Trade and Unemployment with Heterogeneous Firms: How Good Jobs Are Lost

Lutz Altenburg

Diskussionsbeitrag Nr. 476
November 2011

Trade and Unemployment with Heterogeneous Firms: How Good Jobs Are Lost

Lutz Altenburg *

University of Hagen

November 2011

Abstract

This paper incorporates shirking-based efficiency wages into a heterogeneous firm model of international trade in which firms differ in monitoring ability. In equilibrium firms abler in monitoring ask greater worker effort and therefore have an edge in productivity over less able competitors while all of them pay the same wage. Different effort levels entail different effort costs and hence firm specific rents to identical workers. We show that the opening to trade reduces unemployment and, under a plausible condition, raises wages and aggregate factor productivity. Workers staying in employment and the unemployed gain from trade but jobs carrying the highest rents to workers (the least onerous ones) are lost.

Keywords: Efficiency wages; Trade; Unemployment; Firm Heterogeneity

JEL classification: F12; F16; J41

* I am grateful to Anke Brenken, Thomas Eichner, and Felicitas Nowak-Lehmann for helpful comments and suggestions.

E-mail: lutz.altenburg@fernuni-hagen.de

1 Introduction

Although many people think that trade is good for the economy as a whole, there is widespread concern that it might hurt workers, particularly by destroying jobs. This is one of the findings of empirical work on individual attitudes towards trade (e.g., Scheve and Slaughter, 2006). Recently, the issue of trade and jobs has been addressed anew by a line of theoretical research that incorporates labour market imperfections into models of monopolistic competition and firms with heterogeneous productivity, pioneered by Melitz (2003). [1] In the Melitz model there is no unemployment since labour markets are assumed to be perfect. By reallocating market shares towards the more productive firms, trade leads to aggregate productivity gains and hence higher real wages so that all workers gain from it. Things become quite different when labour market imperfections are allowed for. First, the extended heterogeneous firm models (which will be detailed below) shed new light on how trade openness influences aggregate unemployment. Second, part of them allow to study job specific effects of trade. When firms differ in productivity or some other characteristic, jobs at different firms may have different rents to homogeneous workers because they pay different wages or offer different working conditions. In such cases it is possible to distinguish between good (high-rent) and bad (low-rent) jobs. This raises the question: Are there jobs that are at higher risk of being lost through trade than others and if so, are these good or bad jobs?

To address these questions, we develop a framework that incorporates a variant of the Shapiro and Stiglitz (1984) shirking model with heterogeneous firms into a model of trade between two symmetric countries. A single final product is produced by assembling a range of differentiated intermediate inputs. The latter are produced under monopolistic competition, with labour as the only input. Firms imperfectly observe their workers' effort and fire workers who shirk. For the threat of firing to be an effective incentive device, in equilibrium there is unemployment with the unemployed having lower utility than the employed.

Our variant departs from the Shapiro-Stiglitz model in two main respects. First, we abandon the assumption that firms can require only one fixed positive level of effort from their workers. Instead, in line with the "generic efficiency-wage model" (Romer, 2006), we treat work effort as a fully endogenous variable. Firms are free to offer contracts specifying a real wage and any positive level of effort. The contract offered must still satisfy the no-shirking condition, i.e., a worker's expected utility derived from exerting the required effort must be at least as high as his expected utility derived from shirking. However, firms choose workers' effort levels so as to minimize their cost per efficiency unit of labour, taking into account the no-shirking

[1] There also exists a sizable literature that adds imperfect labour markets to standard trade models with homogeneous firms. Labour market imperfections studied in this literature include minimum wages (Brecher, 1974; and Davis, 1998), efficiency wages (Copeland, 1989; Brecher, 1992; Matusz, 1996; and Hoon, 2001), search and matching (Davidson, Martin, and Matusz, 1988, 1999), and fair wages (Agell and Lundborg, 1995; and Kreickemeier and Nelson, 2006).

condition. [2]

Our second point of departure is the introduction of firm heterogeneity by allowing firms to differ in their efficacy to detect shirking. Specifically, we assume that a firm's monitoring ability is drawn from a known distribution upon entry. In equilibrium, firms with higher monitoring ability ask greater worker effort and hence are more productive than less able firms while all of them pay the same wage. In this respect, differences in monitoring ability play much the same role as differences in inherent firm productivity in the Melitz (2003) model. In addition, diversity in effort levels across firms implies that ex ante identical workers end up with different costs of effort and hence different utility levels. Jobs at firms with high monitoring ability are associated with high effort costs and low rents to workers. On the other hand, jobs at firms with low monitoring ability carry high rents because effort costs to workers are low.

We show that a transition from autarky to trade has the same kind of selection and reallocation effects as in the Melitz (2003) model. It forces input producers with low levels of monitoring ability to exit or contract while inducing the ablest input producers to select into exporting and expand. As a result, market shares are reallocated towards abler firms, which raises the firms' average monitoring ability. One implication of this is an unambiguous decrease in a country's unemployment rate. The intuition behind this result can be explained as follows. As firms on average become abler in monitoring, employed workers on average put forth greater effort on their jobs and therefore have to bear higher effort costs relative to wages. This lowers the average surplus to a worker from employment (the average job rent) and hence, for an unemployed, the average capital gain from becoming re-employed. As a consequence, the value to a worker of being unemployed declines. How does this affect the balance of the gain from and the cost of shirking at an individual firm? As in Shapiro and Stiglitz (1984) the cost of shirking equals the surplus from having a job at the firm (the job rent) since this surplus is forfeited when a worker is caught shirking. On the other hand, the gain from shirking depends positively on the cost of effort avoided and negatively on the probability of being caught shirking. In equilibrium the gain from and the cost of shirking just balance. As the opening to trade triggers a decline in the value of being unemployed, at the initial wage each firm sees its workers' job rent getting larger than needed to deter shirking. That means, for each active firm the opening to trade leads to slack in the no-shirking constraint. In our model the only way of removing it and restoring the balance of incentives is by a decline in the aggregate unemployment rate: lower unemployment makes it more likely for an unemployed to find a new job. This raises the expected capital gain from re-employment and hence the value of being unemployed, thereby lowering the firm job rent while leaving the gain from shirking unchanged.

Under a plausible parameter restriction the higher aggregate employment level is accompanied by an increase in a country's wage rate. Furthermore, it is shown that the unemployed

[2] Shirking models with endogenous effort have been used in a variety of contexts (e.g., Moselle, 1996; Mehta, 1998; Walsh, 1999; Rasmussen, 2002; Allgulin and Ellingsen, 2002; Alexopoulos, 2003). The present paper builds on Altenburg and Brenken (2008) who integrate shirking with endogenous effort into a monopolistic competition model of trade with homogeneous firms.

as well as all of the workers who stay in employment will gain from trade. However, workers who under autarky happen to be employed by less able firms, which exit or contract, may lose because of becoming jobless. The jobs lost are just the ones which under autarky require the lowest levels of effort, thereby providing the highest rents to workers.

Another main result is that exposure to trade raises the countries' aggregate factor productivity. As just mentioned, the surviving firms are on average abler in monitoring workers. As a result, workers on average put forth greater effort at given wages. This effect is magnified by a rise in within-firm productivity: the effort level workers provide at a firm with a given monitoring ability is increasing in the wage they are paid, and as trade pushes up wages, each surviving firm can elicit greater effort from its workers. The latter effect is consistent with evidence on the importance of trade-induced within-firm productivity gains. [3] It is also compatible with evidence suggesting that work intensity has kept on increasing over the past decades. [4] All of these results are derived without relying on a specific form of the distribution of monitoring ability levels.

The paper closest to ours is Davis and Harrigan (2011). Like ours, it merges a variant of Shapiro-Stiglitz (1984) efficiency wages with Melitz (2003) type firm heterogeneity to study the impact of trade liberalization between symmetric countries on job loss and aggregate unemployment. Yet the model developed by Davis and Harrigan differs from ours in important ways, as do its implications. First, they retain the assumption of the Shapiro-Stiglitz model that the level of effort for a non-shirking worker can take on only one fixed positive value rather than being an endogenous variable as in the present framework. Second, unlike ours their model features two sources of variations in marginal costs across firms instead of one: monitoring ability and inherent productivity, both of which are drawn from a known joint distribution. Differences across firms in monitoring ability translate into differences in wages such that the wage a firm has to pay its workers to deter shirking varies inversely with its monitoring ability. This has the following implications. As in our model there arise firm specific rents to homogeneous workers: jobs at firms with high monitoring ability have low rents, whereas jobs at firms with low monitoring ability carry high rents. But now the variation in rents is due to diversity in wages, not in effort costs.

Since in the Davis-Harrigan model marginal costs vary across firms because of differences in both productivity and wages, a variety of jobs are lost through trade: jobs at low-productivity firms that pay low wages just as jobs at high-productivity firms that pay high wages. Nonetheless, a key result of the Davis-Harrigan paper is that conditioning on firm productivity, trade always destroys jobs that pay the highest wages, while expanding only the number of the lowest paying jobs. This resembles our result that trade destroys jobs that require the lowest levels of effort, while expanding the number of jobs with the highest effort requirements. Simulat-

[3] See the surveys by Tybout (2003) and Bernard, Jensen, Redding, and Schott (2007).

