterogeneity, so that their estimates for the exporter wage premia exclude from consideration the wage premium that arises as a consequence of the selection of high-productivity firms into exporting.

7 we investigate to what extent the definition of skill groups influences the results from our analysis. The last section concludes with a summary of the most important results.

2 The closed economy

2.1 Preliminaries

We consider an economy with a mass N of producers, determined endogenously, and an exogenous mass L of workers. The population of workers is heterogeneous in two dimensions, which we call skill and ability. Skill refers to the formal education an individual has received, and according to this criterion we split the mass of workers into two groups, skilled workers L_s and unskilled workers L_u. Ability on the other hand refers to the exogenous productivity of an individual, for which we assume a continuous distribution within each skill group. Abilities of skilled and unskilled workers are denoted α and β, respectively.

Firms are monopolistically competitive, and they produce horizontally differentiated goods, for which they face the demand function

$$x(v) = Ap(v)^{-\sigma}, \tag{1}$$

where A is a variable capturing market size, v indexes the firm, and $\sigma > 1$ is the constant price elasticity of demand. Firm-level inputs of skilled and unskilled labour are denoted by l_s and l_u, respectively. We assume a Cobb-Douglas production function of the form

$$q(v) = \hat{\varphi}(v) l_s(v)^{\nu} l_u(v)^{1-\nu} \tag{2}$$

where $\hat{\varphi}(v)$ denotes the overall total factor productivity of firm v, which depends on its baseline productivity $\varphi(v)$ and on the average abilities of the skilled and unskilled workers in its workforce, $\tilde{\alpha}(v)$ and $\tilde{\beta}(v)$, respectively:

$$\hat{\varphi}(v) \equiv \varphi(v)^{\eta} \left[\tilde{\alpha}(v)^{\nu} \tilde{\beta}(v)^{1-\nu}\right]^{1-\eta} \tag{3}$$

Eqs. (2) and (3) imply that $\tilde{\alpha}(v)^{1-\eta}$ and $\tilde{\beta}(v)^{1-\eta}$ give the efficiency units per skilled and unskilled physical labour unit, respectively, as hired by firm v.

We assume that each firm can productively employ only workers above a threshold ability level. This minimum viable ability is higher in firms with a higher baseline productivity, reflecting in a stylised way the fact that more sophisticated firms require more sophisticated workers. For simplicity, and to save on notation in the following, we assume that the firm-specific minimum viable ability for both types of labour is equal to the exogenous baseline productivity of the respective firm.[5] Abilities are perfectly observable to the firm. Firms are therefore indifferent

[5]Sampson (2014) shows that under the assumption of *strict* log-supermodularity, which requires complementarity that is stronger than in the Cobb-Douglas case, positive assortative matching with a separating equilibrium

between hiring workers of different abilities within the same skill group if and only if the wage differential exactly compensates for the ability differential, resulting in a wage per efficiency unit of the respective type of labour that is the same for all abilities.

We assume that the baseline productivity φ, the ability of skilled workers α, and the ability of unskilled workers β are all Pareto distributed, with distribution functions $G(\varphi) = 1 - \varphi^{-g}$, $S(\alpha) = 1 - \alpha^{-s}$, and $U(\beta) = 1 - \beta^{-u}$, respectively, where we have normalised the lower bound for all three distributions to unity.[6]

2.2 Pooling equilibria and separating equilibria

For each skill type of workers, our model allows for two scenarios, which are distinguished by the elasticity of the wage schedule with respect to worker ability in this skill group. If this elasticity is equal to $1-\eta$, the wage rate per efficiency unit of labour is equalised across abilities, and firms are therefore indifferent between hiring workers of different ability within this skill group, provided their ability is higher than the firm-specific threshold level. We label this case the *pooling equilibrium*, since firms employ workers of different ability in equilibrium. If the wage schedule increases in worker ability with an elasticity greater than $1 - \eta$, the wage rate for an efficiency unit of labour increases in worker ability, and firms will therefore only employ workers of the minimum admissible ability in this skill group. We label this case the *separating equilibrium*.

The elasticity of the two wage schedules is of course endogenous, and it depends on the shape parameters of the Pareto distributions of firm baseline productivity and skill-specific worker ability. Intuitively, a steeply increasing wage profile (and, therefore, a separating equilibrium for workers in a given skill group) is more likely, ceteris paribus, if the relative supply of high-ability workers in this skill group is small, i.e., if the shape parameter of the respective ability distribution is high. We only consider a subset of all possible parametrisations of our model, thereby ruling out less interesting or less plausible cases. With two skill types and two types of equilibrium per skill type our model in principle allows for four different autarky regimes. We eliminate on theoretical grounds the two cases featuring either a pooling equilibrium for both skill types or a separating equilibrium for both skill types. In the former case, wages per labour efficiency unit are equalised across firms for all workers, and firms therefore play no interesting role for income inequality. In the latter case, firms face vertical supply curves for all factors of production, and therefore their output is pinned down by the available supply of a specific ability type, leaving no scope for adjustments in firm size in response to economy-wide shocks, such as trade liberalisation. Out of the two remaining cases, we focus on the one featuring a separating

in the labour market arises. We have a weaker assumption on technology (by using a Cobb-Douglas function that is log-supermodular but not strictly so), but therefore need the additional assumption of the firm-specific ability threshold to generate the same result. The Cobb-Douglas specification for effective firm productivity helps us in terms of analytical tractability, in particular, since we have two skill groups of workers.

[6] It is not essential that the distributions of abilities and baseline productivities all have the same domain. For instance, if the floor of one of the ability distributions were lower than the floor of the productivity distribution, some low-ability workers would be non-employable and, hence, would not show up in our dataset.

equilibrium for skilled workers and a pooling equilibrium for unskilled workers. This can be motivated by the observation documented in Section 5.2 that wage dispersion measured by the Theil index is less pronounced for unskilled workers than for skilled workers. With pooling of both skill types, skilled and unskilled wage profiles would have the same elasticity in our model and a larger ability dispersion of unskilled workers could then be directly inferred from a higher Theil index of wages for this skill type. With a Pareto distribution, a larger dispersion of abilities implies lower scarcity of high abilities. Since it is the excess demand for high abilities under pooling that establishes a separating equilibrium in our model, the skill-specific Theil indices reported below provide strong arguments for associating autarky with a pooling equilibrium of unskilled and a separating equilibrium of skilled workers. We assume the following parameter constraint in order to induce this equilibrium:

Assumption 1. $s \geq \bar{s} \equiv g + 1 - \eta\sigma > u$.

Lemma 1. *Assumption 1 establishes for the closed economy a separating equilibrium for skilled workers and a pooling equilibrium for unskilled workers.*

Proof. See the Appendix. □

An intuitive explanation for this parameter constraint is straightforward. As we show in the Appendix, $\bar{s}$ is (the absolute value of) the elasticity of labour demand with respect to φ in a knife-edge equilibrium, in which firms hire only workers *just* compatible with the ability threshold, but in which the wage per efficiency unit of labour is equalised across abilities. If and only if s, the elasticity of the skilled labour supply with respect to α (also in absolute value), is larger than this threshold value, the supply of skilled workers decreases faster with higher abilities than the demand for these skilled workers in the knife-edge equilibrium, and we therefore see wages per efficiency unit that are increasing in skilled worker ability. For an analogous reason, the parameter constraint in Assumption 1 induces a pooling equilibrium for unskilled workers under autarky.

2.3 Firm-specific variables

We now derive the profiles across firms for key firm-level variables, namely skilled employment, unit costs, skilled wages and output. In analogy to Melitz (2003), in the closed economy those firm-level variables increase in firm-specific baseline productivity φ with a constant elasticity, and we can therefore index firms in the following by φ alone.

The distribution of skilled employment across firms is determined by the labour market clearing condition, which for skilled workers under Assumption 1 must hold separately for each ability level. Formally, we require that the demand for skilled workers by firms with a baseline productivity up to $\bar{\varphi}$, denoted by $L_s^d(\bar{\varphi})$, must be equal to the supply of skilled workers with an ability up to $\bar{\alpha} = \bar{\varphi}$, denoted by $L_s(\bar{\varphi})$. We have

$$L_s^d(\bar{\varphi}) = \frac{N}{1 - G(\varphi_c)} \int_{\varphi_c}^{\bar{\varphi}} l_s(\varphi) dG(\varphi) \qquad \text{and} \qquad L_s(\bar{\varphi}) = L_s \left[1 - \left(\frac{\bar{\varphi}}{\varphi_c} \right)^{-s} \right],$$

where φ_c is the cutoff baseline productivity defining the least productive firm that is actually producing. The employment profile satisfying the condition $L_s^d(\bar{\varphi}) = L_s(\bar{\varphi})$ is given by

$$\frac{l_s(\varphi_1)}{l_s(\varphi_2)} = \left(\frac{\varphi_1}{\varphi_2} \right)^{g-s}, \tag{4}$$

where φ_1 and φ_2 are two arbitrary baseline productivity levels. Hence, firm-level employment of skilled workers can be higher or lower in more productive firms, depending on the relative size of g and s: If and only if the ability distribution of skilled labour has a fatter tail than the baseline productivity distribution of firms ($g > s$), firm-level employment of skilled labour is increasing in firm productivity. This is the empirically relevant case, because in our dataset for Germany revenues and skilled employment are positively correlated with a coefficient of 0.83, and revenues are increasing in firm-level productivity in our model under a parameter constraint introduced below.

Cost minimisation implies that firms choose input bundles such that the cost of skilled labour is a constant fraction ν of total cost. Hence, we have $\nu q(\varphi) c(\varphi) = w_s(\varphi) l_s(\varphi)$, where $c(\varphi)$ is the unit cost of firm φ and $w_s(\varphi)$ is the wage firm φ pays to its skilled workers, and consequently

$$\frac{l_s(\varphi_1)}{l_s(\varphi_2)} = \frac{c(\varphi_1)}{c(\varphi_2)} \frac{q(\varphi_1)}{q(\varphi_2)} \frac{w_s(\varphi_2)}{w_s(\varphi_1)}. \tag{5}$$

A goods-market equilibrium in the presence of constant markup pricing implies the standard link between relative unit costs and relative outputs:

$$\frac{q(\varphi_1)}{q(\varphi_2)} = \left(\frac{c(\varphi_1)}{c(\varphi_2)} \right)^{-\sigma}. \tag{6}$$

The unit cost function $c(\varphi)$ is given, under Assumption 1, by

$$c(\varphi) = \mu \varphi^{-\eta} [\hat{w}_s(\varphi)]^{\nu} [\hat{w}_u]^{1-\nu}$$

where $\mu \equiv \nu^{\nu}(1-\nu)^{1-\nu}$, $\hat{w}_s(\varphi) \equiv w_s(\varphi)/\varphi^{1-\eta}$ is the wage for an efficiency unit of skilled labour paid by firm φ, and $\hat{w}_u$ is the wage for an efficiency unit of unskilled labour, which is the same for all firms. The unit cost function implies the following link between relative unit costs, relative skilled wages and relative baseline productivities:

$$\frac{c(\varphi_1)}{c(\varphi_2)} = \left(\frac{w_s(\varphi_1)}{w_s(\varphi_2)} \right)^{\nu} \left(\frac{\varphi_2}{\varphi_1} \right)^{\nu(1-\eta)+\eta}. \tag{7}$$

Eqs. (5), (6) and (7) can be solved for relative unit costs, relative outputs, and relative skilled wages as functions of relative baseline productivities and relative skilled employment. Taking into account the employment profile required for a labour market equilibrium in Eq. (4), we get

$$\frac{c(\varphi_1)}{c(\varphi_2)} = \left(\frac{\varphi_1}{\varphi_2}\right)^{-\Delta}, \qquad \frac{q(\varphi_1)}{q(\varphi_2)} = \left(\frac{\varphi_1}{\varphi_2}\right)^{\sigma\Delta}, \qquad \frac{w_s(\varphi_1)}{w_s(\varphi_2)} = \left(\frac{\varphi_1}{\varphi_2}\right)^{\Phi}, \tag{8}$$

where

$$\Delta \equiv \eta - \frac{s - \bar{s}}{\sigma + (1-\nu)/\nu}, \qquad \Phi \equiv \frac{\nu + (1-\nu)\eta - \Delta}{\nu}. \tag{9}$$

From Assumption 1, we have $s \geq \bar{s}$, which implies $\Delta \leq \eta$. We furthermore assume that Δ is strictly positive, in which case it can be interpreted as the constant elasticity with which firm-level marginal costs decrease in firm-level baseline productivity. Firm-level output and firm-level revenues then increase in baseline productivity with elasticities $\sigma\Delta$ and $(\sigma-1)\Delta$, respectively. Φ is the elasticity of the skilled wage schedule with respect to firm-level baseline productivity. It can be seen that Φ is increasing in s, and that under the parameter constraint imposed by Assumption 1 elasticity Φ is always larger than $1-\eta$, as required for a separating equilibrium in the market for skilled labour. We furthermore assume $\Phi < s$ to ensure that the aggregate wage bill of skilled workers has a finite positive value (see below).

As explained above, firms are indifferent between hiring unskilled workers of the minimum admissible ability $\beta = \varphi$ or ones of higher ability, since the wage per efficiency unit of unskilled labour is equalised across ability levels. This has two implications. First, we can express relative wages of unskilled workers as a function of their relative abilities, *not* as a function of relative firm productivity, as

$$\frac{w_u(\beta_1)}{w_u(\beta_2)} = \left(\frac{\beta_1}{\beta_2}\right)^{1-\eta}, \tag{10}$$

with $1-\eta < u$ to ensure that the aggregate wage bill of unskilled workers has a finite positive value. Second, the number of physical units of unskilled labour hired by a firm of a given productivity is not determined (only the number of efficiency units is), and therefore it is not possible to derive a solution for the unskilled employment profile across firms.

2.4 General equilibrium

Having derived the constant-elasticity profiles for the relevant firm-level variables, we now use general equilibrium constraints to solve for the respective variables of the marginal firm in terms of model parameters, thereby anchoring the respective firm profiles. We also seek to find a solution for the endogenous mass of firms N.

All goods produced by monopolistically competitive firms serve as inputs into a homogeneous final good, for which we assume a CES production function without external scale economies, as in Egger and Kreickemeier (2009):

$$Y = \left[N^{-\frac{1}{\sigma}} \int_{\varphi_c}^{\infty} q(\varphi)^{\frac{\sigma-1}{\sigma}} \frac{dG(\varphi)}{1-G(\varphi_c)} \right]^{\frac{\sigma}{\sigma-1}}.$$

We choose the final good as the numeraire, and therefore the market size variable in Eq. (1) is equal to $A \equiv Y/N$, where Y is real GDP. Market entry occurs via a Melitz-style productivity lottery with fixed market entry cost f_e, paid in units of the final good, and we assume that all firms die after one period. Following Ghironi and Melitz (2005), we abstract from fixed production cost, and therefore all firms that enter the productivity lottery also produce. This implies $\varphi_c = 1$. In equilibrium, the expected profits of entering the lottery have to equal the fixed market entry cost: $R/(N\sigma) = f_e$, where R denotes aggregate revenues. Average revenues are linked to revenues of the marginal firm by

$$\frac{R}{N} = \int_1^{\infty} r(\varphi) dG(\varphi) = r(1)\Theta, \qquad \text{with} \qquad \Theta \equiv \frac{g}{g - \Delta(\sigma-1)} > 1$$

being a constant factor of proportionality.[7] We can therefore write revenues of the marginal firm as $r(1) = \sigma f_e \Theta^{-1}$. Using the goods market equilibrium, markup-pricing, and the equality $R = Y$, the implied marginal cost is given by $c(1) = \Theta^{\frac{1}{\sigma-1}}(\sigma-1)/\sigma$, and the output of the marginal firm is determined as $q(1) = \sigma f_e \Theta^{-\frac{\sigma}{\sigma-1}}$. As an intermediate step towards deriving the wage rates of the marginal firm, $w_s(1)$ and $w_u(1)$, we compute aggregate wage bills for skilled and unskilled workers as

$$W_s = L_s \int_1^{\infty} w_s(\alpha) dS(\alpha) = \frac{s\Theta}{g} L_s w_s(1) \tag{11}$$

and

$$W_u = L_u \int_1^{\infty} w_u(\beta) dU(\beta) = \Gamma L_u w_u(1), \tag{12}$$

respectively, with

$$\Gamma \equiv \frac{u}{u + \eta - 1} > 1$$

as the factor of proportionality linking the wage paid to unskilled workers with the lowest ability to the average wage paid to unskilled workers, and the composite term $s\Theta/g$ playing a similar role in the context of skilled workers. With a Cobb-Douglas production function, relative aggregate wage bills are constant and given by $W_s/W_u = \nu/(1-\nu)$. At the same time, wage rates $w_s(1)$ and $w_u(1)$ are linked via the unit-cost function of the marginal firm:

$$c(1) = \mu w_s(1)^{\nu} w_u(1)^{1-\nu}. \tag{13}$$

Using these conditions, and substituting for $c(1)$ from above, we get explicit solutions for $w_s(1)$

[7] It can easily be checked that $g > \Delta(\sigma - 1)$ is equivalent to $\Phi < s$, which is a parameter constraint introduced above.

and $w_u(1)$:

$$w_s(1) = \frac{\sigma-1}{\mu\sigma}\Theta^{\frac{1}{\sigma-1}}\left(\frac{\Gamma g\nu}{\Theta s(1-\nu)}\frac{L_u}{L_s}\right)^{1-\nu}$$
$$w_u(1) = \frac{\sigma-1}{\mu\sigma}\Theta^{\frac{1}{\sigma-1}}\left(\frac{\Theta s(1-\nu)}{\Gamma g\nu}\frac{L_s}{L_u}\right)^{\nu}$$

As expected, an increase in the relative mass of skilled workers L_s/L_u reduces the skilled wage paid by the marginal firm and increases the unskilled wage paid by that firm.