[4] Using survey data for twelve countries, Green and McIntosh (2001) provide evidence of substantial increases in work intensity during the 1990s; see also Green (2006).

ing their model, Davis and Harrigan find that the transition from autarky to trade destroys a sizable fraction of good jobs paying above average wages, has substantial effects on aggregate productivity and real GDP but has a negligible impact on aggregate unemployment.

There are other papers adopting the Melitz (2003) model of firm heterogeneity that address issues similar to those in the present one but consider a different type of labour market imperfection. One branch of research deals with search and matching frictions, including Helpman and Itskhoki (2010), Helpman, Itskhoki, and Redding (2010), and Felbermayr, Prat, and Schmerer (2011). Helpman and Itskhoki (2010) study a two-country, two-sector model in which one sector produces a homogeneous consumption good and the other produces differentiated consumption goods. Search frictions in the labour market are allowed to vary across sectors and countries. Helpman and Itskhoki find, among other things, that trade liberalization raises welfare and aggregate factor productivity in both countries but has an ambiguous effect on unemployment even in the symmetric case (in which search frictions are the same across countries). The reason for the latter result is that trade liberalization raises the share of workers searching for jobs in the differentiated sector, thereby changing the sectoral composition of unemployment. Such a compositional effect on unemployment is absent in Felbermayr, Prat, and Schmerer (2011) who study an economy with symmetric countries, each of which produces a single final product assembled from differentiated intermediate inputs as in our model. Their model predicts that trade liberalization reduces unemployment and raises real wages if at the same time it increases a measure of aggregate productivity.

Helpman, Itskhoki, and Redding (2010) construct a model that features dispersion of firm wages by allowing for differences in worker ability. The abilities of the workers a firm employs contribute to its profitability but cannot be observed directly. This gives firms an incentive to screen workers. As more productive firms employ more effort in screening, they end up with on average more capable workers, paying them higher wages than less productive firms. One of their main findings is to provide conditions under which exposure to trade raises wage inequality. Another is that the opening of trade has again an ambiguous effect on unemployment.

A further line of research combines firm heterogeneity with fair wages, including Egger and Kreickemeier (2008, 2009) and Amiti and Davis (2011). In these models wages are firm specific, varying with firm productivity or profits by assumption. Like us, Egger and Kreickemeier (2009) consider a symmetric-country model with one final product produced with differentiated intermediate inputs. One of their findings is that the response of unemployment to the opening of trade depends on the relative size of fixed costs of production and fixed trade costs.

The rest of the paper is organized as follows. Section 2 presents the setup of the model for a closed economy. Section 3 describes an equilibrium under autarky. Section 4 considers a trading world with two countries and describes a trading equilibrium. Section 5 deals with the impact of a transition from autarky to trade. Section 6 looks at the effects of a decrease in variable trade costs and simulates the model in order to assess their magnitude. Section 7 concludes.

2 The Model

We develop in this section a closed economy version of our model. After describing the setup of the model, we turn in the next section to the determination of an equilibrium in the closed economy.

2.1 Workers

Consider an economy with a fixed number of identical workers denoted by L, each supplying one physical unit of labour. Workers are infinitely-lived, risk-neutral, and maximize their expected present-discounted lifetime utilities. Time is continuous. The analysis is confined to steady states.

At each instant of time a worker can either be employed or unemployed, and if employed can either work or shirk. When unemployed, a worker receives unemployment benefits b, paid by the government. To finance them, the government imposes a per capita tax T on employed and unemployed workers alike. The instantaneous utility of an employed worker is $w - T - e^{\gamma}/\gamma$ with $\gamma > 1$, where w is the real wage, e the level of effort, and e^{γ}/γ the cost of exerting effort. Assuming that when unemployed a worker does not experience any disutility of effort, the instantaneous utility of an unemployed is $b - T$.

When employed, a worker is asked by the firm to exert effort at a level which for now is taken as fixed. No matter whether or not he exerts effort at that level, a worker faces a risk of being separated from his job at rate δ per unit time. δ is the rate at which a firm is hit by a bad shock, forcing it to close down production so that all its jobs break up. In addition, shirkers, i.e., workers who provide less than the required effort level, face a risk of being detected and fired at rate q per unit time. While the break-up risk affects all firms and workers at the same rate, q is taken as a parameter that varies across firms, measuring the firm specific efficacy in detecting shirking (the "monitoring ability"), as in Davis and Harrigan (2011). Both the common parameter δ and the firm specific parameter q are the arrival rates of two independent Poisson processes and are defined on the set of positive real numbers. For a shirker it is then optimal to provide zero effort, while a nonshirker's best choice is to comply exactly with the required effort standard e.

Let r be the discount rate, $w(q)$ the wage paid by a firm with monitoring ability q, and $e(q)$ the level of effort required by that firm. Then the values of a worker employed at a q-firm who exerts effort, $V_N(q)$, and an employed worker who shirks, $V_S(q)$, satisfy the following asset equations:

$$\begin{aligned} rV_N(q) &= w(q) - T - e(q)^{\gamma}/\gamma + \delta[V_U - V_N(q)], \\ rV_S(q) &= w(q) - T + (\delta + q)[V_U - V_S(q)], \end{aligned} \tag{1}$$

where V_U represents a worker's value of being unemployed. For an employee to choose not to shirk at given values of $e(q)$, V_U, and T the firm has to pay a wage high enough so that

$V_N(q) \geq V_S(q)$. Using (1), this implies that

$$V_N(q) - V_U \geq \frac{e(q)^\gamma/\gamma}{q}. \tag{2}$$

As is well-known, in this form the no-shirking condition states that for a worker to exert effort, his surplus from having a job at a q-firm and working (his job rent) must be as least as high as his gain from shirking (being equal to the cost of effort avoided, e^γ/γ, times $1/q$ as a discount factor). Solving the first equation in (1) and (2) for $w(q)$, the no-shirking condition can be written as

$$w(q) \geq \widetilde{w}(e(q), V_U, T, q) \equiv rV_U + T + \left(\frac{q+r+\delta}{q}\right)\frac{e(q)^\gamma}{\gamma}. \tag{3}$$

It gives the minimum wage $\widetilde{w}(\cdot)$ a firm with monitoring ability q must pay its workers to induce them to provide a given level of effort $e(q)$. Since there is no need to pay a wage higher than $\widetilde{w}(\cdot)$, the firm chooses the wage so that (3) holds with equality, implying that $V_N(q) = V_S(q) \equiv V(q)$.

2.2 Production

A single homogeneous final good is produced by assembling a continuum of differentiated intermediate inputs. We assume that the production function for the final good has a CES form [5]

$$Y = \left[\int_{i\in I} x(i)^\theta di\right]^{1/\theta}, \quad 0 < \theta < 1,$$

where Y is final output, $x(i)$ the amount of intermediate input i, I the set of available input varieties, and the parameter θ is related to the elasticity of substitution σ between inputs by $\sigma = 1/(1-\theta) > 1$. [6] We assume that the final good sector is perfectly competitive and choose the final product as the numéraire.

The price index associated with Y, being equal to the unit cost of the final product, is given by

$$P = \left[\int_{i\in I} p(i)^{1-\sigma} di\right]^{\frac{1}{1-\sigma}},$$

where $p(i)$ denotes the price of intermediate input i. With the final product chosen as the numéraire, we have $P = 1$. Correspondingly, cost minimization in the final good sector implies that aggregate demand for input i is

$$x(i) = Yp(i)^{-\sigma}.$$

Intermediate inputs are produced with labour. We assume that the output $x(q)$ produced by a q-type intermediate input producer equals his effective labour input: $x(q) = e(q)l(q)$, where

[5] An alternative form of a CES production function for the final good, due to Blanchard and Giavazzi (2003), is employed by Egger and Kreickemeier (2009) and Felbermayr, Prat and Schmerer (2011). With that specification the mass of available input varieties does not influence upon productivity in final good production.

[6] As is well-known from Ethier (1982), this specification of the final good production function implies that the average productivity of intermediate inputs increases with the mass of available input varieties.

$l(q)$ denotes the employment level at that firm. In addition, production of each intermediate good requires a fixed overhead input f in terms of the final product, being the same for all intermediate input producers. [7] Each input variety will be produced by a single firm so that the market for intermediate goods is monopolistically competitive.

Facing intermediate input demand and having learnt how effective its monitoring technology q is, an active firm chooses its price and its effort standard at each moment so as to maximize its operating profit, taking (3) as a constraint. In doing so, it takes Y, V_U, and T as given. Thus, the firm's problem is

$$\max_{p(q),e(q)} \left\{ \left[p(q) - \frac{\widetilde{w}(e(q), V_U, T, q)}{e(q)} \right] Y p(q)^{-\sigma} - f \right\}.$$

The first-order conditions are

$$p(q) = \frac{1}{\theta} \frac{w(q)}{e(q)}, \tag{4}$$

$$\frac{w(q)}{e(q)} = \frac{\partial \widetilde{w}(\cdot)}{\partial e(q)}. \tag{5}$$

A firm's marginal cost equals the cost per efficiency unit of labour, $w(q)/e(q)$, and according to (4) the price of an intermediate good is a constant mark-up on marginal cost. Equation (5) is the Solow condition for the firm's subproblem of choosing $e(q)$ to minimize the cost per efficiency unit of labour. Using (3), we see that (5) implies

$$e(q) = \left(\frac{q}{q + r + \delta} \right)^{1/\gamma} w(q)^{1/\gamma}.$$

Substituting this back into (3) yields

$$w(q) = \frac{\gamma}{\gamma - 1} (r V_U + T). \tag{6}$$

Facing the same values of V_U and T, in equilibrium all intermediate input producers choose the same wage, independently of their monitoring ability. Hence, the index q in $w(q)$ can be dropped. [8] Using this fact, we get for the level of effort at a firm

$$e(m) = m w^{1/\gamma}, \tag{7}$$

[7] Taking fixed costs as being measured in units of final output departs from the Melitz model in which they are measured in units of labour. This departure is made here for analytical convenience only. It has also been used by Egger and Kreickemeier (2009) and Felbermayr, Prat, and Schmerer (2011) while Helpman and Itskhoki (2010) have fixed costs borne in units of a homogeneous consumption good produced by a second sector. By contrast, Davis and Harrigan (2011) adopt the Melitz formulation of fixed costs in terms of labour.