Skilled employment of the marginal firm follows from applying Shephard's Lemma to the unit cost function (13) and multiplying the resulting labour input coefficient by firm-level output:

$$l_s(1) = \nu\mu\left(\frac{w_u(1)}{w_s(1)}\right)^{1-\nu} q(1).$$

The mass of firms N follows from the full-employment condition $L_s = N\int_1^\infty l_s(\varphi)dG(\varphi)$:

$$N = \frac{L_s}{l_s(1)}\frac{s}{g}$$

This completes the characterisation of the closed economy.

3 The open economy

We now consider a trade equilibrium between two identical countries. Firms willing to export have to pay a fixed cost f_x, expressed in units of final output Y, and a variable cost of the iceberg type, with $\tau > 1$ units shipped for one unit to arrive abroad.

Given the parameter constraint in Assumption 1, two possible types of equilibria may emerge in the open economy. Both feature a separating equilibrium for skilled workers, since exporting firms are positively selected and therefore the relative demand for high-ability workers of both skill types is higher in the open economy than in autarky. The difference arises in the case of unskilled workers. It is possible that for them the pooling equilibrium from the closed economy survives, but there is another possibility as well: the extra demand from exporters for high-ability unskilled workers may lead to an equilibrium in which high-productivity firms have to pay a higher wage per efficiency unit of unskilled labour than low-productivity firms. In this case, the most able unskilled workers are only hired by high-productivity exporting firms, which have no use for workers of low ability, while the low-ability workers are hired by non-exporting firms, resulting in pooling of unskilled workers over (two) subsets of firms. Within both groups, the wage per efficiency unit is equalised across abilities.

3.1 Firm-level variables and the decision to export

With two identical countries, the gain in operating profits from exporting for each firm is equal to $\tau^{1-\sigma}r(\varphi)/\sigma$, where $r(\varphi)$ is the domestic revenue of a firm with productivity φ. For a firm that is indifferent between exporting and not exporting, this gain in operating profits has to be equal to the fixed export cost f_x:

$$\tau^{1-\sigma}r(\varphi) = \sigma f_x \tag{14}$$

While this indifference condition is of course standard, it is straightforward to see that in our model – different from Melitz (2003) and most of the papers in the literature featuring positive export selection – there cannot be a strict separation of exporters and non-exporters in terms of their baseline productivity. Suppose there were such a sharp separation, and focus on the labour market equilibrium for skilled workers. With a baseline productivity threshold dividing exporting from non-exporting firms, there would have to be an upward jump in the skilled wage profile at the export baseline productivity cutoff, since exporting firms have a discretely higher labour demand than non-exporters, while the ability-specific labour supply is fixed. With the distribution of baseline productivity being continuous, this would entail an upward jump in the marginal cost profile at the cutoff productivity level, which would be incompatible with positive exporter selection in a model such as ours, in which firms treat wages parametrically.[8]

Rather than one baseline productivity threshold, our model has two, the *lower export cutoff* φ_1^x and the *upper export cutoff* φ_2^x. They are the boundaries of a baseline productivity interval with a co-existence of exporters and non-exporters, all of which are indifferent between exporting and non-exporting, and hence – according to Eq. (14) – all of which earn the same domestic revenues $r(\varphi)$. The share of exporters at each baseline productivity level within this interval is implicitly defined by a labour market clearing condition analogous to Eq. (4) for the closed economy, which is given by

$$\frac{l_s(\varphi)}{l_s(1)} = \frac{\varphi^{g-s}}{1+\tau^{1-\sigma}\chi(\varphi)}, \quad \forall\, \varphi \in [\varphi_1^x, \varphi_2^x). \tag{15}$$

There, $l_s(\varphi)$ is the skilled employment used for the output that is sold domestically, and $\chi(\varphi)$ is the fraction of firms exporting among all the firms with baseline productivity φ. We have $\chi(\varphi_1^x) = 0$ and $\chi(\varphi_2^x) = 1$, and, as will be shown below, $\chi'(\varphi) > 0\ \forall\ \varphi \in [\varphi_1^x, \varphi_2^x)$.

The corresponding wage profile follows directly from the fact that in the interval $[\varphi_1^x, \varphi_2^x]$ the firm-specific skilled wage bill $w_s(\varphi)l_s(\varphi)$ is constant, since with Cobb-Douglas production technology it is proportional to firm-specific revenues $r(\varphi)$. Using Eq. (8), we get

$$\frac{w_s(\varphi)}{w_s(1)} = (\varphi_1^x)^{\Phi+g-s}\left[1+\tau^{1-\sigma}\chi(\varphi)\right]\varphi^{s-g}, \quad \forall\, \varphi \in [\varphi_1^x, \varphi_2^x). \tag{16}$$

[8]Discontinuities in the marginal cost profile across firms are compatible with a single export baseline productivity cutoff in models featuring wage setting by firms, as in Helpman et al. (2010) and Egger and Kreickemeier (2012).

The two export cutoffs are implicitly defined by the condition that Eq. (14) holds, given the respective skilled wages

$$w_s(\varphi_1^x) = (\varphi_1^x)^{\Phi}\, w_s(1),$$
$$w_s(\varphi_2^x) = (\varphi_1^x)^{\Phi}\, w_s(1) \left(\frac{\varphi_1^x}{\varphi_2^x}\right)^{g-s} [1+\tau^{1-\sigma}].$$

The employment profile in Eq. (15) and the wage profile in Eq. (16) are not fully determined yet, since $\chi(\varphi)$ is endogenous. However, using the solution for $\chi(\varphi)$ derived below one can show that the employment profile is strictly decreasing in $[\varphi_1^x, \varphi_2^x)$, while the skilled wage profile is increasing with an elasticity larger than Φ. This is intuitive, because similar to other Melitz-type models, more productive firms in our setting have a higher propensity to export, leading to an additional demand for skilled workers with high abilities by firms with higher productivity. In our setting, this drives up wages per efficiency unit of skilled labour and, at the same time, leads to less employment of skilled workers in domestic production in the interval $[\varphi_1^x, \varphi_2^x)$. At baseline productivities outside of that interval the wage profile and the employment profile are increasing with the same constant elasticities as in autarky. For productivities $\varphi < \varphi_1^x$ the respective profiles are given by Eqs. (4) and (8); for productivities $\varphi \geq \varphi_2^x$ they are given by:

$$\left.\begin{aligned} \frac{l_s(\varphi)}{l_s(1)} &= \left(\frac{1}{1+\tau^{1-\sigma}}\right)\varphi^{g-s} \\ \frac{w_s(\varphi)}{w_s(1)} &= \left(\frac{\varphi_1^x}{\varphi_2^x}\right)^{\Phi+g-s}\left(1+\tau^{1-\sigma}\right)\varphi^{\Phi} \end{aligned}\right\} \quad \forall \varphi \geq \varphi_2^x.$$

Figures 1 and 2 show graphical representations of the employment and wage profiles just derived.[9] Both profiles depend on the endogenous variables φ_1^x, φ_2^x and $\chi(\varphi)$, but conditional on those variables, they are independent of the existence of an exporter wage premium for unskilled workers.

In the following we sketch the open economy equilibrium for the more complicated case, in which an exporter wage premium for unskilled workers exists. The case where such a premium is absent follows as a special case, as we will point out. If an exporter wage premium for unskilled workers exists, there is an interval $[\varphi_1^u, \varphi_2^u)$ in which the wage per efficiency unit of unskilled labour increases with constant elasticity, while being constant both below φ_1^u and above φ_2^u:

$$\frac{w_u(\beta)}{w_u(1)} = \begin{cases} \beta^{1-\eta} & \text{if} \quad \beta < \varphi_1^u \\ \left(\frac{\beta}{\varphi_1^u}\right)^{\nu(u+\eta/\nu-s)}\beta^{1-\eta} & \text{if} \quad \beta \in [\varphi_1^u, \varphi_2^u) \\ (\varphi_2^u/\varphi_1^u)^{\nu(u+\eta/n-s)}\,\beta^{1-\eta} & \text{if} \quad \beta \geq \varphi_2^u. \end{cases} \tag{17}$$

Therefore, a firm with productivity $\varphi < \varphi_1^u$ only employs unskilled workers with abilities in the

[9]While linearity below φ_1^x and above φ_2^x follows from the model, linearity between the two export cutoffs is used for convenience, and not implied by the model.

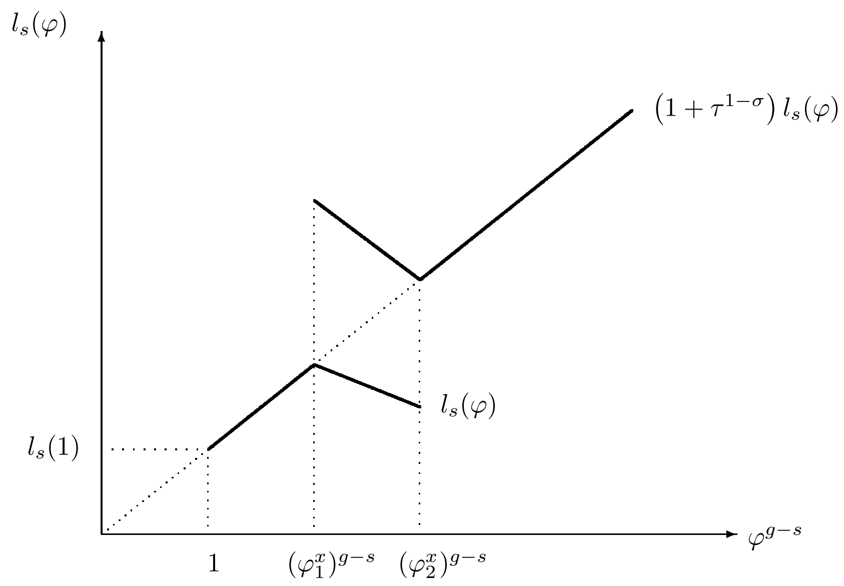


Figure 1: The Employment Profile in the Open Economy

interval $[\underline{\varphi}, \varphi_1^u)$, while a firm with productivity $\varphi \in [\varphi_1^u, \varphi_2^u)$ only employs workers with ability exactly matching its productivity φ. A firm with productivity above φ_2^u is indifferent between hiring unskilled workers with abilities equal to or higher than its baseline productivity. In other words, there is a separating equilibrium for unskilled labour in the interval $[\varphi_1^u, \varphi_2^u)$ and two pooling equilibria, one above and one below this interval. As in the case of skilled workers, the existence of a separating equilibrium for unskilled workers is the result of scarcity of high abilities, which can only materialise in our model if the supply of high-ability unskilled workers is not too large (relative to high-ability skilled workers). Specifically, a necessary condition for an interval $[\varphi_1^u, \varphi_2^u)$ to exist is $u + \eta/\nu - s > 0$.[10]

Both φ_1^u and φ_2^u lie strictly inside the interval $[\varphi_1^x, \varphi_2^x)$, and equipped with this result we can

[10] To get an intuition for the role played by the sign of $u + \eta/\nu - s$, consider the excess demand for unskilled workers with an ability equal to the firm-specific ability threshold, evaluated at an equal wage per efficiency unit for these workers. In the closed economy, the thus defined excess demand is monotonically decreasing in φ, whereas in the open economy it increases over the interval $[\varphi_1^x, \varphi_2^x)$ if $u + \eta/\nu - s > 0$. This condition is necessary but *not* sufficient for a separating equilibrium of unskilled workers over a subdomain of φ, because it is not guaranteed a priori that the large supply of unskilled workers with high abilities is exhausted by the additional demand from exporting firms in the open economy.

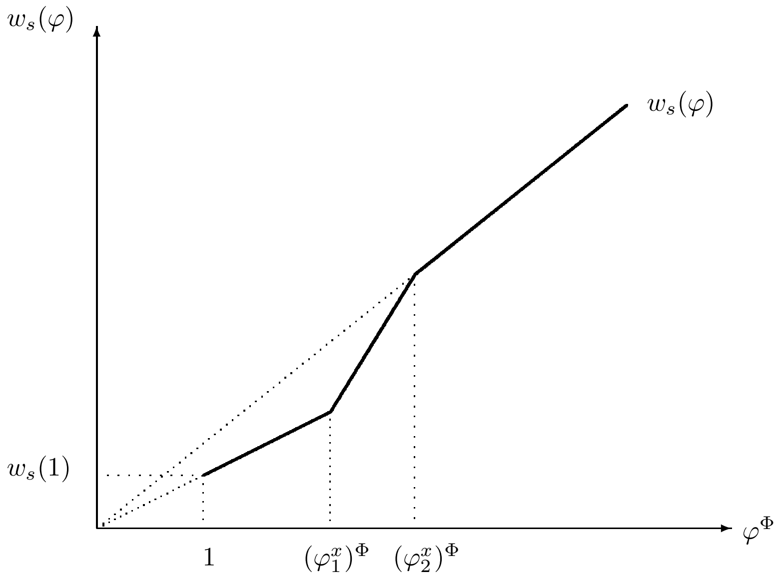


Figure 2: The Skilled Wage Profile in the Open Economy

fully characterise $\chi(\varphi)$ as a function of the cutoff productivities φ_1^u, φ_2^u and φ_1^x:

$$\chi(\varphi) = \begin{cases} 0 & \text{if} \quad \varphi < \varphi_1^x \\ \left[\left(\frac{\varphi}{\varphi_1^x}\right)^{\frac{\Delta[1+\nu(\sigma-1)]}{\nu}} - 1\right]\tau^{\sigma-1} & \text{if} \quad \varphi \in [\varphi_1^x, \varphi_1^u) \\ \left[\left(\frac{\varphi}{\varphi_1^u}\right)^{1+g-\nu s-(1-\nu)u}\left(\frac{\varphi_1^u}{\varphi_1^x}\right)^{\frac{\Delta[1+\nu(\sigma-1)]}{\nu}} - 1\right]\tau^{\sigma-1} & \text{if} \quad \varphi \in [\varphi_1^u, \varphi_2^u) \\ \left[\left(\frac{\varphi}{\varphi_1^x}\right)^{\frac{\Delta[1+\nu(\sigma-1)]}{\nu}}\left(\frac{\varphi_2^u}{\varphi_1^u}\right)^{(1-\nu)[s-u-\eta/\nu]} - 1\right]\tau^{\sigma-1} & \text{if} \quad \varphi \in [\varphi_2^u, \varphi_2^x) \\ 1 & \text{if} \quad \varphi \geq \varphi_2^x \end{cases} \tag{18}$$

It follows from Eq. (18) that $\chi(\varphi)$ increases monotonically in φ over the interval $[\varphi_1^x, \varphi_2^x)$.[11] Since $\chi(\varphi_2) = 1$, we can express the ratio of export cutoffs φ_2^x/φ_1^x as a function of trade costs τ and of $\lambda \equiv \varphi_2^u/\varphi_1^u$:

$$\frac{\varphi_2^x}{\varphi_1^x} = \left(1+\tau^{1-\sigma}\right)^{\frac{\nu}{\Delta[1+\nu(\sigma-1)]}} \lambda^{\frac{\nu(1-\nu)(u+\eta/\nu-s)}{\Delta[1+\nu(\sigma-1)]}}, \tag{19}$$

with $\lambda = 1$ if there is pooling of unskilled workers over the whole range of φ, and $\lambda > 1$, otherwise. Hence λ can be interpreted as a measure for the scarcity of high-ability unskilled workers, with higher values of λ reflecting more scarcity.

[11] It can be checked that under Assumption 1 the exponent $1 + g - \nu s - (1-\nu)u$ is strictly positive.

3.2 Economy-wide variables

Our key variables of interest in the open economy are the exporter wage premia for unskilled and skilled workers. They are defined in our paper in terms of *normalised wages*, which are the wages paid by non-marginal firms relative to the respective wage paid by the marginal firm. Using this concept, we define the exporter wage premium for skilled workers as the percentage increase in the average normalised wage paid by firms above the export cutoff φ_1^x relative to the average normalised wage paid by the same firms in a counterfactual situation in which no firm exports. The exporter wage premium for unskilled workers is defined analogously, the only difference being that the productivity cutoff used is φ_1^u, not φ_1^x, since for unskilled workers this is the threshold above which firms start paying a wage premium. The exporter wage premia thus defined are related to the respective aggregate wage bills by the conditions

$$W_s = L_s \int_1^\infty w_s(\alpha) dS(\alpha) = \frac{s\Theta}{g} L_s w_s(1) \left[1 + (\varphi_1^x)^{-\frac{g}{\Theta}} \omega_s\right] \tag{20}$$

and

$$W_u = L_u \int_1^\infty w_u(\beta) dU(\beta) = \Gamma L_u w_u(1) \left[1 + (\varphi_1^u)^{-\frac{u}{\Gamma}} \omega_u\right], \tag{21}$$

where we can write the exporter wage premium for either skill type as a function of model parameters and a single endogenous variable, λ:

$$\omega_s \equiv \omega_s(\lambda) \qquad \text{and} \qquad \omega_u \equiv \omega_u(\lambda). \tag{22}$$

Whereas the solutions for ω_s and ω_u involve lengthy expressions that are delegated to the Appendix, the impact of λ on the exporter wage premia is intuitive. A larger λ reflects more scarcity of unskilled workers with high abilities, and, hence, results in a larger exporter wage premium for this skill group. With a Cobb-Douglas technology a higher return to unskilled workers implies a lower return to skilled workers, and, hence, the exporter wage premium of skilled workers decreases in λ. For $\lambda = 1$, the expressions in Eq. (22) collapse to

$$\omega_s(1) = \frac{1}{(\Xi - 1)[1 + \nu(\sigma - 1)]} \left[1 - \left(1 + \tau^{1-\sigma}\right)^{-(\Xi-1)}\right] \qquad \text{and} \qquad \omega_u(1) = 0, \tag{23}$$

respectively, with

$$\Xi \equiv \frac{g\nu}{\Delta[1 + \nu(\sigma - 1)]}.$$

The free entry condition in the open economy is given by $R/(N\sigma) = f_e + \chi f_x$, where χ is the overall share of firms that are exporting. The link between average revenues and revenues of the marginal firm can be computed in analogy to the closed economy as

$$\frac{R}{N} = \Theta r(1) \left[1 + (\varphi_1^x)^{-\frac{g}{\Theta}} \omega_s\right] \tag{24}$$

Finally, the exporter indifference condition for the firm with productivity φ_1^x can be written as

$$r(1) = \sigma f_x \tau^{\sigma-1} \left(\varphi_1^x\right)^{-(\sigma-1)\Delta} \tag{25}$$

The free entry condition can be combined with Eqs. (24) and (25) to an implicit function $A(\omega_s, \chi, \varphi_1^x, \tau, f_x) = 0$, linking the three endogenous variables ω_s, χ, and φ_1^x.

The share of exporting firms is computed by integrating over the productivity-specific exporting shares $\chi(\varphi)$:

$$\begin{aligned} \chi &= \int_{\varphi_1^x}^{\varphi_2^x} \chi(\varphi) dG(\varphi) + \int_{\varphi_2^x}^{\infty} dG(\varphi) \\ &= \tau^{\sigma-1}(\varphi_1^x)^{-g} \left\{ \Theta\omega_s + (\Theta - 1) \left[1 - (1 + \tau^{1-\sigma}) \left(\frac{\varphi_2^x}{\varphi_1^x} \right)^{-g} \right] \right\}. \end{aligned} \tag{26}$$

Using Eqs. (19), (22), and (26) to substitute for ω_s and χ in $A(\cdot) = 0$, we arrive at the implicit function

$$\Omega(\varphi_1^x, \lambda; \tau, f_x) = 0, \tag{27}$$

giving combinations between the endogenous variables λ and φ_1^x that are compatible with free entry as well as an indifference about exporting of the marginal exporting firm. As we show in the Appendix, the partial derivatives of $\Omega(\cdot)$ with respect to τ and f_x are positive, and the partial derivatives with respect to φ_1^x and λ are negative. Hence, we have

$$\left.\frac{d\lambda}{d\varphi_1^x}\right|_{\Omega=0} < 0, \qquad \left.\frac{d\varphi_1^x}{d\tau}\right|_{\Omega=0} > 0, \qquad \text{and} \qquad \left.\frac{d\varphi_1^x}{df_x}\right|_{\Omega=0} > 0$$

for parameter values that lead to a solution for $\Omega = 0$ with $\varphi_1^x > 1$ and $\lambda > 1$, which requires that the fixed costs of exporting are sufficiently high:

$$\frac{f_x \tau^{\sigma-1}}{f_e} > \left[1 + (\Theta - 1) \left(1 + \tau^{1-\sigma}\right)^{-(\Xi-1)} \right]^{-1}. \tag{28}$$

For the parameter domain in (28), there exists a well-defined lower productivity bound $\underline{\varphi}_1^x \geq 1$ and a well-defined upper productivity bound $\overline{\varphi}_1^x > 1$, such that $\Omega = 0$ has a solution with $\lambda \geq 1$ if $\varphi_1^x \in [\underline{\varphi}_1^x, \overline{\varphi}_1^x]$.[12] In this case, $\Omega = 0$ establishes a negative relationship between φ_1^x and λ. A higher λ reflects a stronger scarcity of unskilled workers with high abilities and thus higher labour costs for high-productivity firms. Since the least-productive exporters do not have to pay a premium for unskilled workers due to $\varphi_1^u > \varphi_1^x$, they benefit from the now higher labour costs of their high-productivity competitors at home and abroad, and, hence, more firms find it attractive to start exporting, which induces a decline in φ_1^x. Higher costs of exporting increase the minimum productivity required to survive in the export market, leading to a higher φ_1^x.

[12] The lower bound of this interval is equal to 1 if $f_x \tau^{\sigma-1} \leq f_e$, and it is larger than unity otherwise. The upper bound is implicitly determined by $\Omega(\varphi_1^x, 1; \tau, f_x) = 0$.

A second implicit relationship between φ_1^x and λ can be derived from two labour market equilibrium conditions for unskilled workers in the case where there is an exporter wage premium for this group of workers, i.e., if λ is greater than 1. In this case, there is a pooling of unskilled workers with ability $\beta \leq \varphi_1^u$ in firms with baseline productivity $\varphi \in [1, \varphi_1^u)$ and a pooling of unskilled workers with ability $\beta \geq \varphi_2^u$ in firms with baseline productivity $\varphi \in [\varphi_2^u, \infty)$. We show in the Appendix that the condition for both labour market equilibrium conditions to hold at the same time can be expressed as an implicit function $\Phi(\lambda, \varphi_1^x; \tau) = 0$, and the resulting comparative statics are given by

$$\left.\frac{d\lambda}{d\varphi_1^x}\right|_{\Phi=0} < 0 \qquad \text{and} \qquad \left.\frac{d\lambda}{d\tau}\right|_{\Phi=0} < 0.$$

$\Phi(\cdot) = 0$ has a solution in the relevant parameter domain only, if

$$\bar{\lambda} > 1, \tag{29}$$

where

$$\bar{\lambda} \equiv \begin{cases} 0 & \text{if} \quad u + \frac{\eta}{\nu} - s \leq 0 \\ \left[\left(1 + \tau^{1-\sigma}\right)\left(\frac{\Theta\Delta}{g\nu}\frac{u}{\Gamma(u+\eta/\nu-s)}\right)^{-\frac{1}{\Xi-1}}\right]^{\frac{1}{g+1-\nu s-(1-\nu)u}} & \text{if} \quad u + \frac{\eta}{\nu} - s > 0 \end{cases} \tag{30}$$

is defined to cover cases with a small or a large supply of unskilled workers with high abilities. The negative relationship between φ_1^x and λ established by $\Phi = 0$ follows from the observation that a lower propensity to export, i.e., a higher φ_1^x, reduces the scarcity of high-ability unskilled workers and, therefore, reduces λ. Higher variable trade costs reduce the export sales for any given cutoff baseline productivity φ_1^x. This reduces the scarcity of unskilled workers with high abilities, reflected in a decline of λ.

In the Appendix, we show that our model features a unique and stable equilibrium with a selection of firms into exporting under the sufficient condition that

$$\frac{f_x \tau^{\sigma-1}}{f_e} > \left[1 + (\Theta - 1)\left(1 + \tau^{1-\sigma}\right)^{-(\Xi-1)} \max\left\{1, \bar{\lambda}^{-\Xi(1-\nu)(u+\eta/\nu-s)}\right\}\right]^{-1}. \tag{31}$$

The roles played by the various parameter constraints are illustrated in Figure 3. Conditions (28) and (29) imply that both the $\Omega(\cdot) = 0$ locus and the $\Phi(\cdot) = 0$ locus have support in the relevant parameter range. Condition (31) ensures that the $\Omega(\cdot) = 0$ locus lies above the $\Phi(\cdot) = 0$ locus at $\underline{\varphi}_1^x$. This enforces an equilibrium with $\varphi_1^x > 1$ and, thus, one with selection into exporting. The solid curves in Figure 3 represent the implicit functions $\Omega(\cdot) = 0$ and $\Phi(\cdot) = 0$ for a case in which low trade costs establish an equilibrium with $\lambda > 1$, as reflected in the intersection point A. In view of condition (31), it must then be true that the $\Omega(\cdot) = 0$ locus cuts the $\Phi(\cdot) = 0$

locus from above, implying

$$\left.\frac{d\lambda}{d\varphi_1^x}\right|_{\Omega=0} < \left.\frac{d\lambda}{d\varphi_1^x}\right|_{\Phi=0}. \tag{32}$$

This establishes stability, because starting from an off-equilibrium point such as A', (Walrasian) market forces bring the economy to the equilibrium point A.

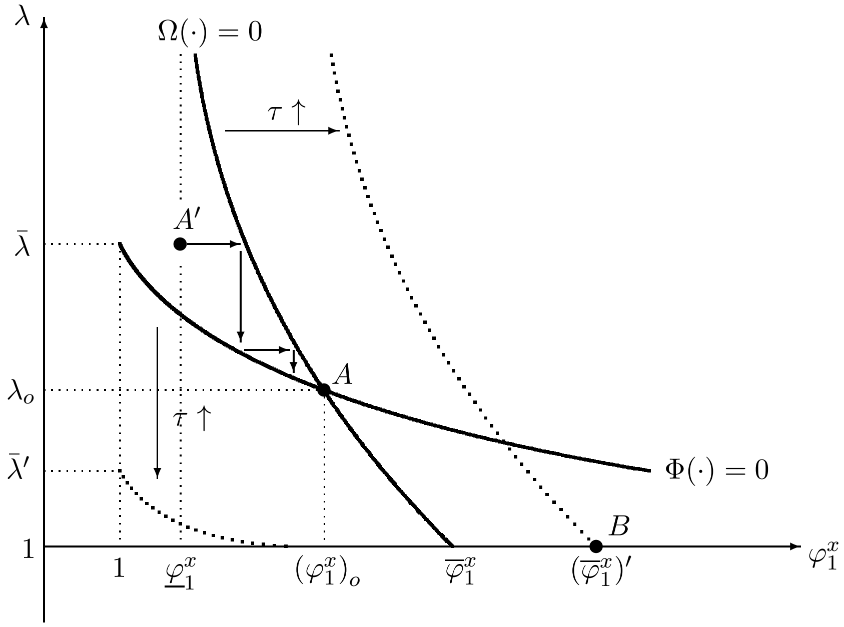


Figure 3: The open economy equilibrium

In Figure 3, we also show that an increase in variable trade costs shifts locus $\Omega(\cdot) = 0$ rightwards and locus $\Phi(\cdot) = 0$ downwards. If the increase in τ is sufficiently large, the economy ends up in an equilibrium such as B, in which high abilities of unskilled workers are no longer scarce and, thus, $\lambda = 1$. In this case, an exporter wage premium for unskilled workers does not exist and the cutoff baseline productivity is given by $\overline{\varphi}_1^x$. Although not explicitly shown, it is easily confirmed from Figure 3 that a further increase in τ would eventually lead to an outcome in which $\Phi(\cdot) = 0$ has no solution for $\varphi_1^x, \lambda \geq 1$.

We round off the discussion in this section by summarising the insights from the analysis above regarding the implications of trade for skilled and unskilled wages.

Proposition 1. *Under the parameter constraint in* (31) *international trade exacerbates the scarcity of high-ability skilled workers, thereby increasing their return relative to low-ability workers and generating an exporter wage premium for this skill group. High ability unskilled workers become scarce in the open economy only if their supply is sufficiently small and trade costs are low. If this is the case, trade will lead to an exporter wage premium also for unskilled workers, but the premium will be smaller than the one for skilled workers.*

Proof. Analysis in the text. □

The wage effects in Proposition 1 are instrumental for the effects of trade on the economy-wide inequality level, which are studied in detail in the next section.

4 Trade and income inequality

To measure income inequality we use the Theil index, which for the distribution of individual labour income w over the interval $[\underline{w}, \overline{w}]$ can be computed as

$$T = \int_{\underline{w}}^{\overline{w}} \frac{w}{\mu_w} \ln \frac{w}{\mu_w} dF(w),$$

where μ_w is average income, and $F(w)$ is the cumulative distribution function of w. A higher value of T is associated with higher inequality, and the Theil index has a minimum value of zero, if income is equally distributed among all workers. A nice feature of the Theil index is its decomposability, which allows us to split the economy-wide inequality into inequality within the groups of skilled and unskilled workers, measured by indices T_s and T_u, respectively, and inequality between the two subgroups, measured by index T_b (see Shorrocks, 1980). For our model, we can write the economy-wide Theil index as

$$T = \nu T_s + (1-\nu)T_u + T_b,$$

with

$$T_b \equiv \nu \ln\left(\frac{(L_u + L_s)\nu}{L_s}\right) + (1-\nu)\ln\left(\frac{(L_u + L_s)(1-\nu)}{L_u}\right).$$

Parameters ν and $1-\nu$ measure the income shares of skilled and unskilled workers, respectively, which due to the Cobb-Douglas technology in Eq. (2) are constant. This implies that access to exporting has an impact on inequality only through its effect on the Theil sub-indices T_s and T_u. These subindices are given by

$$T_u = \frac{L_u}{W_u}\int_1^{\infty} w_u(\beta) \ln\left[\frac{L_u w_u(\beta)}{W_u}\right] dU(\beta) \tag{33}$$

and

$$T_s = \frac{L_s}{W_s}\int_1^{\infty} w_s(\alpha) \ln\left[\frac{L_s}{W_s} w_s(\alpha)\right] dS(\alpha) \tag{34}$$

Based on the Theil index, trade has the following distributional effects:

Proposition 2. *Trade increases the income inequality of skilled workers. The income inequality of unskilled workers remains unaffected, if there is pooling of unskilled workers over all firms, whereas it increases in response to trade, if there is pooling of unskilled workers over subsets of firms. In both cases, the economy-wide inequality is higher in the open than the closed economy.*

Proof. See the Appendix. □

With a pooling of unskilled workers over all firms, the wage profile of unskilled workers is pinned down by the ability of these workers. As a consequence, the Theil index measuring the inequality among unskilled workers is unaffected by trade and given by its autarky level: $T_u = \Gamma - 1 - \ln \Gamma$. In this case, we can focus on T_s when studying the consequences of trade on income inequality. As formally shown in the Appendix, T_s is always larger in the open than in the closed economy, where $T_s = s\Theta/g - 1 - \ln[s\Theta/g]$. This is because trade renders skilled workers with high abilities a scarce resource in our model, which generates an exporter wage premium and augments income inequality within this skill group. With a pooling of unskilled workers over subsets of firms, high-ability workers of both skill types become a scarce resource in the open economy, leading to a wage premium of skilled and unskilled workers employed by exporting firms. The existence of exporter wage premia gives rise to higher inequality within the sub-groups of skilled and unskilled workers, adding up to higher economy-wide income inequality.

We complete the discussion in this section by noting that despite a steeper wage profile for skilled than unskilled workers, it is not clear *a priori* in our model that $T_s > T_u$. The reason is that the Theil index weighs individual wages (relative to the mean) by the relative frequency of workers receiving this wage. Hence, the scarcity of high abilities among skilled workers, which is responsible for larger wage differences among skilled than unskilled workers in the first place, leads to lower weights of high wages in the skilled compared to the unskilled population of workers, thereby counteracting the direct impact of higher wage differences on skill-specific Theil indices.

5 Empirical analysis

In this section, we empirically implement the theoretical model outlined above to estimate the key parameters step by step. For this purpose, we use detailed information on German firms from the linked employer-employee dataset (LIAB), provided by the Institute for Employment Research (IAB).[13] This dataset matches representative firm-level survey information from the IAB establishment panel, which is drawn from a stratified sample of establishments with the stratum being defined over 16 (manufacturing and service) industries, 10 categories of establishment size, and 16 German states.[14] The survey is conducted by Infratest Sozialforschung on behalf of the IAB, yielding a high response rate of about 80 percent. The dataset includes important firm characteristics such as the export status as well as total, domestic, and foreign sales.

LIAB matches the establishment data with individual information on workers who are employed full-time in one of the establishments in this sample as of 30 June in the respective year. The individual data comprise information on educational attainment, distinguishing six levels

[13] Alda et al. (2005) give detailed information on the LIAB dataset, which is confidential but not exclusive, i.e., it is available for non-commercial research by visiting the research data centre of the German Federal Employment Agency at the IAB in Nuremberg, Germany (see also http://fdz.iab.de/en.aspx).

[14] We use the terms firm and establishment synonymously in this section.

of education by the highest degree attained by a worker: 1) lower secondary school, intermediate school leaving certificate or less and no vocational training; 2) lower secondary school, intermediate school leaving certificate or less and vocational training; 3) high-school degree and no vocational training; 4) high-school degree and vocational training; 5) college degree; and 6) university degree. We classify individuals from the first education group as unskilled and individuals with a degree from a college or university as skilled workers (see Südekum, 2008). With regard to the three remaining groups, we follow Klein et al. (2013) and use information on the employment status of a worker to classify her as either skilled (if employed as a white-collar worker) or unskilled (if employed as a blue-collar worker).[15] Our sample covers the years 1996 to 2008, which is the time period before the Subprime Crisis. For this period, we have reliable information on establishments all over Germany. Since our model is not informative about adjustment processes, we do not use the time dimension of the data and instead work with a cross-section of averages by firm.

We drop all establishments that lack sales or employment information. We also drop establishments featuring a value added of less than 10 percent of their sales, because some establishments in the data record negative or other implausible values of value added. This gives a sample of 22,149 establishments, with 7,024 of these establishments ($\approx$ 31.71 percent) being active in the export market.

Descriptive statistics of the establishments are summarised in Table 1. As shown in the table, the average firm in our dataset employs 90.42 workers, with 40.31 percent of its workforce being skilled according to the adopted definition of this group. Not all firms in the data employ both skilled and unskilled workers. The skill premium on wages amounts to 28.81 percent. The wages reported in Table 1 are computed using firm-level averages for the respective skill groups per annum. The sales level of the average firm in the sample is 26 million Euro. There is a substantial heterogeneity in firm size, with sales varying between 5.12 thousand and 19.60 billion Euro, and the workforce varying between 0.04 and 43,225.16 employees.[16]

The descriptive statistics in Table 1 confirm the insight from previous studies that exporters are significantly larger than non-exporters, supporting a selection of the best firms into exporting (cf. Bernard and Jensen, 1995; Bernard et al., 2012). On average, exporters have 127.31 percent higher sales and employ 116.79 percent more workers than the average (exporting or non-exporting) firm in the data. At the same time, exporters in the data display a lower skill-intensity than the average firm. This is in contrast to evidence for Chilean establishments reported by Harrigan and Reshef (2015). However, it is well in line with our theoretical model, which explains a low skill intensity of exporters by the scarcity of high-ability skilled workers. Finally, the descriptives in Table 1 show that the skill premium paid by German exporters is

[15]In a robustness analysis below, we consider alternative definitions of the groups of skilled and unskilled workers. Since wages are right-censored, we follow the literature and impute wages for the best-paid employees in our dataset (see, for instance, Schafer, 1997; Schank et al., 2007). Details on this imputation are available on request.