[8] The feature of all firms choosing the same wage is an implication of the constant elasticity form of the disutility of effort. With functional forms that exhibit a variable elasticity this feature would disappear. Heterogeneous firm models in which equal wages across firms also emerge, though for quite different reasons, include Helpman and Itskhoki (2010) and Felbermayr, Prat and Schmerer (2011).

where

$$m \equiv \left(\frac{q}{q+r+\delta}\right)^{1/\gamma}$$

is a transformed measure of monitoring ability. Note that m is strictly increasing in q. Note also that while q as a Poisson arrival rate can be arbitrarily large, m by definition takes on values between zero and one. In particular, m tends to one as q goes to infinity. This is the extreme case of perfect monitoring in which shirking is instantaneously detected so that a worker's job rent becomes zero. For convenience, we will henceforth use m instead of q to index intermediate input producers. Using (7) and the fact of equal wages across firms, the pricing rule (4) can then be rewritten as

$$p(m) = \frac{w}{\theta e(m)} = \frac{w^{(\gamma-1)/\gamma}}{\theta m}. \tag{8}$$

Revenue $R(m)$ and operating profits $\pi(m)$ of a m-firm are

$$R(m) = Yp(m)^{1-\sigma}; \qquad \pi(m) = \frac{R(m)}{\sigma} - f.$$

Equation (8) implies that $p(m_1)/p(m_2) = m_2/m_1$, where m_1 and m_2 are any two levels of monitoring ability. Accordingly, we have

$$\frac{x(m_1)}{x(m_2)} = \left(\frac{m_1}{m_2}\right)^{\sigma}; \qquad \frac{R(m_1)}{R(m_2)} = \frac{l(m_1)}{l(m_2)} = \left(\frac{m_1}{m_2}\right)^{\sigma-1}. \tag{9}$$

To get the latter equality in (9), use has been made of $l(m) = x(m)/e(m)$ and (7). Thus, we find that on the production side of our model monitoring ability plays a similar role as does productivity in Melitz (2003). A firm with higher monitoring ability asks greater effort of its workers (and hence is more productive), charges a lower price, has greater size in terms of output, employment, and revenue, and earns higher profits than a firm with lower monitoring ability.

Also note that workers employed by abler firms have higher effort costs and hence lower utility levels. Workers would therefore prefer to work in firms whose monitoring ability is low rather than in highly efficacious ones because a job in the former is less onerous. However, outsiders trying to get an easier job at a less able firm cannot credibly commit to exert more effort at the going wage than incumbents, in the same way as unemployed workers cannot credibly promise not to shirk at less than the going wage.

2.3 Firm Entry and Exit

Entry and exit of intermediate input producers are also modeled in a similar way as in Melitz (2003). In the intermediate goods sector there is an unbounded mass of prospective entrants. A firm that wants to enter the market has to pay an entry cost f_E in terms of the final good, which is sunk. Thereafter, the firm draws its monitoring ability q from a continuous cumulative distribution $G_q(q)$, defined for $0 < q < \infty$. As shown above, with r, δ, and γ taken as fixed parameters, the transformed measure m of monitoring ability is strictly increasing in q. Hence, a

q-draw from $G_q(q)$ is equivalent to drawing the corresponding m assigned to q from a cumulative distribution $G(m)$ with density $g(m)$ and support over $(0,1)$. [9]

Once some m has been drawn, each firm's monitoring ability remains unchanged over time, and so does its operating profit $\pi(m)$. If this profit level is negative, the firm will choose not to produce and immediately exit. If $\pi(m) \geq 0$, the firm will take up production. When producing, each firm faces a risk of being hit by a bad shock at rate δ per unit time, which forces it to shut down, lay off its workers, and exit. A firm's value function is then given by $v(m) = \max\{0, \pi(m) \int_0^\infty \exp[-(r+\delta)t]dt\} = \max\{0, \pi(m)/(r+\delta)\}$. By the same reasoning as in Melitz (2003), the lowest level m^* of monitoring ability (the cutoff level) at which it is worth starting production is given by $\pi(m^*) = 0$. [10] This is the zero cutoff profit condition.

The density function of monitoring ability in equilibrium, denoted by $\mu(m)$, is conditional on a successful draw ($m \geq m^*$):

$$\mu(m) = \begin{cases} \dfrac{g(m)}{1-G(m^*)} & \text{if } m \geq m^*, \\ 0 & \text{otherwise,} \end{cases}$$

where $1 - G(m^*)$ is the probability of a successful draw. This allows us to define - analogously to Melitz (2003) - the aggregate monitoring ability as

$$\widetilde{m}(m^*) \equiv \left[\int_0^1 m^{\sigma-1}\mu(m)dm\right]^{\frac{1}{\sigma-1}} = \left[\frac{1}{1-G(m^*)}\int_{m^*}^1 m^{\sigma-1}g(m)dm\right]^{\frac{1}{\sigma-1}}. \tag{10}$$

Aggregate operating profits Π are given by $\Pi = \int_0^1 \pi(m)M\mu(m)dm$. Using (9) and (10) and defining $\widetilde{m} \equiv \widetilde{m}(m^*)$, Π can be rewritten as $\Pi = M\pi(\widetilde{m})$. This implies that average operating profit $\overline{\pi} = \Pi/M$ equals the operating profit of the firm with monitoring ability $\widetilde{m}$ (the representative firm), $\pi(\widetilde{m})$. As in the Melitz model, this makes it possible to rewrite average operating profit as

$$\overline{\pi} = f\left[\left(\frac{\widetilde{m}(m^*)}{m^*}\right)^{\sigma-1} - 1\right], \tag{11}$$

where use has been made of (9) and the fact that the zero cutoff profit condition $\pi(m^*) = 0$ is equivalent to $R(m^*) = \sigma f$. A firm enters the market if the present-discounted value of average operating profit, conditional on successful entry, is at least as high as the entry cost f_E. This leads to the free entry condition

$$[1 - G(m^*)]\frac{\overline{\pi}}{r+\delta} = f_E. \tag{12}$$

[9] In the numerical version of our model we parametrize $G(m)$ as a truncated Pareto distribution defined for $0 < m_{\min} \leq m \leq 1$, where $m_{\min}$ is the lower bound of m in the support of $G(m)$ (to which is assigned some minimum level of q).

[10] The existence of a monitoring ability cutoff m^* follows from the facts that $\pi(0) = -f < 0$ and $\pi(m)$ is increasing in m. When there is some minimum level $m_{\min} > 0$, m^* exists if $\pi(m_{\min}) = R(m_{\min})/\sigma - f < 0$.

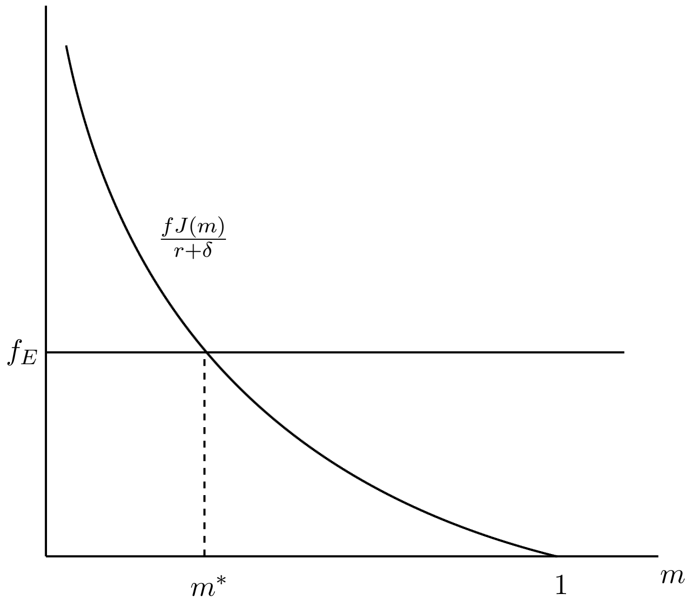

FIGURE 1. Determination of the equilibrium monitoring ability cutoff

3 Equilibrium in a Closed Economy

We begin with the determination of the cutoff monitoring ability level m^*. Substituting (11) into (12) yields

$$\frac{f}{r+\delta} J(m^*) = f_E, \tag{13}$$

where $J(m) \equiv [1 - G(m)][(\widetilde{m}(m)/m)^{\sigma-1} - 1]$. When $G(m)$ has support over (0,1), $J(m)$ monotonically decreases from infinity to zero on (0,1). This ensures the existence of a unique solution to (13), depicted in Figure 1. [11]