[16]Fractional values for the number of workers are the result of computing full-year equivalents of the number of days employed and of averaging establishment information over the covered years.

Table 1: Descriptive statistics

	Obs.	Mean	Std. Dev.	Min.	Max.
		– All firms –			
Nr. of employees	22,149	90.42	465.46	0.04	43,225.16
Nr. of unskilled	22,149	53.97	293.36	0.00	25,968.37
Nr. of skilled	22,149	36.45	196.98	0.00	17,256.80
Wages unskilled	18,062	24,550.78	8,205.45	7,821.95	82,754.99
Wages skilled	18,679	31,640.59	12,220.96	7,796.40	95,512.98
Sales (in 1,000)	22,149	26,000	215,000	5.12	19,600,000
		– Exporters –			
Nr. of employees	7,024	196.02	780.92	0.33	43,225.16
Nr. of unskilled	7,024	123.03	493.54	0.00	25,968.37
Nr. of skilled	7,024	72.99	323.63	0.00	17,256.80
Wages unskilled	6,221	27,555.06	7,915.31	8,538.78	72,693.95
Wages skilled	6,674	37,670.19	11,470.78	7,843.38	95,512.98
Sales (in 1,000)	7,024	59,100	363,000	9,742.79	19,600,000

Notes: Figures in this table are based on averages at the firm-level over 1996 to 2008. Wages and sales levels are in real Euro values (base year 2000).

7.83 percentage points (or 27.12 percent) higher than the skill premium paid by the average German firm. This provides further support for our theoretical argument that exporting increases particularly the scarcity of skilled workers with high abilities.

5.1 Estimation strategy and results

For the fully-fledged quantitative model, we need to estimate seven parameters, namely $\{\nu, \sigma, g, s, u, \eta, \tau\}$. It is shown in the following that ν, σ, and τ as well as the composite parameters Δ/g and Φ/g can be estimated directly from firm-level information, using OLS. Combining the estimates of the two composite parameters allows us to recover estimates for g and s as functions of η. Parameters u and η cannot be estimated from firm level information without knowing the realization of $\lambda \geq 1$. Therefore, we use information on unskilled wages in a grid-search routine to estimate parameters u and η in consideration of the general equilibrium constraints imposed by our model. Since the estimation results for u and η do not influence the other parameter estimates, we can conduct our estimation in two steps.[17]

[17] A two-stage estimation procedure similar to ours, with linear regressions in the first stage and GMM in the second stage, has recently been used by Antràs et al. (2017). Also related, Coşar et al. (2016) set several parameters in accordance to earlier empirical studies and then estimate the remaining parameters using GMM.

Step one: OLS estimation of the parameters $\{\nu, \sigma, \tau, \Delta/g, \Phi/g\}$

We first estimate the three parameters ν, σ and τ. Parameter ν measures the cost share of skilled workers in the total wage bill of a firm. We can estimate ν as an average from a regression

$$\frac{w_{s,v} l_{s,v}}{w_{s,v} l_{s,v} + w_{u,v} l_{u,v}} = \nu + error_v. \tag{35}$$

where v is used as a firm index, and subscripts u and s refer to skilled and unskilled workers. This yields an estimate of $\hat{\nu} = 0.466$ with a standard error of 0.003. This estimate is slightly higher than the skilled labour shares for Mexico reported by Feenstra and Hanson (1997), which range between 0.37 and 0.41. In contrast to them, we do not distinguish skilled and unskilled workers by their employment as non-production and production workers, but instead use the level of educational attainment to distinguish between skill groups.

We estimate σ from the ratio of firm-specific total sales, r_v, and operating profits (defined as value-added net of wage bill), ψ_v, from a regression of the form

$$\frac{r_v}{\psi_v} = \sigma + error_v, \tag{36}$$

Using this procedure, we obtain an estimate of $\hat{\sigma} = 8.261$ with a standard error of 2.933. This estimate of σ is somewhat higher than the average estimate reported by Broda and Weinstein (2006), which amounts to 6.5. However, $\hat{\sigma}$ lies within the range of parameter estimates typically found on the basis of gravity equations (cf. Anderson and van Wincoop, 2004, for an overview).

The parameter τ can be estimated according to

$$\frac{r_v^x}{r_v^d} = t_0 + error_v \tag{37}$$

where r_v^d and r_v^x are the domestic and foreign sales of an exporting firm, respectively, and $t_0 \equiv \tau^{1-\sigma}$ is a constant. Under the present assumptions, we obtain an estimate of $\hat{t}_0 = 0.732$ with a standard error of 0.034. In view of the σ-estimate reported above, this corresponds to an average iceberg-transport-cost parameter of $\hat{\tau} = 1.044$. This estimate is lower than the trade-cost estimates typically obtained from gravity equations. For instance, Novy (2013) reports a τ-estimate for Germany in 2000 of 1.69, whereas Milner and McGowan (2013) estimate τ-levels for Germany that range from 1.39 to 1.53 between 1996 and 2004.[18] However, one should keep in mind that when relying on the textbook gravity model in a model with heterogeneous firms, the estimated trade costs comprise both fixed and variable barriers to international trade, and, hence, it is not surprising that our estimate for τ is lower than the overall trade-cost estimates from gravity models.[19]

[18] We would like to thank Chris Milner for providing country-specific estimates of τ.

[19] We cannot rule out that our assumption of symmetric countries leads us to underestimate the size of the export market, thereby leading to a downward bias in the estimate of the iceberg-trade-cost parameter. One way to avoid this potential downward bias in the estimation of τ would be to follow Egger et al. (2013) in formulating a model with potentially asymmetric countries. However, since our dataset has no information about the destination

We now turn to the estimation of composite parameters Δ/g and Φ/g. For this purpose, we adopt an idea outlined by Arkolakis (2010): With Pareto-distributed productivities and CES demand, if relative firm-level revenues are a monotonic function of these firms' productivities (a feature that our model has, along with most other models building on Melitz, 2003), the link of a firm's revenue to its percentile position p in the revenue distribution is informative about the shape parameter of the underlying Pareto productivity distribution. In our model with firm-specific skilled wages, an analogous argument holds for the link between the skilled wage paid by a firm and its percentile position p in the wage distribution.

Since both the revenue distribution and skilled wage distribution across firms are distributed Pareto in their upper tails, i.e., for all lower truncations $\bar{\varphi} > \varphi_2^x$, with shape parameters $(\sigma - 1)\Delta/g$ and Φ/g, respectively, we can therefore learn about these composite parameters from linking revenues and skilled wages of exporters with baseline productivity $\varphi > \bar{\varphi}$ to their percentile positions in the revenue and skilled wage distribution. Denoting by φ_p the baseline productivity at percentile p, we can write $\varphi_p/\bar{\varphi} = (1 - Pr_p/100)^{-1/g}$, where $Pr_p \in \{1, ..., 99\}$ is the percentile position of a firm with baseline productivity φ_p in the productivity distribution of exporters with $\varphi > \bar{\varphi}$, and $rank_p \equiv 1 - Pr_p/100$ is the respective firm's rank in this distribution.[20] Denoting by $\tilde{r}_p$ and $\tilde{w}_{ps}$ average revenues and average skilled wages of firms that are more productive than firm p, we can derive the following two regression equations:

$$\ln \tilde{r}_p = \alpha_0 + \alpha_1 \ln rank_p + error_p, \tag{38}$$

$$\ln \tilde{w}_{ps} = \beta_0 + \beta_1 \ln rank_p + error_p \tag{39}$$

where α_0 and β_0 are constants and $\alpha_1 \equiv -(\sigma - 1)\Delta/g$ and $\beta_1 \equiv -\Phi/g$ are functions of the vector of model parameters $\mathbf{x} \equiv (\nu, \sigma, s, g, \eta)$.[21]

Except for the requirement that firms considered for the estimation of α_1 and β_1 must have a productivity of $\varphi > \varphi_2^x$, there is no further constraint confining the choice of $\bar{\varphi}$. In the absence of a strong prior, we estimate Eq. (38) for different realizations of $\bar{\varphi}$, which is linked to $\bar{\chi}$, the share of exporters with a productivity greater than $\bar{\varphi}$, by $\bar{\chi} \equiv (\bar{\varphi})^{-g}$. The different realizations of $\bar{\varphi}$ are generated by ordering exporters by their revenues, dividing them into subgroups of 100 producers, and then estimating Eq. (38) for different combinations of these subgroups. In particular, we estimate parameter α_1 for the 100 largest exporters, add the next 100 exporters, re-estimate parameter α_1, and continue this process until the set of subgroups of exporters

of exports, we would not be able to disentangle relative market size from the trade-cost parameter itself without adding a new, arbitrary identification assumption. For this reason, and also because the estimation of trade costs is not the main focus of this analysis, we stick to the more parsimonious model variant with symmetric countries and discuss the aptitude of the model to capture important features of the data below.

[20] Since observed ranks of firms in the revenue and skilled wage distribution are positively but not perfectly correlated, we associate $rank_p$ with a firm's rank in the revenue distribution and therefore approximate the productivity distribution by the observed distribution of revenues in our dataset.

[21] Instead of using averages as left-hand side variables, we could simply link each exporter's revenue and skilled wage to its rank in the respective distribution in order to estimate α_1 and β_1. However, working with averages helps guarding against mis-measurement of revenues and wages at the firm level and thus gives more reliable results.

is exhausted. Since our dataset comprises 7,024 exporters, this process gives us 70 different α_1-estimates, which vary over the interval $(-0.815, -0.608)$.

To select a specific parameter estimate, we use the general-equilibrium relationship implied by the model, according to which total revenues of all exporters, R^x, are linked to total revenues of exporters with $\varphi > \bar{\varphi}$, $\bar{R}^x$, by

$$\frac{R^x}{\bar{R}^x} = 1 + \left(\frac{\chi}{\bar{\chi}} - 1\right)\frac{1}{\Theta}. \tag{40}$$

Since R^x, $\bar{R}^x$, χ, and $\bar{\chi}$ are all observable, we can use Eq. (40) to compute 70 values of Θ, based on the same subgroups of exporters used earlier for estimating Eq. (38). Accounting for $\alpha_1 = 1/\Theta - 1$ this implies 70 different values for α_1. All thus computed values for α_1 lie below the interval of estimated values for α_1 from Eq. (38). The largest computed α_1 in Eq. (40) – i.e. the one closest to the interval $(-0.815, -0.608)$ – is generated by $\bar{\chi} = 0.05$, and hence we pick the α_1-estimate from Eq. (38) that is based on the subgroup of largest exporters that make up 5 percent of the total firm population. The corresponding estimate is $\hat{\alpha}_1 = -0.693$ with a standard error of 0.002.[22]

This estimate of α_1 is close to the Pareto shape parameter reported by Arkolakis (2010) for foreign sales of French exporters, which amounts to -0.67 when considering uniform fixed costs and to -0.61 when allowing for fixed costs to be firm-specific. For $\bar{\chi} = 0.050$, we estimate a shape parameter of the skilled wage distribution of $\hat{\beta}_1 = -0.047$ with a standard error of 0.001. The parameter estimates for β_1 lie in the interval $(-0.085, -0.035)$, which suggests that the estimate of β_1 is not too sensitive to the specific choice of $\bar{\chi}$.

With the parameter estimates $\hat{\alpha}_1$ and $\hat{\beta}_1$ at hand, we can in a next step use the definitions of Δ and Φ to solve for estimates of the shape parameters of the baseline productivity distribution, g, and of the ability distribution of skilled workers, s:

$$\hat{g} = -\frac{[\hat{\nu} + (1-\hat{\nu})\eta](\hat{\sigma}-1)}{\hat{\alpha}_1 + \hat{\nu}(\hat{\sigma}-1)\hat{\beta}_1} = 3.975 + 4.546\eta, \tag{41}$$

$$\hat{s} = -\frac{[\hat{\nu} + (1-\hat{\nu})\eta](\hat{\sigma}-1)[1-\hat{\beta}_1+\hat{\alpha}_1]}{\hat{\alpha}_1 + \hat{\nu}(\hat{\sigma}-1)\hat{\beta}_1} = 1.409 + 1.612\eta. \tag{42}$$

The positive impact of η on $\hat{g}$ and $\hat{s}$ is intuitive. All other things equal, a larger parameter η increases the marginal cost differential of two firms resulting from a differential in their baseline productivities, according to Eq. (7). However, this increase in the marginal cost gap is only compatible with a given revenue profile over percentiles and thus a given estimate $\hat{\alpha}_1$, if it is compensated by a lower relative frequency of high productivity producers and thus a lower (higher) average productivity in high (low) percentiles of the revenue distribution of exporters. The reasoning behind the positive relationship of η and s is similar. A larger η allows productivity differences to exert a larger impact on relative skilled wages (because higher wages become a

[22]We have checked that $\bar{\chi} < (\varphi_2^x)^{-g}$ holds for the estimated values of φ_2^x and g, as required by the model.

less important cost factor for firms). However, this is compatible with a given skilled wage profile over percentiles and thus a given estimate of $\hat{\beta}_1$ only, if the higher skilled wage gap is compensated by a lower relative frequency of skilled workers with high abilities. From Eqs. (41) and (42) we can also infer that $\hat{s}/\hat{g} = 1 - \hat{\beta}_1 + \hat{\alpha}_1 = 0.355$ is independent of η.

One could alternatively formulate the least-squares problems to estimate the parameters $\{\nu\}$, $\{\sigma\}$, $\{t_0\}$, $\{\alpha_0, \alpha_1\}$, and $\{\beta_0, \beta_1\}$ in terms of a generalised-method-of-moments (GMM) approach either separately for each parameter type or for all parameters jointly. When allowing for heteroskedasticity, the GMM estimator is fully equivalent to OLS for each equation as above (see Greene, 2003, p. 221).

Step two: Estimation of the two remaining parameters η and u

Together with the parameter restrictions from the model, the estimates from above confine the parameter space of possible combinations of η and u. To see this, note first that condition $s > g + 1 - \eta\sigma$ is equivalent to the constraint

$$\eta > \frac{\hat{\alpha}_1}{\hat{\alpha}_1 + \hat{\beta}_1(\hat{\sigma} - 1)} \equiv \underline{\eta} = 0.669. \tag{43}$$

Furthermore, combining the conditions $u < g+1-\eta\sigma$ and $u+\eta-1 > 0$ establishes $1-\eta < g+1-\eta\sigma$ and, thus, $\eta < \bar{\eta}$, with

$$\bar{\eta} \equiv \begin{cases} -\frac{\hat{\nu}}{1-\hat{\nu}+\hat{\alpha}_1+\hat{\nu}(\hat{\sigma}-1)\hat{\beta}_1} \equiv \bar{\eta}_0 & \text{if} \quad \bar{\eta}_0 \in (0,1) \\ 1 & \text{otherwise} \end{cases}. \tag{44}$$

For the parameter estimates from above, we have $\hat{\eta}_0 < 0$. This establishes $(0.669, 1)$ as the relevant domain of η. For parameter u, the relevant domain is confined by the conditions $u + \eta - 1 > 0$ and $u < g + 1 - \eta\sigma$, and this domain is not empty, if $\eta < \bar{\eta}$. With our parametrisation, we then have $u \in (1 - \eta, 4.975 - 3.715\eta)$.

To select a unique (η, u)-tuple from the possible parameter space, we conduct the following grid-search procedure. We first divide the interval $[\underline{\eta} + 10^{-7}, 1 - 10^{-7}]$ into 100 symmetric subintervals and then pick the bounds of these subintervals as possible parametrisations of η. This gives 101 η-values. For all of these η-values, we then specify a u-interval $[1 - \eta + 10^{-7}, \hat{g} + 1 - \eta\hat{\sigma} - 10^{-7}]$, divide this interval into 100 subintervals, and choose their boundaries as possible parametrisations of u. This gives 101 u-values for each η parameter. Altogether, there are then 10,201 candidate (η, u)-tuples that are supported by our model. For all these parameter tuples, we solve for φ_1^x and λ as described in the theory section, and then compute Theil subindices T_u and $T_w = T - T_b$ as measures of wage inequality for unskilled workers and the within component of the overall wage inequality of skilled and unskilled workers.

Using the definitions

$$t_u(w_u) \equiv \frac{L_u w_u}{W_u} \ln \frac{L_u w_u}{W_u}, \qquad t(w) \equiv \frac{(L_u + L_s)w}{W_u + W_s} \ln \frac{(L_u + L_s)w}{W_u + W_s}, \tag{45}$$

we can express the two Theil subindices as functions of $t_u(w_u)$ and $t(w)$ over the pool of unskilled workers and the pool of all workers, respectively:

$$T_u = \int_{\underline{w}_u}^{\overline{w}_u} t_u(w_u) dF_u(w_u), \qquad T_w = \int_{\underline{w}}^{\overline{w}} t(w) dF(w), \tag{46}$$

where $dF_u(w_u)$ is the relative frequency of an unskilled wage of w_u in the interval $(\underline{w}_u, \overline{w}_u)$, and $dF(w)$ is the relative frequency of any (skilled or unskilled) wage w in the interval $(\underline{w}, \overline{w})$. The Theil subindices T_u and T_w can therefore be interpreted as two data moments that are based on individual wage observations, and we can use them as inputs for a GMM estimation of η and u based on the aforementioned grid.[23]

Whereas targeting two data moments would be sufficient in principle for estimating the two remaining parameters u and η, we face the problem that T_u and T_w are collinear in our model if $\lambda = 1$. This implies that our econometric model would be underidentified if $\lambda = 1$, and, hence, we would not be able to determine a unique combination of η and u in this case. To overcome this problem, we target the average wage of unskilled workers with an ability higher than φ_2^x, denoted $\tilde{w}_u^x$, relative to the average wage of all unskilled workers, W_u/L_u,

$$\frac{L_u \tilde{w}_u^x}{W_u} = \frac{(\varphi_2^x)^{1-\eta} \lambda^{\nu(u+\eta/\nu - s)}}{1 + (\varphi_1^u)^{u+\eta-1} \omega_u}, \tag{47}$$

as an additional data moment. With the solution for the general equilibrium at hand, we can compute theory-consistent measures of $L_u \tilde{w}_u^x / W_u$ for all (η, u)-combinations. Crucially, if $\lambda = 1$ (and, thus, $\omega_u = 0$), Eq. (47) reduces to $L_u \tilde{w}_u^x / W_u = (\varphi_2^x)^{1-\eta}$, and, hence, only depends on η but not on u. This implies that $L_u \tilde{w}_u^x / W_u$ and T_u are not collinear, if $\lambda = 1$. Finally, we can express $L_u \tilde{w}_u^x / W_u$ as the (conditional) mean of observables:

$$\frac{L_u \tilde{w}_u^x}{W_u} = \int_{\underline{w}_u^x}^{\overline{w}_u} \frac{L_u w_u}{W_u} dF_u^x(w_u), \tag{48}$$

where $\underline{w}_u^x$ is the lower bound of wages received by unskilled workers with abilities $\beta \geq \varphi_2^x$ and $dF_u^x(w_u^x)$ is the conditional relative frequency of unskilled wages over interval $(\underline{w}_u^x / \overline{w}_u)$, i.e., $dF_u^x(w_u) = dF_u(w_u)/[1 - F_u(\underline{w}_u^x)]$, with $1 - F_u(\underline{w}_u^x) = (\phi_2^x)^{-u}$.

Taking stock, we have identified three moment conditions in two parameters, which we denote by $m_u(\eta, u)$ for targeting T_u, $m_w(\eta, u)$ for targeting T_w, and $m_r(\eta, u)$ for targeting the average

[23] To be consistent with the previous estimations and with the results of the model, when constructing $t(w)$ we impose the restriction that firms pay the same (average) wage to all their skilled workers.

wage of skilled workers with $\beta > \varphi_2^x$. We select the preferred (η, u)-tuple as the solution to

$$\arg\min_{(\eta,u)} \mathbf{m}'\mathbf{\Psi}\mathbf{m}, \tag{49}$$

where $\mathbf{m}$ is a 3x1 vector of the moment conditions and $\mathbf{\Psi}$ is a 3x3 positive semi-definite moments weighting matrix. To give more precisely measured moments a higher weight in the estimation, we weigh the moment conditions by the inverse variance of the data underlying the construction of the respective averages and estimate parameter values of $\hat{\eta} = 0.987$ and $\hat{u} = 0.052$ with standard errors of 0.001 and 0.005, respectively. These parameter estimates represent interior solutions, i.e. neither $\hat{\eta}$ nor $\hat{u}$ lie at their interval bounds, and they support an outcome with $\lambda = 1$ and, thus, a pooling of unskilled workers over all firms.[24]

To get an intuition for the estimates of η and u, one should bear in mind that for $\lambda = 1$, $\hat{\eta}$ is disciplined by the third moment condition, which links average wages from an upper tail of the ability distribution of unskilled workers in the model to the average wage of unskilled workers employed by high productivity firms in the data. Since the ability distribution in the model does not have an upper bound, a good fit of the model with this specific moment requires that the elasticity of the wage profile of unskilled workers is sufficiently small, so that the impact of ability differences on wages is not too high.[25] The low estimate of u is then the result of bringing a high $\hat{\eta}$ and thus a relatively flat wage profile in accordance with the sizable Theil index of unskilled workers' incomes observed in the data. The parameter estimate of η allows us to compute implied estimates of the shape parameters of the productivity distribution and the ability distribution of skilled workers, which we estimate at $\hat{g} = 8.461$ and $\hat{s} = 3.000$, respectively.

We complete the discussion in this section by summarising the point estimates for our parameters and the main composites of these parameters in Table 2. To account for remaining interdependencies of the parameter estimates, which are not adequately addressed by OLS, we have computed standard errors from 50 bootstrap samples of the data, which differ slightly from the estimated standard errors in the text. We report the bootstrapped standard errors in parentheses.

5.2 Model fit

With the parameter estimates derived in the previous section, we can quantify economy-wide inequality and its components, and contrast the respective results with the empirical patterns in our dataset. Table 3 reports the computed and observed measures of inequality (with bootstrapped standard errors in parentheses). Our model underestimates the various forms of wage

[24]The weighting matrix only matters if there are more moment conditions than parameters to estimate, which is the case in our application if $\lambda > 1$. Hence, with $\lambda = 1$, our parameter estimates do not change when giving equal weight to all moment conditions, or when using the inverse of the estimated variance-covariance matrix of the moment conditions for specifying $\mathbf{\Psi}$, as recommended by Hansen (1982).

[25]Note that in our model $1 - \eta$ captures the effect of abilities on effective firm productivity *conditional on* abilities being higher than the minimum viable level. Therefore our high estimate for η is compatible with the underlying model assumption that abilities are important for determining this productivity.

Table 2: Parameter estimates

parameter values		composite values	
ν	0.466 (0.002)	Δ	0.807 (0.066)
σ	8.261 (2.615)	Φ	0.399 (0.121)
τ	1.044 (0.036)	Θ	3.253 (0.531)
g	8.461 (2.411)	Γ	1.340 (0.022)
s	3.000 (0.993)	Ξ	1.098 (0.151)
η	0.987 (0.031)		
u	0.052 (0.108)		

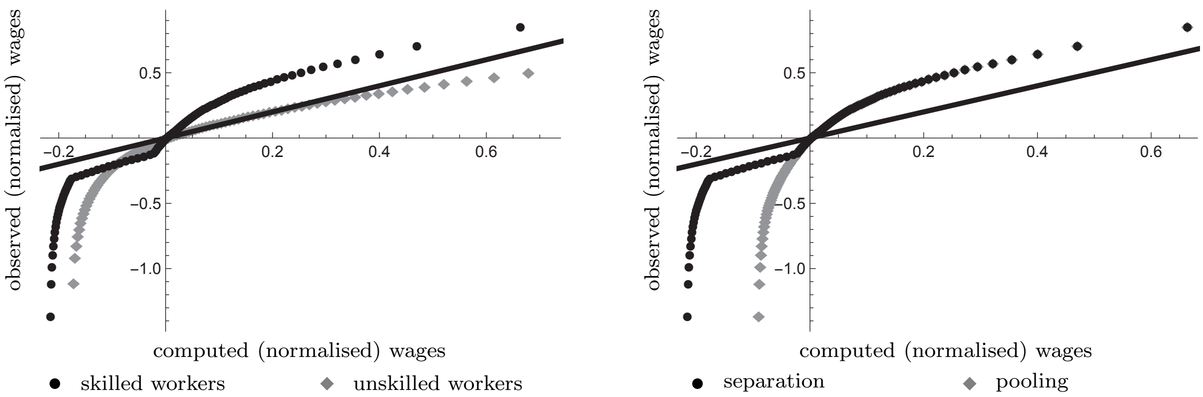
inequality in our dataset, except for the inequality within the group of unskilled workers, which has been targeted in the GMM estimation. Although the downward bias in the T_s is sizable, the model predicts the correct ranking of T_s and T_u. Furthermore, from the inspection of T, we see that despite abstracting from other sources of wage inequality, such as rent-sharing, our model of ability differences of workers can explain a substantial share of overall wage inequality in our dataset.

Table 3: Income inequality

	observed	computed
T_s	0.035	0.017 (0.008)
T_u	0.045	0.047 (0.006)
T_w	0.040	0.033 (0.005)
T_b	0.016	0.008 (0.001)
T	0.056	0.041 (0.005)

To illustrate how the model performs in explaining the observed distribution of wages, we plot observed against computed wages in Figure 4. For illustrative purposes, we have determined averages of 100 wage groups (in logs) and normalised these averages by dividing them by their median value. The observed values come directly from the data and the computed ones are determined by calculating for 100 income groups lower and upper ability bounds, determining the average wages within these bounds using the wage profiles from our model, and dividing the thus determined averages by their median. In Panel A, we show to what extent our model is capable of capturing the wage profile for skilled and unskilled workers in the data. Black dots refer to skilled wages and grey diamonds to unskilled wages, respectively. Clearly, by excluding sources of wage inequality other than ability differences, our model cannot explain all of the within-group variation of individual wages in the data. Accounting for the scarcity of high abilities in the population of skilled workers helps targeting wage differences in the middle of the wage distribution, but it cannot explain the strong variation of wages at the tails of the distribution.

In Panel B, we further elaborate to what extent accounting for the scarcity of high-ability skilled workers improves the aptitude of our model to explain the variation of skilled wages in



Panel A: Skilled and Unskilled Wages

Panel B: Separation and Pooling of Skilled Workers

Figure 4: Observed versus computed wages

the data. For this purpose, we contrast the separating equilibrium supported by the data (black dots) with a pooling equilibrium of skilled workers, featuring an elasticity of the wage profile that is equal to $1 - \eta$. Relying on our parameter estimates, we see that ignoring the scarcity of high abilities would lead our model to underestimate the high variation of skilled wages to a much larger extent than the preferred specification with a separating equilibrium of skilled workers.

Figure 4 suggests that accounting for the the scarcity of skilled workers with high abilities improves the fit of the data by our model. We can go one step further and quantify this improvement by computing the divergence between the observed and computed distribution of skilled wages. For this purpose, we can rely on an index of divergence introduced by Kullback and Leibler (1951), which measures the information loss from approximating the true distribution of wages by the distribution from our model.[26] The Kullback-Leibler divergence gives the expected logarithmic difference between the true probability distribution of variable x, $\widehat{Pr}(x)$, and its theoretical approximation, $Pr(x)$. For discrete distributions it is given by

$$\mathcal{D}(\widehat{Pr}||Pr) = \sum_x \widehat{Pr}(x) \left[\ln \left(\frac{\widehat{Pr}(x)}{Pr(x)} \right) \right], \tag{50}$$

where $\widehat{Pr}(x) = 0$ whenever $Pr(x) = 0$. The Kullback-Leibler divergence takes on a value of zero, if two distributions are identical, and it is larger than zero otherwise. Lower values of $\mathcal{D}(\widehat{Pr}||Pr)$ indicate a smaller information loss and thus a better approximation of the true distribution of x by Pr.

We can now associate x_i with a realization of wages in wage group i, and therefore interpret $\widehat{Pr}(x_i)$ as the probability that a random draw of a skilled worker from our dataset gives an observation in wage group i. Since the 100 wage groups are disjoint and of (almost) equal size,

[26] Mrázova et al. (2016) provide a detailed discussion about the properties of the Kullback-Leibler index and its relationship to Shannon's (1948) entropy.

we have $\widehat{Pr}(x_i) \approx 0.01$ for all $i \in \{1, ..., 100\}$.[27] For constructing the values of $Pr(x_i)$, we consider observed lower and upper bounds of skilled wages for each wage group and construct ability bounds from the wage profile of skilled workers in our model. The thus computed ability bounds are then used to determine $Pr(x_i) = \int_{\underline{\alpha}_i}^{\overline{\alpha}_i} \alpha^{-s} dS(\alpha)$. Using $\widehat{Pr}(x_i)$ and $Pr(x_i)$ in Eq. (50) gives

$$\mathcal{D}_s(\widehat{Pr}||Pr) = 0.001 \qquad \text{and} \qquad \mathcal{D}_p(\widehat{Pr}||Pr) = 2.300 \tag{51}$$

for the baseline specification with a separating equilibrium for skilled workers and a counterfactual situation with a pooling equilibrium of skilled workers, respectively. The sizable difference between the two divergence measures confirms our conjecture from Figure 4 that accounting for the scarcity of high-ability skilled workers improves the fit of our model with the observed distribution of skilled wages considerably.

6 The exporter wage premium and its effect on inequality

When it comes to quantifying the exporter wage premium, the structural approach employed in this paper, unlike previous studies, allows us to compute skill-specific premia. The parameter estimates derived above imply $u + \eta/\nu - s < 0$, which means that there is no exporter wage premium for unskilled workers. By contrast, the exporter premium for skilled workers is large and amounts to 12.137 percent. Weighting the skill-specific wage premia by the share of skilled and unskilled workers employed in exporting firms, we arrive at an average exporter wage premium of 4.896 percent. This average is at the lower bound of results from previous empirical work, which reports estimates that vary between four and eighteen percent (cf. Bernard and Jensen, 1999; Mayer and Ottaviano, 2008; Frias et al., 2009; Egger et al., 2013; Hauptmann and Schmerer, 2013), when not distinguishing between skilled and unskilled workers. Using the same dataset as we do, Klein et al. (2013) estimate an average exporter wage premium for Germany of 4.4 percent in a reduced-form approach.[28]

To provide a better understanding of how the existence of an exporter wage premium affects economy-wide wage inequality in our model, we can quantify the impact of trade on wage inequality in Germany by contrasting the Theil indices for the observed exposure to trade with the respective indices in the counterfactual situation without any trade. Moving the German economy to autarky means eliminating the exporter wage premium by design. Since exporters do not pay a wage premium to their unskilled workforce, this comparative static exercise only affects the wage distribution of skilled workers, making it more equal. Specifically, the Theil index measuring the wage distribution of skilled wages, T_s, declines by 35.975 percent, whereas

[27] Due to indivisibilities, small deviations from $\widehat{Pr}(x_i) = 0.01$ are possible.

[28] Schank et al. (2007) report much smaller exporter wage premia for Germany. According to their findings, plants with an export to sales ratio of 60 percent pay 1.8 percent higher wages to unskilled and 0.9 percent higher wages to skilled workers than otherwise identical non-exporting plants.

the overall wage inequality, as measured by the Theil index T, falls by 6.783 percent.[29]

As an alternative counterfactual exercise, we consider the case of abolishing all variable trade costs, while leaving the fixed export costs constant. At the observed exposure to trade, $\chi = 0.317$, we can compute an estimate of f_x/f_e by employing the parameter estimates in Eq. (27). This establishes $\widehat{f_x/f_e} = 0.414$. Equipped with this fixed cost ratio, we can then apply the implicit function in Eq. (27) to compute the share of exporting firms if τ falls to unity. This gives $\chi = 0.466$ and indicates that eliminating iceberg trade costs has a considerable effect on the share of exporting firms. Regarding the effect on inequality, there are now two potentially counteracting effects: An increase in the exporter wage premium, which increases inequality, and an increase in the share of the skilled worker population being paid this premium. The latter effect is non-monotonic, becoming negative if χ is sufficiently large. In line with this reasoning, the distributional effects of our counterfactual exercise are small. The exporter wage premium for skilled workers would increase by 2.720 percent and would be counteracted and dominated by an increase in the share of skilled workers employed by exporters, so that wage inequality within the group of skilled workers would be reduced if variable trade costs fell to zero. Specifically, we find that the Theil index of skilled workers falls by 1.729 percent, whereas the Theil index for all workers falls by 0.326 percent.

7 Robustness analysis

In this section, we analyse the robustness of our estimation results regarding the distinction of skilled and unskilled workers. For this purpose, we consider two alternative definitions of skill groups: a narrow definition, in which we associate the skilled with those workers who hold a college or university degree; and a broad definition, in which we classify only workers as unskilled if their highest degree attained is one of a lower secondary education and if they lack vocational training. The parameter estimates for these two alternative skill definitions are summarised in Table 4. To facilitate the discussion, we have added our estimates from the benchmark scenario, whereas we do not report the estimates for σ, τ, and Θ, as they are independent of the chosen skill definitions.

From the ν-estimates in Table 4, we can conclude that the considered changes in the skill definitions have large effects on the income shares of the two skill groups. However, the effects on the other parameter estimates are rather small and, relying on the bootstrapped standard errors from the benchmark scenario, the differences in most cases are statistically insignificant. Furthermore, where a direct economic interpretation of the parameters is applicable, the observed changes in the point estimates seem to be well in line with standard economic reasoning.

For instance, classifying as skilled only those workers who hold a college or university degree makes abilities of skilled workers more similar, and therefore leads to a higher estimate of s,

[29]With constant expenditure shares due to the Cobb-Douglas technology and full employment of both skill types, there are no effects on the relative return of skilled to unskilled workers on average, leaving the between component of the Theil index, T_b, unaffected.

Table 4: Parameter estimates for alternative skill definitions

parameter	benchmark	narrow skill	broad skill
ν	0.466	0.106	0.841
g	8.461	9.914	7.438
s	3.000	3.480	2.632
η	0.987	0.991	0.997
u	0.052	0.037	0.013
Δ	0.807	0.946	0.709
Φ	0.399	0.432	0.345
Γ	1.340	1.346	1.326
Ξ	1.098	0.630	1.241

whose value approaches infinity in the limiting case of fully symmetric agents. Of course, one would also expect a higher elasticity of the skilled wage schedule when skilled workers with high abilities are a scarcer resource, i.e., when s is higher. This reasoning is in line with a higher estimate of Φ for the scenario with a narrow skill definition. Finally, since the markup per efficiency unit paid to skilled workers with higher abilities lowers the effect that differences in the baseline productivity exert on marginal costs, one would expect that the elasticity of firm-level marginal costs with respect to productivity declines, when skilled workers become a less important cost factor as fewer of them are employed by firms. This conjecture is confirmed by a higher estimate of Δ when considering the narrow skill definition.