Next we turn to the determination of the equilibrium unemployment rate. Recall that when unemployed, a worker has an instantaneous utility of $b - T$. Let α be the rate per unit time at which he becomes re-employed and $\overline{V}$ be the expected value to a worker of being employed. The asset equation for an unemployed worker then is

$$rV_U = b - T + \alpha(\overline{V} - V_U). \tag{14}$$

It seems natural to calculate $\overline{V}$ as the employment-weighted average of the firm value to a worker of being employed. Denoting the aggregate unemployment rate by u and using the fact that aggregate employment is $(1-u)L = \int_0^1 l(m)M\mu(m)dm$, we can write $\overline{V}$ as

$$\overline{V} = \frac{M}{(1-u)L} \int_0^1 V(m)l(m)\mu(m)dm,$$

[11] When the support of $G(m)$ is bounded from below by some $m_{\min} > 0$, $J(\cdot)$ is monotonically decreasing from $J(m_{\min})$ to zero on $[m_{\min}, 1]$. Then a unique intersection of the curve $fJ(m)/(r+\delta)$ with the f_E line exists if and only if $J(m_{\min}) > (r+\delta)f_E/f$. The graphical representation of the determination of the equilibrium cutoff in Figure 1 is adapted from Demidova (2008).

where, from (1),

$$V(m) = \frac{1}{r+\delta}\left(w - T - \frac{e(m)^\gamma}{\gamma} + \delta V_U\right).$$

The expression for $\overline{V}$ accounts for the fact that firms with a higher monitoring ability have higher employment levels. This implies that when an abler firm is hit by a bad shock, more workers are displaced from their jobs. It also means: the abler a newly entering firm is, the more job offers will arrive at the pool of the unemployed. [12]

The only component of $V(m)$ that differs across firms is the effort cost per unit time $e(m)^\gamma/\gamma$. For calculating $\overline{V}$, it therefore suffices to calculate the employment-weighted average of $e(m)^\gamma$, denoted by $\overline{e^\gamma}$. Using (7), (9), and (10), we obtain (see Appendix):

$$\overline{e^\gamma} = w\psi(m^*), \qquad \text{where} \quad \psi(m^*) \equiv \frac{\int_{m^*}^1 m^{\gamma+\sigma-1} g(m)dm}{\int_{m^*}^1 m^{\sigma-1} g(m)dm} < 1 \quad \text{for} \quad m^* < 1. \tag{15}$$

Using this, we get for a worker's average surplus from being employed (referred to as average job rent)

$$\overline{V} - V_U = \frac{1}{r+\delta}\left[w\left(1 - \frac{\psi(m^*)}{\gamma}\right) - T - rV_U\right]. \tag{16}$$

Note that $\psi(m^*)/\gamma$ represents the ratio of the average cost of effort to the wage, $(\overline{e^\gamma}/\gamma)/w$, implying that the latter is solely determined by the monitoring ability cutoff m^*.

In equilibrium no worker shirks because each firm sets the wage and its specific effort standard so as to deter shirking. Accordingly, at each instant of time, $\delta(1-u)L$ employed workers become unemployed because their jobs break up, while $\alpha u L$ unemployed get re-employed. In steady state the flows into and out of unemployment must balance, which implies

$$\alpha = \frac{\delta(1-u)}{u}. \tag{17}$$

Substituting (17) for α into (14) and combining (14) and (16), we obtain

$$\overline{V} - V_U = \frac{w\left(1 - \frac{\psi(m^*)}{\gamma}\right) - b}{r + \delta/u}. \tag{18}$$

The equation tells us that the average job rent equals a worker's average flow surplus from having a job times a discount factor, $1/(r+\delta/u)$. The average flow surplus from employment is increasing in w and decreasing in ψ. A higher unemployment rate raises the average job rent because it lowers the rate at which workers discount the average flow surplus from employment. Intuitively, when unemployment rises, there are less workers that are separated from their jobs. It is therefore more difficult for the unemployed to get re-employed. This raises the value of employment to a worker relative to unemployment.

In equilibrium the no-shirking condition (2) holds with equality, stating that a worker's job rent at a firm equals the gain from shirking at this firm. As a consequence, it must also hold

[12] Calculating $\overline{V}$ instead as the simple expected value of $V(m)$, $EV = \int_0^1 V(m)\mu(m)dm$ would fail to account for the differences across firms in the capacity of creating jobs.

that the average job rent is equal to the employment-weighted average gain from shirking (for short referred to as average gain from shirking). Rewriting (2) in terms of m, using (7), and taking employment-weighted averages yields [13]

$$\overline{V} - V_U = \frac{w[1-\psi(m^*)]}{\gamma(r+\delta)}. \tag{19}$$

Note that the average gain from shirking is increasing in w and decreasing in ψ. The first property is an immediate consequence of the fact that a higher wage allows each firm to require greater effort from its workers. The second property derives from the fact that a worker's gain from shirking is decreasing in the firm's monitoring ability. To see the reason for this, note that a higher detection rate q implies that workers are discounting the flow benefit from shirking at a higher rate, which lowers the gain from shirking. On the other hand, a higher q allows a firm to require greater effort at a given wage, thereby raising the gain from shirking. However, given our specification of the disutility of effort, the latter effect is never strong enough to outweigh the former. [14]

From (18) and (19) it follows that

$$\frac{w\left(1-\frac{\psi(m^*)}{\gamma}\right)-b}{r+\delta/u} = \frac{w[1-\psi(m^*)]}{\gamma(r+\delta)}. \tag{20}$$

We call this the aggregate no-shirking condition, borrowing the term introduced by Shapiro and Stiglitz (1984). Solving (20) for u yields

$$u = \frac{\delta w(1-\psi(m^*))}{rw(\gamma-1)+\delta w(\gamma-\psi(m^*))-(r+\delta)\gamma b} = \frac{\delta(1-\psi(m^*))}{\delta(1-\psi(m^*))+(r+\delta)[\gamma(1-\varrho)-1]}, \tag{21}$$

where $\varrho = b/w$ is the benefit replacement rate ($0 < \varrho < 1$). [15] In the sequel we assume that ϱ is a fixed policy parameter, being a measure of a country's generosity of unemployment compensation. With ϱ fixed , in (20) w cancels out. As a consequence, (20) alone determines

---

[13] By using (7) the firm-level gain from shirking (the right-hand side of (2)) can be rewritten as

$$\frac{e(q)^\gamma}{\gamma q} = \frac{(1-m^\gamma)e(m)^\gamma}{\gamma(r+\delta)m^\gamma} = \frac{(1-m^\gamma)w}{\gamma(r+\delta)}.$$

The employment-weighted average gain from shirking is calculated as

$$\frac{w}{\gamma(r+\delta)}\frac{M}{(1-u)L}\int_0^1 (1-m^\gamma)l(m)\mu(m)dm = \frac{w(1-\psi(m^*))}{\gamma(r+\delta)},$$

where use has been made of (9), (10), and $(1-u)L = Ml(\widetilde{m})$. Alternatively, (19) can be derived as follows. From (6), $rV_U = [(\gamma-1)/\gamma]w - T$. Substitute this for rV_U into (16) to get (19).

[14] An increase in q by one percent, raises e^γ/γ by $1-m^\gamma < 1$ percent so that $e^\gamma/(\gamma q)$ falls.

[15] The term $\gamma(1-\varrho)-1$ in the the second expression for u is positive. To see this, note that using (6) and (14) we get $\gamma(1-\varrho)-1 = (\gamma\alpha/w)(\overline{V}-V_U) > 0$: for all active firms the job rent $V(m)-V_U$ must be positive because otherwise workers would not be willing to work. Hence, the average job rent $\overline{V}-V_U$ must be positive too.

the equilibrium unemployment rate. [16] This implies that the equilibrium unemployment rate is independent of the size of the labour force L, and, given ϱ, is (through ψ) solely determined by the monitoring ability cutoff m^*. In addition, it is increasing in ϱ and decreasing in the value of ψ and hence in m^*. [17] The former property is quite intuitive. The average flow surplus from employment declines as ϱ increases while the average gain from shirking remains unaltered. For keeping the average job rent at its initial level, the unemployment rate must rise. To see the intuition for u being decreasing in ψ, it helps to contrast the relative changes in the average job rent and the average gain from shirking resulting from an increase in the value of ψ. In the case of a fixed ϱ the relative change in the average job rent is

$$\frac{d(\overline{V}-V_U)}{\overline{V}-V_U} = \frac{dw}{w} - \left(\frac{\psi}{\gamma(1-\varrho)-\psi}\right)\frac{d\psi}{\psi} + \left(\frac{\delta}{ru+\delta}\right)\frac{du}{u}.$$

The relative change in the average gain from shirking (AGS) is given by

$$\frac{d\,AGS}{AGS} = \frac{dw}{w} - \left(\frac{\psi}{1-\psi}\right)\frac{d\psi}{\psi}.$$

The expressions show that, holding u and w constant, an increase in ψ reduces the average job rent by a smaller percentage than the average gain from shirking, giving rise to slack in the aggregate no-shirking condition. [18] However, maintaining the aggregate no-shirking condition as an equality requires average job rent and average gain from shirking to change by the same percentage. For this to occur, the unemployment rate must decline. The reason is that a lower u raises the rate at which the average flow surplus from a job is discounted. Intuitively, a lower u makes it easier for an unemployed to get a job because with fewer unemployed workers more existing jobs are destroyed. This lowers the average job rent while leaving the average gain from shirking unaffected.