Choosing alternative definitions of skill groups also affects the observed measures of wage inequality. This can be seen from Table 5, where we report the various Theil indices discussed above for the benchmark scenario as well as the two alternative skill definitions. As expected, the broader the definition of a skill group is, the larger is the wage inequality in this group. Interestingly, we observe for both the narrower as well as the broader skill definition that the share of wage inequality attributed to the within component increases in the data. Our model is not overly successful in capturing the between-component of the Theil index, but it does a fairly good job in predicting wage inequality within the subgroups of skilled and unskilled workers, in particular when considering the narrow skill definition.

Table 5: Theil indices for alternative skill definitions

	benchmark		narrow skill def.		broad skill def.	
	observed	computed	observed	computed	observed	computed
T_s	0.035	0.017	0.029	0.028	0.040	0.014
T_u	0.045	0.047	0.058	0.049	0.038	0.044
T_w	0.040	0.033	0.053	0.047	0.040	0.018
T_b	0.016	0.008	0.016	0.001	0.007	0.044
T	0.056	0.041	0.069	0.048	0.047	0.063

With the parameter estimates from Table 4 at hand, we can finally compute the exporter wage premium. For skilled workers this premium amounts to 34.367 percent when considering a narrow skill definition, which is almost three times as high as in the benchmark scenario. Since $u + \eta/\nu - s > 0$ holds, the necessary condition for an exporter wage premium for unskilled workers is met, but we estimate $\bar{\lambda} < 1$, which enforces $\lambda = 1$ and thus $\omega_u = 0$. The average exporter wage premium for all skilled and unskilled workers amounts to 4.520 percent for the narrow skill definition, which is of the same order of magnitude as in the benchmark scenario. Relying on the broad skill definition, we estimate an exporter wage premium for skilled workers of 7.244 percent. Whereas $u+\eta/\nu-s < 0$ does not leave scope for an exporter wage premium of unskilled workers, the average exporter wage premium for skilled and unskilled workers amounts to 4.930. This confirms the insight from above that the estimate of the average exporter wage premium is robust to changes in the definition of skill groups.

8 Conclusion

We have proposed a structural empirical model of an open economy in which the populations of firms and workers are heterogeneous. Firms hire skilled and unskilled workers for production, and firms with a higher (innate) baseline productivity require skilled and unskilled workers of higher ability, leading them to pay higher wages for both types of workers. With more productive firms selecting into exporting, our model allows for the existence of an exporter wage premium due to extra demand by exporting firms for high-ability workers that are in fixed supply. Whether or not an exporter wage premium materialises in our model depends on the supply of and the demand for workers with high abilities, and, therefore, on the ability distributions of skilled and unskilled workers as well as on the baseline productivity distribution of firms. With two types of workers present, we can decompose the exporter wage premium into its skill-specific components.

We structurally estimate the parameters of the model using the matched employer-employee dataset LIAB of the Institute for Employment Reserch (IAB), which provides detailed information on German firms and workers for the years 1996 to 2008. The dataset provides information on the educational background of workers as well as their employment as blue- or white-collar workers. We use this information to find the empirical equivalent of the skilled and unskilled worker categories in our theoretical model. Using the estimated parameter values, we find an average exporter wage premium of about 5 percent, which is within the range of values reported in other studies. But we also find that this average masks a lot of heterogeneity, with exporting firms paying no wage premium to their unskilled workers, while the exporter wage premium for the subgroup of skilled workers is higher than 12 percent. This suggests to us that it is empirically relevant to dissect the exporter wage premium as we have done in this paper.

Through its effect on the exporter wage premium, trade in the proposed model has considerable effects on the economy-wide distribution of income, which is more unequal in the open than the closed economy. However, the respective effects are not monotonic, implying that traveling further along the road of trade liberalisation will eventually reduce income inequality, despite an increase in the exporter wage premium. The reason for this is that less productive firms start exporting, employing workers with lower abilities and thus from lower segments of the wage distribution. In two extensions, we show that the main insights from our analysis are robust to quite substantial changes in the definitions of skilled and unskilled workers.

References

Alda, H., S. Bender, and H. Gartner (2005): "The Linked Employer-employee Dataset Created from the IAB Establishment Panel and the Process-produced Data of the IAB (LIAB)," *Schmollers Jahrbuch*, 125, 327–336.

Amiti, M. and D. R. Davis (2012): "Trade, Firms, and Wages: Theory and Evidence," *Review of Economic Studies*, 79, 1–36.

Anderson, J. E. and E. van Wincoop (2004): "Trade Costs," *Journal of Economic Literature*, 42, 691–751.

Antràs, P., T. Fort, and F. Tintelnot (2017): "The Margins of Global Sourcing: Theory and Evidence from U.S. Firms," *American Economic Review*, forthcoming.

Arkolakis, C. (2010): "Market Penetration Costs and the New Consumers Margin in International Trade," *Journal of Political Economy*, 118, 1151–1199.

Armenter, R. and M. Koren (2015): "Economies of Scale and the Size of Exporters," *Journal of the European Economic Association*, 13, 482–511.

Bernard, A. B. and J. B. Jensen (1995): "Exporters, Jobs, and Wages in U.S. Manufacturing: 1976-1987," *Brookings Papers on Economic Activity. Microeconomics*, 67–119.

——— (1999): "Exceptional Exporter Performance: Cause, Effect, or Both?" *Journal of International Economics*, 47, 1–25.

Bernard, A. B., J. B. Jensen, S. J. Redding, and P. K. Schott (2012): "The Empirics of Firm Heterogeneity and International Trade," *Annual Review of Economics*, 4, 283–313.

Broda, C. and D. E. Weinstein (2006): "Globalization and the Gains from Variety," *Quarterly Journal of Economics*, 121, 541–585.

Coşar, A. K., N. Guner, and J. Tybout (2016): "Firm Dynamics, Job Turnover, and Wage Distributions in an Open Economy." *American Economic Review*, 106, 625–663.

Davis, D. R. and J. Harrigan (2011): "Good Jobs, Bad Jobs, and Trade Liberalization," *Journal of International Economics*, 84, 26–36.

Egger, H., P. Egger, and U. Kreickemeier (2013): "Trade, Wages, and Profits," *European Economic Review*, 64, 332 – 350.

Egger, H. and U. Kreickemeier (2009): "Firm Heterogeneity And The Labor Market Effects Of Trade Liberalization," *International Economic Review*, 50, 187–216.

——— (2012): "Fairness, Trade, and Inequality," *Journal of International Economics*, 86, 184–196.

Feenstra, R. C. and G. H. Hanson (1997): "Foreign Direct Investment and Relative Wages: Evidence from Mexico's Maquiladoras," *Journal of International Economics*, 42, 371–393.

Frias, J. A., D. S. Kaplan, and E. Verhoogen (2009): "Exports and Wage Premia: Evidence from Mexican Employer-Employee Data," unpublished manuscript.

Ghironi, F. and M. J. Melitz (2005): "International Trade and Macroeconomic Dynamics with Heterogeneous Firms," *Quarterly Journal of Economics*, 120, 865–915.

Greene, W. (2003): *Econometric Analysis*, Prentice Hall, New Jersey, 5th ed.

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

Harrigan, J. and A. Reshef (2015): "Skill Biased Heterogeneous Firms, Trade Liberalization and the Skill Premium," *Canadian Journal of Economics*, 48, 1024–1066.

Hauptmann, A. and H.-J. Schmerer (2013): "Do Exporters Pay Fair-wage Premiums?" *Economics Letters*, 121, 179–182.

Helpman, E., O. Itskhoki, and S. Redding (2010): "Inequality and Unemployment in a Global Economy," *Econometrica*, 78, 1239–1283.

Irrazabal, A., A. Moxnes, and K. H. Ulltveit-Moe (2013): "Heterogeneous Firms or Heterogeneous Workers? Implications for Exporter Premiums and the Gains from Trade," *Review of Economics and Statistics*, 95, 839–849.

Klein, M., C. Moser, and D. Urban (2013): "Exporting, Skills and Wage Inequality," *Labour Economics*, 25, 76–85.

Kullback, S. and R. A. Leibler (1951): "On Information and Sufficiency," *The Annals of Mathematical Statistics*, 22, 79–86.

Mayer, T. and G. I. Ottaviano (2008): "The Happy Few: The Internationalisation of European Firms – New Facts based on Firm-level Evidence," *Intereconomics*, 44, 149–158.

Melitz, M. J. (2003): "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71, 1695–1725.

Milner, C. and D. McGowan (2013): "Trade Costs and Trade Composition," *Economic Inquiry*, 51, 1886–1902.

Mrázova, M., P. Neary, and M. Parenti (2016): "Sales and Markup Dispersion: Theory and Empirics," *Unpublished Manuscript*.

Novy, D. (2013): "Gravity Redux: Measuring International Trade Costs with Panel Data," *Economic Inquiry*, 51, 101–121.

Sampson, T. (2014): "Selection into Trade and Wage Inequality," *American Economic Journal: Microeconomics*, 6, 157–202.

Schafer, J. (1997): *Analysis of Incomplete Multivariate Data*, Chapman and Hall/CRC.

Schank, T., C. Schnabel, and J. Wagner (2007): "Do Exporters Really Pay Higher Wages? First Evidence from German Linked Employer-Employee Data," *Journal of International Economics*, 72, 52–74.

Shannon, C. E. (1948): "A Mathematical Theory of Communication," *Bell System Technical Journal*, 27, 379–423.

Shorrocks, A. F. (1980): "The Class of Additively Decomposable Inequality Measures," *Econometrica*, 48, 613–625.

Südekum, J. (2008): "Convergence of the Skill Composition Across German Regions," *Regional Science and Urban Economics*, 38.

A Appendix

A.1 Proof of Lemma 1

To prove Lemma 1, we proceed in two steps. First, we show that with pooling of unskilled workers over all firms, $s \geq g + 1 - \eta\sigma$ is necessary and sufficient for segmentation of skilled workers by firms. Second, we show that with segmentation of skilled workers by firms $u < g + 1 - \eta\sigma$ is sufficient for pooling of unskilled workers by firms. Together, the two conditions then establish Lemma 1.

Let us first consider pooling of unskilled workers over all firms. In this case, wage differences reflect ability differences, and the wage paid to a worker with ability β are linked to the wage paid to the least able unskilled worker, with ability β_c, who finds employment in the least productive firm, featuring productivity φ_c, by $w_u(\beta) = (\beta/\beta_c)^{1-\eta} w_u(\beta_c)$. Let us further assume that skilled workers were also just compensated for their ability differences, implying that skilled wages would be given by $w_s(\alpha) = (\alpha/\alpha_c)^{1-\eta} w_s(\alpha_c)$, with $\alpha_c = \varphi_c$. In this case, our model would mimic a textbook Melitz (2003) model with the productivity of firms given by φ^η and, since all producers pay the same wage per efficiency unit for labour input, marginal costs are inversely proportional to to this productivity. If firms in such a situation were only employing those skilled workers that exactly match their productivity threshold, i.e., if $\alpha = \varphi$ would hold for all φ, the assumption of a Cobb-Douglas technology would give rise to $w_s(v)l_s(v) = \nu c(v)q(v)$. Accordingly, firm-level demand for skilled workers could be expressed by $l_s(v) = (\varphi(v)/\varphi_c)^{\eta\sigma-1} l_s^c$, where l_s^c and φ_c refer to the number of skilled workers employed by and the baseline productivity of the least productive producer, respectively. Suppressing firm index v, total labour demand of all firms with productivity higher than $\bar{\varphi}$ would be given by $N l_s^c [g/(g - \eta\sigma + 1)](\bar{\varphi}/\varphi^c)^{\eta\sigma - g - 1}$, where N is the mass of producers. The supply of workers with ability higher than $\bar{\varphi}$ equals $L_s \bar{\varphi}^{-s}$, and labour market clearing at $\bar{\varphi} = \varphi_c$ therefore requires $N l_s^c [g/(g - \eta\sigma + 1)] = L_s \varphi_c^{-s}$. Putting together, we find that under labour market clearing $g + 1 - \eta\sigma > s$ would lead to excess supply of skilled workers with ability $\beta > \varphi_c$ in the knife-edge case, in which firms employ skilled workers whose ability corresponds to their productivity threshold and pay the same wage per efficiency unit for either skill group. This would result in firms hiring workers with abilities above their productivity threshold and thus establish pooling of skilled workers over firms. In contrast, $s > g + 1 - \eta\sigma$ would lead to excess demand of skilled workers in the thus described knife-edge case and ultimately lead to a separating equilibrium and a skilled wage profile that is steeper than required for a compensation of ability differences.

Let us now consider segmentation of skilled workers by firms and let us hypothesise that firms hire only unskilled workers that exactly match their technology requirements, $\beta = \varphi$, paying them a wage of $w_u(\beta) = (\beta/\beta_c)^{1-\eta} w_u(\beta_c)$. Using Eqs. (5) and (9) from Section 2.3, we can write firm-level employment of unskilled workers as $l_u(\varphi) = (\varphi/\varphi_c)^{\Delta(\sigma-1)-1+\eta} l_u(\varphi_c)$. The demand for unskilled workers of firms with productivity $\varphi > \bar{\varphi}$ is given by $N l_u^c [g/(g - \Delta(\sigma - 1) - 1 + \eta)](\bar{\varphi}/\varphi^c)^{\Delta(\sigma-1)-g+1-\eta}$, whereas supply of unskilled workers with ability higher than $\bar{\varphi}$ equals $L_u \bar{\varphi}^{-u}$. With labour market clearing at $\bar{\varphi} = \varphi_c$, we find that there is an excess supply of unskilled workers in the knife-edge case considered here, if $u < g - \Delta(\sigma - 1) + 1 - \eta$. In this case, we have pooling of unskilled workers over all firms, whereas there is segmentation of unskilled workers over firms if $u > g - \Delta(\sigma - 1) + 1 - \eta$. Noting further that $\Delta - \eta \leq 0$ is established by $s \geq g + 1 - \eta\sigma$, one can confirm that the parameter constraint in Assumption 1 establishes pooling of unskilled workers if there is segmentation of skilled workers by firms. Putting together, we can thus conclude that with this parameter constraint the only possible outcome consistent with labour market clearing is one with segmentation of skilled workers by firms and pooling of unskilled workers over firms. This completes the proof.

A.2 Wage profiles of skilled and unskilled workers in the open economy

Let us consider the productivity thresholds φ_1^x, φ_2^x as defined in the main text. Then, for all $\varphi \in [\varphi_1^x, \varphi_2^x)$, exporters and non-exporters coexist, and, hence, it must be true that $(1 + \tau^{1-\sigma})r(\varphi)/\sigma - r(\varphi)/\sigma = f_x$ and thus $\tau^{1-\sigma}r(\varphi) = \sigma f_x$. Accounting for $r(\varphi) = p(\varphi)q(\varphi)$ and Eq. (8) from the closed economy, we obtain the revenue profile for the open economy:

$$r(\varphi) = \begin{cases} \varphi^{(\sigma-1)\Delta} r(1) & \text{if} \quad \varphi < \varphi_1^x \\ (\varphi_1^x)^{(\sigma-1)\Delta} r(1) & \text{if} \quad \varphi \in [\varphi_1^x, \varphi_2^x) \\ \varphi^{(\sigma-1)\Delta} \left(\frac{\varphi_1^x}{\varphi_2^x}\right)^{(\sigma-1)\Delta} r(1) & \text{if} \quad \varphi \geq \varphi_2^x \end{cases} \tag{A.1}$$

Furthermore, the wage profile of skilled workers can be expressed as

$$\frac{w_s(\varphi)}{w_s(1)} = \begin{cases} \varphi^{-\frac{g}{\Theta}+s} & \text{if} \quad \varphi < \varphi_1^x \\ \varphi^{-g+s}[1 + \tau^{1-\sigma}\chi(\varphi)](\varphi_1^x)^{g\frac{\Theta-1}{\Theta}} & \text{if} \quad \varphi \in [\varphi_1^x, \varphi_2^x) \\ \varphi^{-\frac{g}{\Theta}+s}[1 + \tau^{1-\sigma}] \left(\frac{\varphi_1^x}{\varphi_2^x}\right)^{g\frac{\Theta-1}{\Theta}} & \text{if} \quad \varphi \geq \varphi_2^x \end{cases} \tag{A.2}$$

where we have used Eq. (8) together with $[\nu + (1-\nu)\eta]/\nu - \Delta/\nu = -g/\Theta + s$, and $\chi(\varphi) = 0$ if $\varphi < \varphi_1^x$ to establish the first line in Eq. (A.2). The second line is obtained from our previous result that $r(\varphi) = r(\varphi_1^x)$ holds for all $\varphi \in [\varphi_1^x, \varphi_2^x)$ and the insight that a fraction $\chi(\varphi)$ of firms of productivity φ exports over productivity interval $[\varphi_1^x, \varphi_2^x)$. Skilled labour demand of firms with productivity $\varphi \in (\varphi_1^x, \varphi_2^x]$ is therefore proportional to $[1+\tau^{1-\sigma}\chi(\varphi)]w_s(\varphi)$, and these firms exist with density $g\varphi^{-g-1}$, whereas skilled workers with ability $\alpha = \varphi$ exist with density $s\varphi^{-s-1}$. Equalising supply and demand for all $\varphi \in [\varphi_1^x, \varphi_2^x)$ therefore requires a wage profile over this productivity subdomain as given by the second line of (A.2). Finally, the third line is derived in analogy to the first line, accounting for $\chi(\varphi) = 1$ if $\varphi \geq \varphi_2^x$.

To determine the wage profile of unskilled workers in Eq. (17), let us assume that there exists an interval $[\varphi_1^u, \varphi_2^u)$, such that there is segmentation of unskilled workers by firms over this productivity interval. Under Assumption 1 and a Cobb-Douglas technology, it is immediate that $\varphi_1^u, \varphi_2^u \in (\varphi_1^x, \varphi_2^x)$. From the discussion of the closed economy, it follows that $w_u(\beta)/w_u(1) = \beta^{1-\eta}$ holds for all $\beta < \varphi_1^u$, establishing the first line in Eq. (17). With segmentation of skilled and unskilled workers by firms, the marginal cost of a firm with productivity $\varphi \in [\varphi_1^u, \varphi_2^u)$ is given by

$$c(\varphi) = \mu \frac{w_u(\varphi)}{\varphi} \left(\frac{w_s(\varphi)}{w_u(\varphi)} \right)^{\nu}, \tag{A.3}$$

where $\beta = \varphi$ and thus $w_u(\beta) = w_u(\varphi)$ have been considered. From the assumption of a Cobb-Douglas technology paired with the requirement that labour markets of skilled and unskilled workers clear for all $\varphi \in [\varphi_1^u, \varphi_2^u)$, we further have

$$\frac{\nu}{1-\nu} = \frac{L_s s \varphi^{-s-1} w_s(\varphi)}{L_u u \varphi^{-u-1} w_u(\varphi)}, \tag{A.4}$$

implying $c(\varphi) \propto w_u(\varphi)\varphi^{-1-\nu(u-s)}$. This establishes $w_u(\beta)/w_u(\varphi_1^u) = (\varphi/\varphi_1^u)^{1+\nu(u-s)}$ and thus the second line in Eq. (17). For the third line, we can simply note that $w_u(\beta)/w_u(\varphi_2^u) = (\beta/\varphi_2^u)^{1-\eta}$ holds for all $\beta > \varphi_2^u$, whereas $w_u(\varphi_2^u)/w_u(1) = (\varphi_2^u/\varphi_1^u)^{\nu(u+\eta/\nu-s)}(\varphi_2^u)^{1-\eta}$ follows from above. This completes the formal characterisation of the two wage profiles.

A.3 Derivation of Eq. (18)

Provided that $[\varphi_1^u, \varphi_2^u)$ is not empty, we have to distinguish five productivity domains. The first one is $\varphi < \varphi_1^x$ and implies $\chi(\varphi) = 0$. The second one is $\varphi \geq \varphi_2^x$ and establishes $\chi(\varphi) = 1$. The third productivity domain is given by $\varphi \in [\varphi_1^x, \varphi_1^u)$. In this case, we have segmentation of skilled workers and pooling of unskilled workers, establishing $c(\varphi) \propto w_s(\varphi)^{\nu}\varphi^{-\nu-\eta(1-\nu)}$. Since we know from above that $r(\varphi) = r(\varphi_1^x)$ an thus $c(\varphi) = c(\varphi_1^x)$ holds for all $\varphi \in [\varphi_1^x, \varphi_2^x)$, we can solve for $w_s(\varphi)/w_s(\varphi_1^x) = (\varphi/\varphi_1^x)^{1+\eta(1-\nu)/\nu}$. Equating the latter with $w_s(\varphi)/w_s(\varphi_1^x) = (\varphi/\varphi_1^x)^{-g+s}[1+\tau^{1-\sigma}\chi(\varphi)]$ from the second line of Eq. (A.2) establishes the second line in Eq. (18).

The fourth segment, we have to consider is one with $\varphi \in [\varphi_1^u, \varphi_2^u)$. In this case, we have segmentation of skilled and unskilled workers by firms and thus $c(\varphi) \propto w_u(\varphi)\varphi^{-1-\nu(u-s)}$ (see above) or, equivalently, $c(\varphi) \propto w_s(\varphi)\varphi^{-1-(1-\nu)(s-u)}$. Since we have $c(\varphi) = c(\varphi_1^u)$, we can solve for $w_s(\varphi)/w_s(\varphi_1^u) = (\varphi/\varphi_1^u)^{1+(1-\nu)(s-u)}$. Accounting for $w_s(\varphi_1^u)/w_s(\varphi_1^x) = (\varphi_1^u/\varphi_1^x)^{1+\eta(1-\nu)/\nu}$, we obtain

$$\frac{w_s(\varphi)}{w_s(\varphi_1^x)} = \left(\frac{\varphi_1^u}{\varphi_1^x}\right)^{\frac{\nu+\eta(1-\nu)}{\nu}} \left(\frac{\varphi}{\varphi_1^u}\right)^{1+(1-\nu)(s-u)} \tag{A.5}$$

Equating Eq. (A.5) with $w_s(\varphi)/w_s(\varphi_1^x) = (\varphi/\varphi_1^x)^{-g+s}[1+\tau^{1-\sigma}\chi(\varphi)]$ and accounting for $1 + \eta(1-\nu)/\nu + g - s = \Delta[1+\nu(\sigma-1)]/\nu$ gives the third line in Eq. (18). The fifth productivity segment corresponds to $\varphi \in [\varphi_2^u, \varphi_2^x)$. In this case, we have $c(\varphi) \propto w_s(\varphi)^{\nu}\varphi^{-\nu-\eta(1-\nu)}$ and, in view of $c(\varphi) = c(\varphi_2^u)$: $w_s(\varphi)/w_s(\varphi_2^u) = (\varphi/\varphi_2^u)^{1+\eta(1-\nu)/\nu}$. Evaluating Eq. (A.5) at $\varphi = \varphi_2^u$, we obtain

$$\frac{w_s(\varphi)}{w_s(\varphi_1^x)} = \left(\frac{\varphi}{\varphi_1^x}\right)^{\frac{\nu+\eta(1-\nu)}{\nu}} \left(\frac{\varphi_2^u}{\varphi_1^u}\right)^{(1-\nu)[s-u-\eta/\nu]}. \tag{A.6}$$

Equating with $w_s(\varphi)/w_s(\varphi_1^x) = (\varphi/\varphi_1^x)^{-g+s}[1+\tau^{1-\sigma}\chi(\varphi)]$ and accounting for $1+\eta(1-\nu)/\nu+g-s = \Delta[1+\nu(\sigma-1)]/\nu$ gives the fourth line in Eq. (18) and completes the formal characterisation of $\chi(\varphi)$.

A.4 Derivation of Eqs. (20), (21) and characterisation of $\omega_s(\lambda)$, $\omega_u(\lambda)$

Let us first consider the unskilled wage bill in Eq. (21), which is given by

$$W_u = L_u w_u(1) \int_1^{\varphi_1^u} \beta^{1-\eta} dU(\beta) + L_u w_u(1) \int_{\varphi_1^u}^{\varphi_2^u} \left(\frac{\beta}{\varphi_1^u}\right)^{\nu(u+\eta/\nu-s)} \beta^{1-\eta} dU(\beta)$$
$$+ L_u w_u(1) \int_{\varphi_2^u}^{\infty} \left(\frac{\varphi_2^u}{\varphi_1^u}\right)^{\nu(u+\eta/\nu-s)} \beta^{1-\eta} dU(\beta). \tag{A.7}$$

Solving the integrals establishes Eq. (21), with

$$\omega_u(\lambda) \equiv \frac{\nu(u+\eta/\nu-s)}{\nu s + (1-\nu)u - 1} \left[1 - \lambda^{1-\nu s-(1-\nu)u}\right]. \tag{A.8}$$

It is immediate from Eq. (A.8) that $\omega_u(\lambda) > 0$ if $u + \eta/\nu - s > 0$ and that $\omega_u(1) = 0$ holds in this case.

Let us now turn to the wage bill of skilled workers, which is given by

$$\begin{aligned} W_s &= L_s \int_1^{\infty} ws(\varphi) dS(\varphi) \\ &= L_s w_s(1) \int_1^{\varphi_1^x} \varphi^{-\frac{g}{\Theta}+s} dS(\varphi) + L_s w_s(1) \int_{\varphi_1^x}^{\varphi_1^u} \left(\frac{\varphi}{\varphi_1^x}\right)^{g\frac{\Theta-1}{\Theta}} \varphi^{-g+s} dS(\varphi) \\ &\quad + L_s w_s(1) \int_{\varphi_1^u}^{\varphi_2^u} \left(\frac{\varphi_1^u}{\varphi_1^x}\right)^{\frac{\nu+(1-\nu)\eta-s\nu}{\nu}} \left(\frac{\varphi}{\varphi_1^u}\right)^{1+g-\nu s-(1-\nu)u} (\varphi_1^x)^{g\frac{\Theta-1}{\Theta}} \varphi^{-g+s} dS(\varphi) \\ &\quad + L_s w_s(1) \int_{\varphi_2^u}^{\varphi_2^x} \left(\frac{\varphi_2^u}{\varphi_1^u}\right)^{(1-\nu)(s-\eta/\nu-u)} \left(\frac{\varphi}{\varphi_1^x}\right)^{\frac{\nu+(1-\nu)\eta-s\nu}{\nu}} (\varphi_1^x)^{g\frac{\Theta-1}{\Theta}} \varphi^{-g+s} dS(\varphi) \\ &\quad + L_s w_s(1) \int_{\varphi_2^x}^{\infty} (1+\tau^{1-\sigma}) \left(\frac{\varphi_1^x}{\varphi_2^x}\right)^{g\frac{\Theta-1}{\Theta}} \varphi^{-\frac{g}{\Theta}+s} dS(\varphi). \end{aligned} \tag{A.9}$$

Solving the integrals, the wage bill simplifies to Eq. (20), with

$$\begin{aligned} \omega_s(\lambda) &\equiv \frac{g\nu - \Delta\nu(\sigma-1)}{g\nu - \Delta[1+\nu(\sigma-1)]} \left[1 - \left(\frac{\varphi_1^u}{\varphi_1^x}\right)^{\frac{\Delta[1+\nu(\sigma-1)]-g\nu}{\nu}}\right] \\ &\quad + \frac{g - \Delta(\sigma-1)}{\nu s + (1-\nu)u - 1} \left(\frac{\varphi_1^u}{\varphi_1^x}\right)^{\frac{\Delta[1+\nu(\sigma-1)]-g\nu}{\nu}} \left[1 - \lambda^{1-\nu s-(1-\nu)u}\right] \\ &\quad + \frac{g\nu - \Delta\nu(\sigma-1)}{g\nu - \Delta[1+\nu(\sigma-1)]} \lambda^{(1-\nu)(s-u-\eta/\nu)} \left(\frac{\varphi_2^u}{\varphi_1^x}\right)^{\frac{\Delta[1+\nu(\sigma-1)]-g\nu}{\nu}} \left[1 - \left(\frac{\varphi_2^x}{\varphi_2^u}\right)^{\frac{\Delta[1+\nu(\sigma-1)]-g\nu}{\nu}}\right] \\ &\quad + (1+\tau^{1-\sigma}) \left(\frac{\varphi_2^x}{\varphi_1^x}\right)^{-g} - 1, \end{aligned} \tag{A.10}$$

Rearranging terms and accounting for Eqs. (9), (19), and (A.8), $\omega_s(\lambda)$ simplifies to

$$\omega_s(\lambda) = \frac{\Delta}{g\nu - \Delta[1+\nu(\sigma-1)]}\left[1-(1+\tau^{1-\sigma})\left(\frac{\varphi_2^x}{\varphi_1^x}\right)^{-g}\right] - \frac{(1-\nu)[g-\Delta(\sigma-1)]}{g\nu-\Delta[1+\nu(\sigma-1)]}\left(\frac{\varphi_1^u}{\varphi_1^x}\right)^{\frac{\Delta[1+\nu(\sigma-1)]-g\nu}{\nu}}\omega_u(\lambda). \tag{A.11}$$

Accounting for $\omega_u'(\lambda) > 0$ if $u+\eta/\nu-s > 0$ from above, it follows from Eq. (A.11) that a higher λ (and thus lower ω_u) lowers the exporter wage premium of skilled workers. However, this captures only the effect through an increase in ω_u, leaving aside additional effects from adjustments in productivity ratios φ_2^x/φ_1^x and φ_1^u/φ_1^x. These additional effects and their consequences for ω_s are addressed in the proof of Proposition 1.

A.5 Derivation of Eq. (26)

The share of exporting firms is given by

$$\begin{aligned}\chi &= \int_{\varphi_1^x}^{\varphi_2^x}\chi(\varphi)dG(\varphi) + \int_{\varphi_2^x}^{\infty}dG(\varphi)\\ &= \tau^{\sigma-1}\left\{\int_{\varphi_1^x}^{\varphi_1^u}\left(\frac{\varphi}{\varphi_1^x}\right)^{\frac{\Delta[1+\nu(\sigma-1)]}{\nu}}dG(\varphi) + \int_{\varphi_1^u}^{\varphi_2^u}\left(\frac{\varphi_1^u}{\varphi_1^x}\right)^{\frac{\Delta[1+\nu(\sigma-1)]}{\nu}}\left(\frac{\varphi}{\varphi_1^u}\right)^{1+g-\nu s-(1-\nu)u}dG(\varphi)\right.\\ &\quad\left.+\int_{\varphi_2^u}^{\varphi_2^x}\left(\frac{\varphi_2^u}{\varphi_1^u}\right)^{(1-\nu)(s-u-\eta/\nu)}\left(\frac{\varphi}{\varphi_1^x}\right)^{\frac{\Delta[1+\nu(\sigma-1)]}{\nu}}dG(\varphi) - \int_{\varphi_1^x}^{\varphi_2^x}dG(\varphi)\right\} + \int_{\varphi_2^x}^{\infty}dG(\varphi)\end{aligned} \tag{A.12}$$

Solving the integrals establishes

$$\begin{aligned}\tau^{1-\sigma}(\varphi_1^x)^g\chi &= \frac{\Xi}{\Xi-1}\left[1-\left(\frac{\varphi_1^u}{\varphi_1^x}\right)^{\frac{g(1-\Xi)}{\Xi}}\right] + \frac{g}{\nu s+(1-\nu)u-1}\left(\frac{\varphi_1^u}{\varphi_1^x}\right)^{g\frac{1-\Xi}{\Xi}}\left(1-\lambda^{1-\nu s-(1-\nu)u}\right)\\ &+\frac{\Xi}{\Xi-1}\lambda^{(1-\nu)(s-u-\eta/\nu)}\left(\frac{\varphi_2^u}{\varphi_1^x}\right)^{g\frac{1-\Xi}{\Xi}}\left[1-\left(\frac{\varphi_2^x}{\varphi_2^u}\right)^{g\frac{1-\Xi}{\Xi}}\right] + (1+\tau^{1-\sigma})\left(\frac{\varphi_2^x}{\varphi_1^x}\right)^{-g} - 1.\end{aligned} \tag{A.13}$$

Accounting for the definition of Θ and substituting $\omega_s(\lambda)$ from Eq. (A.11), gives Eq. (26). This completes the proof.

A.6 Characterisation of implicit function $\Omega(\varphi_1^x, \lambda; \tau, f_x) = 0$

Solving Eq. (26) for ω_s and considering the indifference condition for the marginal exporter in Eq. (14), we can rewrite Eq. (24) as follows

$$R = \Theta N r(1) + \sigma N f_x\left\{\chi - (\Theta-1)\tau^{\sigma-1}(\varphi_1^x)^{-g}\left[1-(1+\tau^{1-\sigma})\left(\frac{\varphi_2^x}{\varphi_1^x}\right)^{-g}\right]\right\} \tag{A.14}$$

Accounting for the free entry condition in the open economy, we can solve for the revenues of the marginal exporter

$$r(1) = \frac{\sigma f_e}{\Theta - (\Theta - 1)(\varphi_1^x)^{-\frac{g}{\Theta}}\left[1 - (1+\tau^{1-\sigma})\left(\frac{\varphi_2^x}{\varphi_1^x}\right)^{-g}\right]}. \tag{A.15}$$

Combining Eqs. (25) and (A.15) and accounting for Eq. (19) we can solve for the implicit function

$$\begin{aligned}\Omega(\varphi_1^x, \lambda; \tau, f_x) &= \Theta(\varphi_1^x)^{-g\frac{\Theta-1}{\Theta}} \\ &- (\Theta-1)(\varphi_1^x)^{-g}\left[1 - (1+\tau^{1-\sigma})^{1-\Xi}\lambda^{-\Xi(1-\nu)[u+\eta/\nu-s]}\right] - \frac{f_e}{f_x\tau^{\sigma-1}} = 0,\end{aligned} \tag{A.16}$$

which respects both free entry into the productivity lottery and indifference of the marginal exporter with productivity $\varphi_1^x > 1$. Computing the partial derivatives of $\Omega(\cdot)$ is straightforward and thus left for the interested reader. In view of $\partial\Omega/\partial\varphi_1^x < 0$, $\partial\Omega/\partial\lambda < 0$, and $\lim_{\varphi_1^x\to\infty}\Omega(\cdot) = -\tau^{1-\sigma}f_e/f_x < 0$, it is immediate that $\Omega(1, 1; \tau, f_x) > 0$ is necessary (not sufficient) for an outcome with $\Omega(\cdot) = 0$ that features $\varphi_1^x > 1$ and $\lambda \geq 1$. This establishes the parameter constraint in (28).