Let us now turn to the determination of the equilibrium values of the wage and the mass of firms. The equilibrium wage can be determined from aggregate labour demand, a relationship between aggregate employment, the cutoff m^*, and the wage (see Janiak, 2008). To derive it, recall that aggregate employment is $(1-u)L = \int_0^1 l(m)M\mu(m)dm$. Using (9) and (10), it can be rewritten as $(1-u)L = Ml(\widetilde{m})$. Next consider the equilibrium price index associated with Y, which is given by

$$P = \left[\int_0^1 p(m)^{1-\sigma}M\mu(m)dm\right]^{1/(1-\sigma)}.$$

[16] This also holds true when the unemployed receive no income ($b = 0$), which is the case considered by Davis and Harrigan (2011). When unemployment benefits b are fixed rather than being linked to w, (20) defines a relationship between w and u for a given m^*. In that case the equilibrium values of w and u are determined by (20) and a second relationship between these variables, the aggregate labour demand schedule, introduced shortly.

[17]
$$\psi'(m) = \frac{g(m)m^{\sigma-1}}{\int_m^1 \xi^{\sigma-1}g(\xi)d\xi}\left[\psi(m)-m^{\gamma}\right].$$

Noting $\int_m^1 \xi^{\gamma+\sigma-1}g(\xi)d\xi > m^{\gamma}\int_m^1 \xi^{\sigma-1}g(\xi)d\xi$ for $m<1$, it follows that $\psi'(m) > 0$ for $m < 1$.

[18] This follows from $\gamma(1-\varrho)-1>0$. See footnote 15.

Using $p(m) = (\widetilde{m}/m)p(\widetilde{m})$ yields $P = M^{1/(1-\sigma)}p(\widetilde{m})$; and from $P = 1$, this implies that the equilibrium mass of firms is

$$M = [p(\widetilde{m})]^{\sigma-1}. \tag{22}$$

Using this together with (7)–(9) and the zero cutoff profit condition, we find that aggregate labour demand is given by

$$(1-u)L = \frac{f\theta}{1-\theta}(\theta m^*)^{1-\sigma} w^{\frac{(\gamma-1)(\sigma-1)-\gamma}{\gamma}}. \tag{23}$$

Equation (23) tells us that it depends on the sign of $(\gamma-1)(\sigma-1)-\gamma$ whether aggregate labour demand increases or decreases with the wage. To give an explanation, we use the aforementioned fact that aggregate employment can be expressed as $(1-u)L = Ml(\widetilde{m})$, where $l(\widetilde{m})$ is the employment level at the representative firm. The latter can be written as [19]

$$l(\widetilde{m}) = \left(\frac{\widetilde{m}}{m^*}\right)^{\sigma-1} \frac{f\theta}{(1-\theta)w}.$$

The expression shows that the representative firm's employment level is a decreasing function of the wage (as is the employment level at any active firm), the wage elasticity of firm-level labour demand being equal to -1. Let us now consider the relationship between the mass of firms M and the wage. Substituting the pricing rule (8) into (22) yields

$$M = \frac{w^{\frac{(\gamma-1)(\sigma-1)}{\gamma}}}{(\theta\widetilde{m})^{\sigma-1}},$$

which shows that the mass of firms increases with the wage, with the wage elasticity of M given by $(\gamma-1)(\sigma-1)/\gamma$. This reflects the fact that the assumed CES technology of final good production exhibits external economies of scale with respect to the mass of firms (input varieties) M: the larger the mass of input varieties, the higher the average productivity of intermediate inputs, which equals the price charged by the representative intermediate input producer, $p(\widetilde{m}) = M^{1/(\sigma-1)}$. [20] In what follows we assume that $(\gamma-1)(\sigma-1)/\gamma > 1$, which is equivalent to

Assumption 1. $\sigma > (2\gamma-1)/(\gamma-1)$.

Assumption 1 is tantamount to assuming that the wage is increasing with aggregate labour demand. That is, the positive relationship between the mass of firms and the wage dominates

---

[19] Using (7)–(9) and $R(m^*) = \sigma f$ (from the zero cutoff profit condition), we get for $m \geq m^*$:

$$l(m) = \left(\frac{m}{m^*}\right)^{\sigma-1} \frac{f\theta}{(1-\theta)w},$$

where $f\theta/(1-\theta)w = l(m^*)$ is the employment level at a m^*-firm that just breaks even.

[20] Using (9) in $Y = [\int_0^1 x^\theta M\mu(m)dm]^{1/\theta}$ yields $Y = M^{\sigma/(\sigma-1)}x(\widetilde{m}) = M^{1/(\sigma-1)}Mx(\widetilde{m})$, where $Mx(\widetilde{m})$ and $M^{1/(\sigma-1)}$ represent, respectively, total use and average productivity of intermediate inputs.

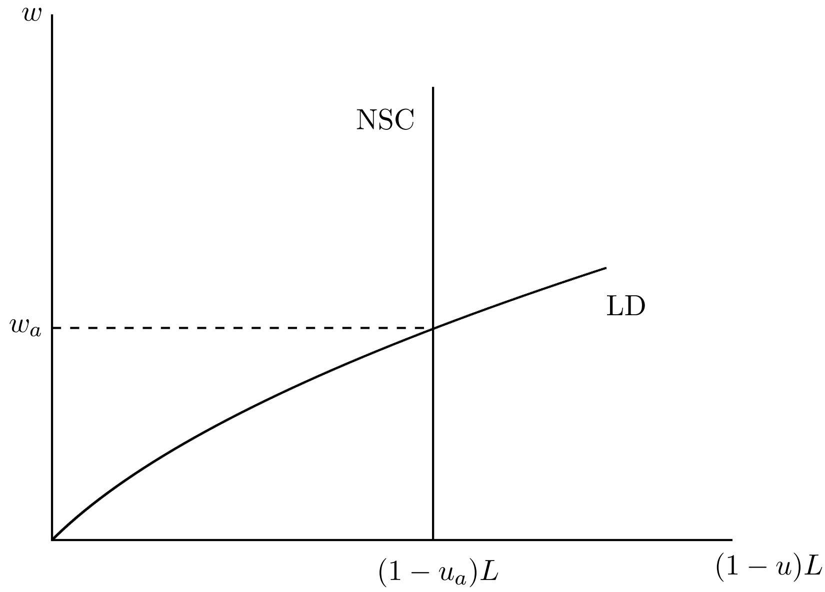

FIGURE 2. Determination of the equilibrium wage

the negative influence of the wage on firm-level labour demand. Assumption 1 requires that $\sigma > 2$. [21]

The determination of the equilibrium wage is illustrated in Figure 2. With a fixed benefit replacement rate the equilibrium unemployment rate is determined by the aggregate no-shirking condition (21) and solely depends on the equilibrium cutoff m^*, as shown above. It corresponds to the vertical line NSC at $(1-u_a)L$, where u_a is the equilibrium unemployment rate. Aggregate labour demand corresponding to m^* (equation (23)) is depicted by the curve LD, which under Assumption 1 slopes upward. The intersection of LD and NSC determines the equilibrium wage w_a.

The aggregate monitoring ability level $\widetilde{m}$ is determined by the cutoff level m^*. Therefore, once the wage is known, the price charged by the representative firm, $p(\widetilde{m})$, can be determined from the pricing rule (8). And, from (22), this pins down the equilibrium mass of firms.

As for the accounting for final output Y, the following is worth noting. First, from $P = 1$, aggregate revenue of the intermediate input producers equals final output Y. [22] Second, when the government's role is limited to providing and financing unemployment benefits, a balanced

---

[21] Assumption 1 seems plausible. The parameter γ may be considered as being related to the intertemporal elasticity of substitution in labour supply, which is given by $1/(\gamma - 1)$. Estimates of the latter are within the -0.07 to 0.45 range, the average of which implies a value of γ greater than 6. See Pencavel (1986, Table 1.22). This corresponds to a value of $(2\gamma - 1)/(\gamma - 1)$ smaller than $11/5$. In their calibration of the US manufacturing sector Bernard, Eaton, Jensen, and Kortum (2003) set $\sigma = 3.8$. Given plausible values of γ, this choice of σ clearly satisfies the requirement $\sigma > (2\gamma - 1)/(\gamma - 1)$. For a discussion of the form of aggregate labour demand in several variants of the Melitz (2003) model see Janiak (2008).

[22] This follows from $Y = MM^{1/(\sigma-1)}x(\widetilde{m})$, and the fact that average revenue is $\overline{R} = R(\widetilde{m}) = p(\widetilde{m})x(\widetilde{m}) = M^{1/(\sigma-1)}x(\widetilde{m})$.

government budget requires that $\varrho wuL = TL$. In the accounting for Y aggregate tax revenue and aggregate unemployment benefits therefore cancel out. As a result, final output equals the sum of the aggregate wage bill, aggregate operating profits, and aggregate fixed costs: $Y = (1-u)Lw + M(\overline{\pi} + f)$. Third, in a stationary equilibrium, aggregate operating profits Π equal the sum of total entry costs and total investment income from financing firms. To see this, note that total entry costs per unit time are $M_E f_E$, where M_E denotes the mass of entrants per unit time. As the mass of firms must remain unchanged over time, we have $[1 - G(m^*)]M_E = \delta M$, i.e., at each instant of time the mass of successful entrants must equal the mass of incumbent firms that exit. Using this, the free entry condition (12), and $\Pi = M\overline{\pi}$, we get $M_E f_E = \delta\Pi/(r+\delta)$ (see Melitz, 2003). Summing up $M_E f_E$ and total investment income, $r\Pi/(r+\delta)$, yields aggregate operating profit Π. This completes the description of the unique equilibrium in a closed economy.