A.7 Characterisation of implicit function $\Phi(\lambda, \varphi_1^x; \tau) = 0$

Implicit function $\Phi(\lambda, \varphi_1^x; \tau) = 0$ only exists if there is segmentation of unskilled workers over interval $[\varphi_1^u, \varphi_2^u]$ – presuming $u + \eta/\nu - s > 0$ – and it is derived from the condition of labour market clearing over the ability segments $\beta \leq \varphi_1^u$ and $\beta \geq \varphi_2^u$. Let us first consider labour market clearing over ability segment $\beta \leq \varphi_1^u$. Due to pooling of unskilled workers over firms with productivity $\varphi \leq \varphi_1^u$, demand for efficiency units of unskilled labour by these firms is given by

$$L_u^{D1} = \frac{(1-\nu)}{\nu}\frac{sL_s w_s(1)}{w_u(1)}\frac{\Theta}{g}\left\{1 - (\varphi_1^x)^{-\frac{g}{\Theta}} + (\varphi_1^x)^{-\frac{g}{\Theta}}\frac{\Xi}{\Theta}\left[1 - \left(\frac{\varphi_1^u}{\varphi_1^x}\right)^{g\frac{1-\Xi}{\Xi}}\right]\right\}, \tag{A.17}$$

according to Eq. (18) and the observation that $[\nu(\sigma-1)/\sigma]Nr(1)dG(1) = L_s w_s(1)dS(1)$. Evaluating Eq. (A.4) at $\varphi = \varphi_1^u$ and setting labour demand equal to labour supply, $L_u\Gamma[1 - (\varphi_1^u)^{-\frac{u}{\Gamma}}]$, finally gives the labour market clearing condition

$$(\varphi_1^x)^{\frac{g}{\Theta}} - \left[\frac{\Xi}{\Theta} - \frac{g}{\Theta}\frac{\Gamma}{u}\right]\left(\frac{\varphi_1^u}{\varphi_1^x}\right)^{\frac{\Delta[1+\nu(\sigma-1)]}{\nu}} + \frac{\Xi}{\Xi-1}\frac{\Delta}{g\nu}\frac{g}{\Theta}\frac{\Gamma}{u}\left(\frac{\varphi_1^u}{\varphi_1^x}\right)^{u+\eta/\nu-s}(\varphi_1^x)^{\frac{u}{\Gamma}} = 0. \tag{A.18}$$

Following the same derivation steps as above for ability domain $\beta \geq \varphi_2^u$, establishes market clearing if

$$\left(\frac{\varphi_2^x}{\varphi_2^u}\right)^{g\frac{\Xi-1}{\Xi}} = \frac{\Theta\Delta}{g\nu}\frac{u}{\Gamma(u+\eta/\nu-s)} \equiv \xi. \tag{A.19}$$

For $u + \eta/\nu - s > 0$, the right-hand side of this expression is larger (smaller) than one if

$g\nu - \Delta[1+\nu(\sigma-1)] > (<)0$, establishing $\varphi_2^x/\varphi_2^u > 1$. Combining market clearing conditions (A.18) and (A.19) with Eq. (19) allows us to formulate the implicit function

$$\Phi(\varphi_1^x, \lambda; \tau) = (\varphi_1^x)^{\frac{g}{\Theta}} + \frac{\Xi}{\Xi - 1}\frac{\Delta}{g\nu}\left[1 - \rho(\lambda;\tau)^{\Delta[1+\nu(\sigma-1)]-g\nu}\right]$$
$$- \frac{g}{\Theta}\frac{\Gamma}{u}\xi^{-\frac{\nu(u+\eta/\nu-s)}{g\nu-\Delta[1+\nu(\sigma-1)]}}\rho(\lambda;\tau)^{\nu(u+\eta/\nu-s)}(\varphi_1^x)^{\frac{u}{\Gamma}} = 0, \qquad \text{(A.20)}$$

where $\rho(\lambda;\tau) \equiv (1+\tau^{1-\sigma})^{\frac{1}{\Delta[1+\nu(\sigma-1)]}}\lambda^{-\frac{g+1-\nu s-(1-\nu)u}{\Delta[1+\nu(\sigma-1)]}}$ has been considered. Since it is a tedious task to show the properties of $\Phi(\lambda, \varphi_1^x; \tau)$, we have deferred a detailed discussion to a Supplement, which is available upon request. There, we show that $\Phi(\cdot) = 0$ has a solution with $\varphi_1^x, \lambda > 1$ if and only if, in addition to $u + \eta/\nu - s > 0$, $1 + \tau^{1-\sigma} > \xi^{\frac{1}{\Xi-1}}$ holds. These two conditions establish $\bar{\lambda} > 1$ in the main text. We also show that if $\Phi(\cdot) = 0$ has a solution with $\varphi_1^x, \lambda > 1$, the partial derivatives of Φ are $\partial\Phi/\partial\lambda > 0$, $\partial\Phi/\partial\varphi_1^x > 0$, and $\partial\Phi/\partial\tau > 0$. As a side product, we also show that $\varphi_1^u/\varphi_1^x > 1$ must hold in this case.

A.8 Existence and uniqueness of an interior equilibrium

As outlined in the main text, if $u + \eta/\nu - s \leq 0$, we have $\lambda = 1$ and in this case the parameter constraint in (28) is necessary and sufficient for an outcome with $\varphi_1^x > 1$ and $\lambda = 1$ – see the discussion below Eq. (A.16). A similar result is obtained if $u + \eta\nu - s > 0$ and $1 + \tau^{1-\sigma} \leq \xi^{\frac{1}{\Xi-1}}$, because in this case $\Phi(\cdot) = 0$ has no solution with $\varphi_1^x, \lambda > 1$, and, hence, under parameter constraint (28) the equilibrium values of φ_1^x and λ are given by $\Omega(\varphi_1^x, 1; \tau, f_x) = 0$. The third possible scenario is one with $u + \eta\nu - s > 0$ and $1 + \tau^{1-\sigma} \geq \xi^{\frac{1}{\Xi-1}}$, and in this case both $\Omega(\varphi_1^x, \lambda; \tau, f_x) = 0$ and $\Phi(\lambda, \varphi_1^x; \tau) = 0$ are consistent with $\varphi_1^x, \lambda > 1$. This does not mean that the implicit functions $\Omega(\cdot) = 0$ and $\Phi(\cdot) = 0$ have an intersection point in the relevant (φ_1^x, λ)-space. However, by parameter constraint (31) it is ensured that $\Omega(1, \bar{\lambda}; \tau, fx) > 0$ and since $\partial\Omega/\partial\lambda < 0$, it must be true that at low levels of φ_1^x, $\Phi(\cdot) = 0$ lies below $\Omega(\cdot) = 0$ in the relevant (φ_1^x, λ)-space. Hence, if there is no intersection of the two implicit function loci, the equilibrium realisations of φ_1^x and λ must again be characterised by $\Omega(\varphi_1^x, 1; \tau, f_x) = 0$. In a final step, we show that if $\Phi(\cdot) = 0$ and $\Omega(\cdot) = 0$ intersect for some $\varphi_1^x, \lambda > 1$, the intersection point must be unique. A sufficient condition for this is that inequality (32) holds in the intersection point. Using the partial derivatives of Ω and Φ, we can rewrite this inequality as follows

$$-\frac{1 + (\varphi_1^x)^{-\frac{g}{\Theta}}[C-1]}{C} < \frac{1}{\xi}\frac{(1-\nu)(u+\eta-1)}{g+1-\nu s-(1-\nu)u}(\varphi_1^x)^{-\frac{g}{\Theta}}B(\rho), \qquad \text{(A.21)}$$

where $C \equiv (1+\tau^{1-\sigma})^{1-\Xi}\lambda^{-\Xi(1-\nu)[u+\eta/\nu-s]}$ and

$$B(\rho) \equiv \frac{(\varphi_1^x)^{\frac{g}{\Theta}} - \xi^{-\frac{\nu(u+\eta/\nu-s)}{g\nu-\Delta[1+\nu(\sigma-1)]}}\rho^{\nu(u+\eta/\nu-s)}(\varphi_1^x)^{\frac{u}{\Gamma}}}{\rho^{\Delta[1+\nu(\sigma-1)]-g\nu} - \xi^{-\frac{\nu(u+\eta-1)}{g\nu-\Delta[1+\nu(\sigma-1)]}}\rho^{\nu(u+\eta/\nu-s)}(\varphi_1^x)^{\frac{u}{\Gamma}}}. \qquad \text{(A.22)}$$

Differentiating the left-hand side of (A.21) with respect to C gives $-[(\varphi_1^x)^{-g/\theta} - 1]/C^2 > 0$. In view of $C \leq 2$, we can thus safely conclude that

$$-\frac{1 + (\varphi_1^x)^{-\frac{g}{\Theta}}[C-1]}{C} < -\frac{1 + (\varphi_1^x)^{-\frac{g}{\Theta}}}{2}. \tag{A.23}$$

Let us now turn to the right-hand side of (A.21). Differentiating $B(\rho)$ gives $B'(\rho) > 0$, implying

$$B(\rho) > B\left(\xi^{\frac{1}{g\nu - \Delta[1+\nu(\sigma-1)]}}\right) = \xi \frac{(\varphi_1^x)^{\frac{g}{\Theta}} - (\varphi_1^x)^{\frac{u}{\Gamma}}}{1 - (\varphi_1^x)^{\frac{u}{\Gamma}}}. \tag{A.24}$$

Since

$$-\frac{1 + (\varphi_1^x)^{-\frac{g}{\Theta}}}{2} < \frac{(1-\nu)(u+\eta-1)}{g+1-\nu s-(1-\nu)u} \frac{(\varphi_1^x)^{\frac{g}{\Theta}} - (\varphi_1^x)^{\frac{u}{\Gamma}}}{1 - (\varphi_1^x)^{\frac{u}{\Gamma}}} (\varphi_1^x)^{-\frac{g}{\Theta}}. \tag{A.25}$$

holds for any $\varphi_1^x > 1$, we can safely conclude that if $\Phi(\cdot) = 0$ and $\Omega(\cdot) = 0$ have an intersection point in the (φ_1^x, λ)-space, this intersection point must be unique.

A.9 Proof of Proposition 1

We consider $\omega_u(1) = 0$ and $\omega_u'(\lambda) > 0$ from Eq. (A.8) and show that $\omega_s(\lambda) > \omega_u(\lambda)$ and thus $\omega_s(\lambda) > 0$ holds for all $\lambda \leq \bar{\lambda}$. If $\lambda = 1$, the exporter wage premium of skilled workers in Eq. (A.11) simplifies to $\omega_s(1)$ in Eq. (23). Since $d\omega_s(1)/d\tau < 0$ and $\lim_{\tau\to\infty} \omega_s(1) = 0$, it follows that $\omega_s(1) > 0$ for all (finite values of) τ, whereas $\omega_u(1) = 0$ is established by Eq. (A.8). If $u + \eta/\nu - s > 0$ allows for $\lambda > 1$, we can make use of φ_2^x/φ_1^x from Eq. (19) and $\varphi_1^u/\varphi_1^x = \varphi_1^u/\varphi_2^u \times \varphi_2^u/\varphi_2^x \times \varphi_2^x/\varphi_1^x = \rho(\lambda;\tau)^\nu \xi^{-\frac{\nu}{g\nu - \Delta[1+\nu(\sigma-1)]}}$ from Eqs. (19), (A.19), and the definition of λ and rewrite the exporter wage premium of skilled workers in Eq. (A.11) as follows

$$\omega_s(\lambda) = \frac{\Xi}{\Xi-1}\frac{\Delta}{g\nu}\Bigg\{1 - (1+\tau^{1-\sigma})^{-1-\Xi}\lambda^{-\Xi(1-\nu)(u+\eta/\nu-s)} \times \left[\frac{g\nu - \Delta[1+\nu(\sigma-1)]}{\nu s+(1-\nu)u-1} + \frac{(1-\nu)(u+\eta-1)}{\nu s+(1-\nu)u-1}\lambda^{\nu s+(1-\nu)u-1}\right]\Bigg\}, \tag{A.26}$$

with

$$\omega_s'(\lambda) = \frac{(1+\tau^{1-\sigma})^{-1-\Xi}\lambda^{-\Xi(1-\nu)(u+\eta/\nu-s)-1}}{1+\nu(\sigma-1)}\Bigg[\frac{g\nu(1-\nu)(u+\eta/\nu-s)}{\nu s+(1-\nu)u-1} - \frac{[g+1-\nu s-(1-\nu)u](1-\nu)[u+\eta-1]}{\nu s+(1-\nu)u-1}\lambda^{\nu s+(1-\nu)u-1}\Bigg], \tag{A.27}$$

Since the bracket term on the right hand side is decreasing in λ and since $\omega_s'(1) \propto -(g+1-u-\eta)$, which is negative according to Assumption 1, we can safely conclude that $\omega_s'(\lambda) < 0$ holds for all possible λ. Using $\omega_u'(\lambda) > 0$ and

$$\omega_s(\bar{\lambda}) = \frac{\Xi}{\Xi-1}\frac{\Delta}{g\nu}\left\{1 - \frac{(1-\nu)(u+\eta-1)}{\nu_s+(1-\nu)u-1}\frac{1}{\xi} - \frac{g\nu - \Delta[1+\nu(\sigma-1)]}{\nu s+(1-\nu)u-1}\bar{\lambda}^{1-/\nu s-(1-\nu)u}\right\}, \tag{A.28}$$

it follows that $\omega_s(\lambda) - \omega_u(\lambda) > \hat{\Delta}_\omega(\bar{\lambda}) \equiv \omega_s(\bar{\lambda}) - \omega_u(\bar{\lambda})$, $\hat{\Delta}'_\omega(\bar{\lambda}) > 0$, and $\hat{\Delta}_\omega(1) = [\Xi/(\Xi - 1)][\Delta/(g\nu)](1 - 1/\xi) > 0$.[30] Hence, we can conclude that $\omega_s(\lambda) > \omega_u(\lambda) \geq 0$ holds for all possible λ. This completes the proof.

A.10 Proof of Proposition 2

Since a detailed proof of Proposition 2 is tedious, we provide a sketch of how such a proof can be organised here and defer a more detailed formal discussion to a Supplement, which is available upon request. Starting point is the observation that the Theil index respects (mean-normalised) second order stochastic dominance, and thus ranks income distribution in the open economy as more unequal than the income distribution in the closed economy, if the latter Lorenz dominates the former. To put it differently, if Lorenz curves do not intersect, we can use them to establish the ranking of income distributions from the Theil index. Regarding the Lorenz curves of unskilled workers, we have to distinguish three segments in the open economy if $\lambda > 1$. The difference between Lorenz curves in the closed and the open economy for segment 1, with $\beta \leq \varphi_1^u$ is given by

$$\Delta_1^u(\mu_u) \equiv [1 - (1 - \mu_u)]^{\frac{1}{\Gamma}} \left[1 - \frac{1}{1 + (\varphi_1^u)^{-\frac{u}{\Gamma}}\omega_u}\right], \tag{A.29}$$

which is positive and decreasing in the share of unskilled workers μ_u over ability interval $(1, \varphi_1^u]$. Let us first consider the case of unskilled workers. The differentials in the second and third segments of the Lorenz curve are given by

$$\Delta_2^u(\mu_u) \equiv 1 - (1 - \mu_u)^{\frac{1}{\Gamma}} - \frac{1}{1 + (\varphi_1^u)^{-\frac{u}{\Gamma}}\omega_u} \left\{ 1 - (\varphi_1^u)^{-\frac{u}{\Gamma}} \right.$$
$$\left. + (\varphi_1^u)^{-\frac{u}{\Gamma}} \frac{u + \eta - 1}{\nu s + (1 - \nu)u - 1} \left[1 - \left[(\varphi_1^u)^{-u}(1 - \mu_u)\right]^{\frac{\nu s + (1-\nu)u - 1}{u}}\right] \right\} \tag{A.30}$$

and

$$\Delta_3^u(\mu_u) \equiv (1 - \mu_u)^{\frac{1}{\Gamma}} \left[\frac{\lambda^{\nu(u + \eta/\nu - s)}}{1 + (\varphi_1^u)^{-\frac{u}{\Gamma}}\omega_u} - 1\right], \tag{A.31}$$

respectively. $\Delta_3^u(\mu_u)$ is positive and decreasing in the share of unskilled workers μ_u over the entire ability interval $[\varphi_2^u, \infty)$. Furthermore, we have $\Delta_1^u(\mu_{u1}^1) = \Delta_2^u(\mu_{u1}^1)$ and $\Delta_2^u(\mu_{u2}^2) = \Delta_3^u(\mu_{u2}^2)$, with $\mu_{u1}^1 \equiv 1 - (\varphi_1^u)^{-u}$ and $\mu_{u2}^2 \equiv 1 - (\varphi_2^u)^{-u}$, respectively. Noting further that $d\Delta_2^u/d(\mu_u) < 0$, it follows that $\Delta_2^u(\mu_u) > 0$ must hold over ability interval $(\varphi_1^u, \varphi_2^u)$. This confirms that distribution of unskilled wages in the closed economy Lorenz dominates the distribution of unskilled wages in the open economy, so that T_u must be larger in the open than the closed economy, whenever $\lambda > 1$. If $\lambda = 1$, the Theil index for unskilled wages is not affected by trade.

Let us now turn to the Lorenz curve for skilled wages, which has five (three) segments if

[30] Rearranging terms in Eq. (A.19), one can show that in a scenario with $u + \eta/\nu - s > 0$, which is a prerequisite for $\lambda > 1$, we have $\xi >, =, < 1$ if $g\nu - \Delta[1 + \nu(\sigma - 1)] >, =, < 0$. This establishes $\hat{\Delta}_\omega(1) > 0$.

$\lambda > (=)1$. For the first segment the difference between the Lorenz curve in closed and the open economy is

$$\Delta_1^s(\mu_s) \equiv [1-(1-\mu_s)]^{\frac{g}{s\Theta}} \left[1 - \frac{1}{1+(\varphi_1^x)^{-\frac{g}{\Theta}}\omega_u}\right], \tag{A.32}$$

which is positive and increasing in the share of skilled workers over ability interval $(1, \varphi_1^x]$. The differential for the fifth segment of the Lorenz curve is given by

$$\Delta_5^s(\mu_s) = (1-\mu_s)^{\frac{g}{s\Theta}} \left\{ \frac{1}{1+(\varphi_1^x)^{-\frac{g}{\Theta}}\omega_s} (1+\tau^{1-\sigma}) \left(\frac{\varphi_1^x}{\varphi_2^x}\right)^{g\frac{\Theta-1}{\Theta}} - 1 \right\}. \tag{A.33}$$

It is positive and decreasing over ability interval $[\varphi_2^x, \infty)$. We do not display the remaining three segments here, but refer the interested reader to our Supplement. There, we show that $\Delta_i^s > 0$ indeed holds for all $i = 1, 2, ..., 5$. Accordingly, the wage distribution of skilled workers in the closed economy Lorenz dominates the wage distribution of skilled workers in the open economy, so that T_s must be larger with trade than under autarky, irrespective of whether λ is larger than or equal to one. This completes the proof.