4 Open Economy

We now consider a world of two identical countries of the type described above that have opened up to trade. [23] We assume that only intermediate goods are tradable while the final good is nontraded. The price index associated with the final good is the same across countries, as are all the aggregate variables. Hence, we can still use the final good as a numéraire, with the price index normalized to one in each country.

4.1 Product Markets

Consider an intermediate input producer that after his monitoring ability draw has chosen to take up production. In an open economy he has to decide whether to export or serve the domestic market only. If he decides to export, he has to bear an extra fixed cost f_x per unit time, measured in units of the final good. In addition, there are per-unit trade costs of the melting-iceberg type: $\tau > 1$ units must be shipped for one unit to arrive abroad.

If the firm chooses to serve both the domestic and the foreign market, it will allocate any given output across markets so as to maximize total revenue. This requires equalization of marginal revenues across markets and, given the constant-elasticity demand for intermediates, the same producer price for domestic and foreign sales, denoted by $p(m)$. Due to variable trade costs, in the foreign country a domestically produced input variety is then priced at $\tau p(m)$, and in order to deliver x units abroad, a firm has to manufacture τx units. Also note that Y represents final output in any country. Variables relating to domestic sales are denoted by subscript d while subscript x is used to denote variables relating to sales abroad. Domestic demand and foreign demand (in consumption units) for a m-type intermediate input are then given by $x_d(m) = Yp(m)^{-\sigma}$ and $x_x(m) = Y[\tau p(m)]^{-\sigma} = \tau^{-\sigma} x_d(m)$, respectively. Correspondingly, revenues from

[23] Limiting the analysis to two countries is not essential to the paper's results. The analysis could easily be extended to a world comprised of more than two symmetric countries, as considered in Melitz (2003).

domestic and export sales are $R_d(m) = Y(p(m))^{1-\sigma}$ and $R_x(m) = Y(\tau p(m))^{1-\sigma} = \tau^{1-\sigma}R_d(m)$.

The labour market conditions faced by an intermediate input firm are as in a closed economy. Thus, in choosing the producer price and the effort to be asked of its workers the firm still takes into account the no-shirking condition (3). The first order conditions for a profit-maximizing choice of these variables are the same as in the closed economy case (equations (4) and (5)), as are its implications. All intermediate input producers pay the same wage, which as before is given by (6). According to (7) and (8), respectively, the effort put forth by workers at a m-firm is $e(m) = mw^{1/\gamma}$, and the producer price it charges (being equal to the customer price for domestic sales) is $p(m) = w^{(\gamma-1)/\gamma}/\theta m$.

A firm's operating profit $\pi(m)$ can be separated into two portions: profits from domestic sales, $\pi_d(m) = R_d(m)/\sigma - f$, and profits from export sales, $\pi_x(m) = R_x(m)/\sigma - f_x$, where in the former the fixed production cost f is accounted for. [24] For firms to produce for the domestic market, their monitoring ability must be at least as high as m^*, defined by $\pi_d(m^*) = 0$. And for firms to export, their monitoring ability must be at least as high as m_x^*, defined by $\pi_x(m_x^*) = 0$, where $m_x^* \geq m^*$. Following Melitz (2003), henceforth we confine attention to equilibria in which $m_x^* > m^*$, i.e., there is a partitioning of firms by export status. Firms with $m \in [m^*, m_x^*)$ serve only the domestic market while firms with $m \in [m_x^*, 1)$ serve both the domestic and the foreign market.

The ex ante probability of successful entry is as before $1-G(m^*)$; and the ex ante probability that a successful entrant will export is $\chi = [1 - G(m_x^*)]/[1 - G(m^*)]$, being equal to the ex post fraction of exporting firms in a country. That is, if M is a country's mass of active intermediate input producers and M_x its mass of exporters, then we have $M_x = \chi M$. On the other hand, $M_t = M + M_x$ represents the total mass of input varieties available to the final good sector in a country.

Applying the formula for the aggregate monitoring ability in (10), define $\widetilde{m} \equiv \widetilde{m}(m^*)$ and $\widetilde{m}_x \equiv \widetilde{m}(m_x^*)$. Then by analogy to the closed economy, $\pi_d(\widetilde{m})$ represents the average operating profit earned by firms in a country from domestic sales, while $\pi_x(\widetilde{m}_x)$ represents the average operating profit earned by exporting firms from sales abroad. Correspondingly, the overall average operating profit across all domestic firms in a country is $\overline{\pi} = \pi_d(\widetilde{m}) + \chi\pi_x(\widetilde{m}_x)$. Using (9) and the zero cutoff profit conditions $\pi_d(m^*) = 0$ and $\pi_x(m_x^*) = 0$, this can be rewritten as

$$\overline{\pi} = f\left[\left(\frac{\widetilde{m}(m^*)}{m^*}\right)^{\sigma-1} - 1\right] + \chi f_x\left[\left(\frac{\widetilde{m}_x(m_x^*)}{m_x^*}\right)^{\sigma-1} - 1\right]. \tag{24}$$

The free entry condition continues to be

$$[1 - G(m^*)]\frac{\overline{\pi}}{r+\delta} = f_E.$$

[24] Separating profits this way is possible because production involves incurring the fixed cost f anyway. Hence, each exporting firm will also serve the domestic market since it brings about extra operating profit; see Melitz (2003).

Inserting (24) for $\overline{\pi}$ gives

$$\frac{f}{r+\delta}J(m^*) + \frac{f_x}{r+\delta}J(m_x^*) = f_E, \tag{25}$$

where as before $J(m) = [1-G(m)][(\widetilde{m}(m)/m)^{\sigma-1}-1]$. The present-discounted value of expected profits upon entry (the left-hand side of (25)) now has two components: the present-discounted values of expected profits from domestic sales and from sales abroad.

A further implication of the zero cutoff profit conditions $\pi_d(m^*) = 0$ and $\pi_x(m_x^*) = 0$ is that the monitoring ability cutoffs m_x^* and m^* are linked as follows:

$$m_x^* = \tau \left(\frac{f_x}{f}\right)^{\frac{1}{\sigma-1}} m^*. \tag{26}$$

From (26) it follows that the partitioning of firms by export status ($m_x^* > m^*$) is equivalent to $\tau(f_x/f)^{1/(\sigma-1)} > 1$, in the same way as in Melitz (2003).

The aggregate monitoring ability representative of all domestic and foreign intermediate input producers active in a country can be defined as

$$\widetilde{m}_t \equiv \left\{\frac{1}{M_t}\left[M\widetilde{m}^{\sigma-1} + M_x\left(\tau^{-1}\widetilde{m}_x\right)^{\sigma-1}\right]\right\}^{\frac{1}{\sigma-1}}, \tag{27}$$

which is the perfect analogue to the aggregate productivity in Melitz (2003). The aggregate monitoring ability $\widetilde{m}_t$ plays a similar role as does $\widetilde{m}$ in the closed economy. That is, the aggregate variables can be expressed as functions of $\widetilde{m}_t$ and the mass of input varieties available in a country, M_t. In particular we find that the price index P, final output Y, aggregate revenue TR, and aggregate employment $(1-u)L$ can be written as (see Appendix):

$$\begin{aligned} P &= M_t^{1/(1-\sigma)}p(\widetilde{m}_t), & Y &= M_t^{\sigma/(\sigma-1)}x_d(\widetilde{m}_t), \\ TR &= M_t R_d(\widetilde{m}_t) = Y, & (1-u)L &= M_t l_d(\widetilde{m}_t). \end{aligned} \tag{28}$$

Given that the final good is taken as the numéraire in each country ($P = 1$), the first equation in (28) implies that

$$M_t = [p(\widetilde{m}_t)]^{\sigma-1}.$$

Using this together with (7)–(9), and $R_d(m^*) = \sigma f$ in $(1-u)L = M_t l_d(\widetilde{m}_t)$, we find that the aggregate labour demand schedule for a country is still given by (23) as in the closed economy. Finally, using the facts that $TR = Y$ and $\overline{R} = \sigma(\overline{\pi} + f + \chi f_x)$, we get [25]

$$M = \frac{TR}{\overline{R}} = \frac{Y}{\sigma(\overline{\pi} + f + \chi f_x)}. \tag{29}$$

---

[25] The expression for $\overline{R}$ obtains by using $\overline{R} = R_d(\widetilde{m}) + \chi R_x(\widetilde{m}_x)$ and $\overline{\pi} = \pi_d(\widetilde{m}) + \chi\pi_x(\widetilde{m}_x)$ together with $R_d(m) = \sigma[\pi_d(m) + f]$ and $R_x(m) = \sigma[\pi_x(m) + f_x]$.

4.2 Aggregate No-Shirking Condition

Let $\mu_x(m)$ be the density conditional on exporting, with $\mu_x(m) = g(m)/[1 - G(m_x^*)]$ if $m \geq m_x^*$ and zero otherwise. Then the employment-weighted average value to a worker of being employed is

$$\overline{V} = \frac{1}{(1-u)L}\left[\int_0^1 V(m) l_d(m) M\mu(m)dm + \int_0^1 V(m) l_x(m) M_x \mu_x(m)dm\right],$$

where the expression for $V(m)$ is the same as in the closed economy, and $l_d(m) = x_d(m)/e(m)$ and $l_x(m) = \tau x_x(m)/e(m)$ are the employment levels at a m-firm for, respectively, domestic sales and (possibly) exports. Once again, the cost of effort $e(m)^\gamma/\gamma$ is the only component of $V(m)$ to vary with m. By using the function $\psi(\cdot)$ defined in (15) the employment-weighted average $\overline{e^\gamma}$ is calculated as (see Appendix):

$$\overline{e^\gamma} = w\psi_t, \tag{30}$$

where

$$\psi_t = \frac{fa(m^*)\psi(m^*) + f_x a(m_x^*)\psi(m_x^*)}{fa(m^*) + f_x a(m_x^*)} = \varpi_d \psi(m^*) + \varpi_x \psi(m_x^*),$$

with $a(m) \equiv m^{1-\sigma}\int_m^1 \xi^{\sigma-1} g(\xi)d\xi$, and $\varpi_d \equiv fa(m^*)/[fa(m^*) + f_x a(m_x^*]$ and $\varpi_x = 1 - \varpi_d$ being a country's shares of domestic sales and exports in revenue. [26]

The remaining steps in deriving the aggregate no-shirking condition are the same as for the closed economy. The asset equation (14) for an unemployed and the flow condition (17) continue to hold in an open economy. The average job rent takes the form $\overline{V} - V_U = (r+\delta)^{-1}[w(1 - \psi_t/\gamma) - T - rV_U]$. Combining these facts and taking account of a fixed benefit replacement rate yields

$$\overline{V} - V_U = \frac{w\left(1 - \frac{\psi_t}{\gamma} - \varrho\right)}{r + \delta/u}.$$

On the other hand, the average job rent must be equal to the average gain from shirking: [27]

$$\overline{V} - V_U = \frac{w(1-\psi_t)}{\gamma(r+\delta)}.$$

Equating the two expressions for $\overline{V} - V_U$ gives the aggregate no-shirking condition for the open economy, and solving it for u yields

$$u = \frac{\delta(1-\psi_t)}{\delta(1-\psi_t) + (r+\delta)[\gamma(1-\varrho) - 1]}. \tag{31}$$

As (31) reveals, the expression for the equilibrium unemployment rate in a country is formally identical with that in (21) holding for the closed economy, with $\psi(m^*)$ being replaced by ψ_t.

---

[26] The use of the functions $a(\cdot)$ is adopted from Demidova (2008).

[27] The average gain from shirking is calculated as

$$\frac{w}{\gamma(r+\delta)(1-u)L}\left[\int_0^1 (1-m^\gamma) l_d(m) M\mu(m)dm + \int_0^1 (1-m^\gamma) l_x(m) M_x \mu_x(m)dm\right] = \frac{w(1-\psi_t)}{\gamma(r+\delta)}.$$

4.3 Equilibrium in an Open Economy

We now turn to the determination of a trading equilibrium. The cutoff levels m^* and m_x^* are determined by equations (25) and (26). Substituting (26) for m_x^* into (25) leads to

$$\frac{f}{r+\delta}J(m^*) + \frac{f_x}{r+\delta}J(Am^*) = f_E, \tag{32}$$

where $A \equiv \tau(f_x/f)^{1/(\sigma-1)}$. Recall that, from (26), partitioning of firms by export status ($m_x^* > m^*$) is equivalent to $A > 1$. Also note that $J(Am^*)$ is defined for $0 < m^* \leq 1/A$ as $m_x^* = Am^*$ is bounded from above by one. Consider the case where $G(m)$ has support over $(0, 1)$, in which case $J(m)$ is monotonically decreasing from infinity to zero on $(0, 1)$. Hence, the left-hand side of (32) also monotonically decreases from infinity to zero on $(0, 1)$. This ensures that there exists a unique cutoff level m^* that solves equation (32). It does not ensure, however, that m^* is low enough to be compatible with an export cutoff m_x^* smaller than one. When the entry cost f_E is relatively small, expected profits have to be small as well, which requires a high level of the domestic cutoff m^*. Then it may well happen that a cutoff $m_x^* < 1$ that makes exporting profitable fails to exist. That is, in such cases a trading equilibrium does not exist so that the countries are bound to remain in the state of a closed economy. In what follows we rule out such cases. [28]

Once the domestic cutoff m^* is known, the export cutoff m_x^* can be calculated from (26). The equilibrium values of the remaining variables are determined in a similar way as for the closed economy. According to (31), at a given replacement rate ϱ a country's unemployment rate depends on ψ_t, which is uniquely determined by m^* and m_x^*. With u determined, so too are, by (23), a country's wage and hence, by $Y = (1-u)Lw/\theta$, its final output. Furthermore, as in the Melitz (2003) model, the cutoffs m^* and m_x^* pin down average operating profit $\overline{\pi}$ and the fraction χ of exporting firms. This, by (29), determines a country's mass of intermediate input producers as well as the mass of input varieties available in a country.

5 The Impact of Trade

This section considers the effects of trade by comparing the equilibria in the closed and open economy. Variables relating to the autarky equilibrium are denoted by subscript a, while variables relating to the trading equilibrium are denoted the same way as in the previous section, with a subscript t added when expedient. We begin by examining how trade affects a country's unemployment rate, wage, and aggregate factor productivity in the intermediate goods sector. Thereafter, we consider the welfare gains and losses of workers.

[28] As $J(Am^*)$ decreases from infinity to zero on $(0, 1/A]$, for $m^* = 1/A$ the left-hand side of (32) takes the value $fJ(1/A)/(r+\delta)$. Hence, a condition necessary and sufficient for m^* to support a trading equilibrium is $f_E > fJ(1/A)/(r+\delta)$.

5.1 The Impact on Unemployment, Wages, and Productivity

A first step is to examine how a country's cutoff level of monitoring ability is affected by the move from autarky to trade. Using (13) and (25), we get

$$fJ(m_a^*)/(r+\delta) = f_E > f_E - f_x J(m_x^*)/(r+\delta) = fJ(m^*)/(r+\delta).$$

The existence of a trading equilibrium requires that $m_x^* < 1$. Since $J(\cdot)$ is a decreasing function, it therefore follows that $m^* > m_a^*$, i.e., the domestic monitoring ability cutoff rises after the opening to trade.

How does the higher monitoring ability cutoff influence a country's unemployment rate and wage? As shown above, with a fixed benefit replacement rate, a country's unemployment rate is determined by the aggregate no-shirking condition, given by (21) in autarky and (31) with trade. Therefore, the impact of trade on unemployment can be inferred from comparing the value of ψ_t with the value of $\psi(m_a^*)$. From the definition of ψ_t in (30), we have $\psi_t = \psi(m^*) + \varpi_x[\psi(m_x^*) - \psi(m^*)]$. Recall that $\psi(\cdot)$ is an increasing function. Using this together with the assumption of partitioning of firms by export status ($m_x^* > m^*$), it follows from $m^* > m_a^*$ that $\psi_t > \psi(m^*) > \psi(m_a^*)$. Hence we conclude that a move from autarky to trade reduces unemployment in a country.

The effects of trade on a country's employment and its wage are shown in Figure 3. The vertical lines NSC_a at $(1-u_a)L$ and NSC_t at the higher employment level $(1-u_t)L$ depict the aggregate no-shirking conditions (21) and (31) in autarky and with trade, respectively. With the change in employment determined, the impact of trade on a country's equilibrium wage can be seen from aggregate labour demand, given by (23). The curve LD_a depicts aggregate labour demand for the cutoff level m_a^* in autarky. Under Assumption 1 the labour demand curve slopes upward for any cutoff level m^*. Moreover, equation (23) reveals that aggregate labour demand is decreasing in m^*. Therefore, by raising the domestic cutoff level, a move from autarky to trade causes an upward shift in aggregate labour demand from LD_a to LD_t. Since the no-shirking condition line shifts to the right, we find that trade leads to a higher wage in a country.

Next, we examine the impact of trade on a country's total factor productivity in the intermediate goods sector. Following Helpman and Itskhoki (2010), we use as a measure the employment-weighted average of firm-level productivity. In our framework a m-firm's productivity level equals the level of effort $e(m)$ provided by its workers. Noting that $x_a(m) = e_a(m)l_a(m)$, in a closed economy our measure of total factor productivity is thus given by

$$TFP_a = \frac{1}{(1-u_a)L}\int_{m_a^*}^{1} x_a(m)M_a\mu(m)dm.$$

As shown in the Appendix, this can be rewritten as

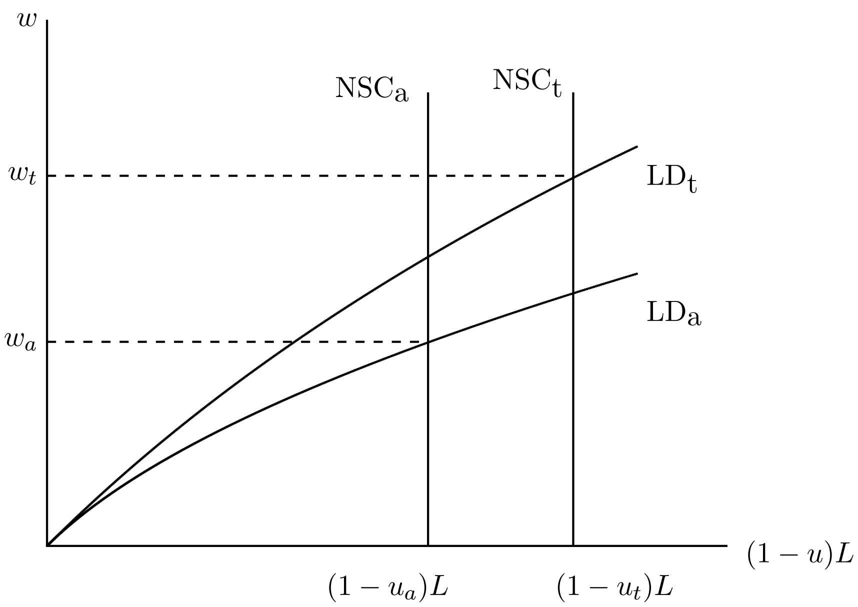

FIGURE 3. Impact of trade on the wage and employment

$$TFP_a = w_a^{1/\gamma}\phi(m_a^*), \tag{33}$$

where

$$\phi(m) \equiv \frac{\int_m^1 \xi^\sigma g(\xi)d\xi}{\int_m^1 \xi^{\sigma-1} g(\xi)d\xi} < 1 \quad \text{for} \quad m < 1.$$

The expression $\phi(m_a^*)$ represents the average monitoring ability of firms in a closed economy.

Similarly, a country's total factor productivity in an open economy is given by

$$TFP = \frac{1}{(1-u)L}\left[\int_{m^*}^1 x_d(m)M\mu(m)dm + \int_{m_x^*}^1 \tau x_x(m)M_x\mu_x(m)dm\right].$$

This can be rewritten as (see Appendix):

$$TFP = w^{1/\gamma}\phi_t, \tag{33a}$$

where

$$\phi_t = \frac{fa(m^*)\phi(m^*) + f_x a(m_x^*)\phi(m_x^*)}{fa(m^*) + f_x a(m_x^*)} = \varpi_d\phi(m^*) + \varpi_x\phi(m_x^*) < 1.$$

The expressions $\phi(m^*)$ and $\phi(m_x^*)$ represent the average monitoring ability of, respectively, firms producing for the domestic market and exporting firms. Overall average monitoring ability ϕ_t is a weighted average of $\phi(m^*)$ and $\phi(m_x^*)$, where the weights are the shares of domestic sales and exports in revenue. [29]

The effect of trade on total factor productivity is revealed by inspection of (33) and (33a). The expression for ϕ_t can be written as $\phi_t = \phi(m^*) + \varpi_x[\phi(m_x^*) - \phi(m^*)]$. Since $\phi(\cdot)$ is an

[29] Note that $\phi(m^*) \neq \widetilde{m}$, $\phi(m_x^*) \neq \widetilde{m}_x$, and $\phi_t \neq \widetilde{m}_t$, i.e., average monitoring abilities differ from the monitoring abilities of the firms being representative of those serving the domestic market, those that export, and of all domestic and foreign firms active in a country, respectively.

increasing function, partitioning of firms by export status ($m_x^* > m^*$) together with $m^* > m_a^*$ implies that $\phi_t > \phi(m^*) > \phi(m_a^*)$. [30] That is, trade leads to an increase in average monitoring ability of firms. In conjunction with the higher wage, this results in greater average worker effort and hence higher total factor productivity in the intermediate goods sector. Also note that, thanks to the higher wage, the level of effort and hence within-firm productivity rises in all firms that stay in the market after the transition to trade. These results are summarized in

Proposition 1. *Suppose that there is partitioning of firms by export status. Then the opening to trade lowers unemployment in each country; under Assumption 1 it raises a country's wage and total factor productivity in the intermediate goods sector.*

The mechanism driving the results is as follows. Trade offers prospects for additional profits to firms with higher monitoring ability through entry into the export market, thereby contributing to an increase in overall expected profits at the entry stage. But as entry cost is fixed, for the free entry condition to be restored, this has to be offset by lower expected profits from domestic sales. These come about because firms anticipate that the wage and hence the cost per unit of effective labour are going to rise. As a result, the domestic cutoff rises above its autarky level ($m^* > m_a^*$). Exposure to trade thus has the same kind of selection effects as in Melitz (2003). The least able firms with $m \in [m_a^*, m^*)$ exit, while firms with $m \in [m_x^*, 1)$ start producing for the export market. Both effects reallocate market shares towards firms with higher monitoring ability.

To see the intuition behind the fall in the unemployment rate caused by trade opening, it is helpful to look at its effects that would occur if both the unemployment rate and the wage were held constant at their initial levels. Since market shares are reallocated towards abler firms, employed workers on average have to work harder and thus incur higher effort costs per unit time relative to wages. As explained above, this reduces both the average job rent and the average gain from shirking. However, average job rent falls by a smaller percentage than average gain from shirking so that slack arises in the aggregate no-shirking constraint. Slack will also arise in the no-shirking constraints for individual firms. For the unemployed, a lower average surplus from employment entails a lower expected capital gain from re-employment and hence a lower flow value of being unemployed, rV_U. As a consequence, at the initial wage each firm sees its workers' job rent getting larger than needed to induce them to put forth effort. For restoring the balance of the gains from and the costs of shirking, both in the aggregate and at the firm level, average job rent and average gain from shirking need to be leveled out again. The only way this can occur is by a fall in the aggregate unemployment rate: lower unemployment makes it more likely for an unemployed to get re-employed. This raises the rate at which workers discount the average flow surplus from being employed, which tends to

---

[30]

$$\phi'(m) = \frac{g(m)m^{\sigma-1}}{\int_m^1 \xi^{\sigma-1} g(\xi) d\xi} \left[\phi(m) - m\right].$$

Using $\int_m^1 \xi^{\sigma-1} g(\xi) d\xi < m^{-1} \int_m^1 \xi^{\sigma} g(\xi) d\xi$ for $m < 1$ leads to $\phi'(m) > 0$ for $m < 1$.

reduce the average job rent, while leaving the average gain from shirking unaffected. Though the average job rent declines, the expected capital gain from re-employment, $\alpha(\overline{V} - V_U)$, rises when the unemployment rate falls. The reason is the positive effect on the job acquisition rate α, which dominates. [31] The flow value of being unemployed, rV_U, therefore rises when the unemployment rate declines, which lowers firm-level job rents. Since the gain from shirking is independent of u, this acts towards restoring the balance of incentives in individual firms.

The increase in aggregate employment comes from two sources: the shift in market shares towards abler firms, having larger payrolls, and the expansion of employment at the firms with the highest monitoring ability, selecting into exporting. It is worth noting that the increase in aggregate employment occurs in spite of forces that work against it. Each surviving firm (with $m \geq m^*$) employs less workers for domestic sales than it does under autarky. [32] That means, each firm with $m \in [m^*, m_x^*)$ will reduce its employment level compared to autarky. In addition, the change in aggregate employment depends on the change in the number of domestic intermediate input producers. Equation (22) that determines the mass of firms under autarky, M_a, can be rewritten as [33]

$$M_a = \frac{Y_a}{\sigma(\overline{\pi}_a + f)}.$$

Comparing this equation with (29) reveals that there are opposing effects of trade on a country's mass of firms. Final output Y is higher in the trading equilibrium, which tends to raise M; but both the higher average profit $\overline{\pi}$ and the extra fixed cost f_x to be borne by the exporting firms tend to reduce M. Simulations we performed using a truncated Pareto distribution suggest that the equilibrium number of domestic firms typically decreases after the move from autarky to trade. [34] This again tends to dampen the increase in aggregate employment.

5.2 Gains and Losses of Workers

We now consider the effects of trade on the welfare of workers employed by firms with different monitoring ability levels and of the unemployed. Welfare is measured by the expected lifetime utility of the respective group members. Recall that the value of being unemployed, V_U, satisfies

---

[31] Using

$$\alpha(\overline{V} - V_U) = \frac{\delta(1-u)}{ru + \delta}\left[w\left(1 - \frac{\psi_t}{\gamma}\right) - b\right]$$

shows that the expected capital gain is decreasing in u.

[32] With trade the employment level for domestic sales of a firm with $m \geq m^*$ can be expressed as $l_d(m) = (m/m^*)^{\sigma-1} f\theta/w(1-\theta)$, while its employment level under autarky is given by $l_a(m) = (m/m_a^*)^{\sigma-1} f\theta/w_a(1-\theta)$. See also footnote 19 above.

[33] From (22),

$$M = [p(\widetilde{m})]^{\sigma-1} = \frac{[p(\widetilde{m})]^{\sigma} x(\widetilde{m})}{p(\widetilde{m})x(\widetilde{m})} = \frac{M^{\sigma/(\sigma-1)} x(\widetilde{m})}{R(\widetilde{m})} = \frac{Y_a}{\sigma(\overline{\pi}_a + f)},$$

where use has been made of the fact that in autarky $Y = M^{\sigma/(\sigma-1)} x(\widetilde{m})$ and $R(\widetilde{m}) = \overline{R} = \sigma(\overline{\pi} + f)$.

[34] In Melitz (2003) trade unambiguously lowers M because in his model aggregate revenue is a constant.

(6) in autarky and with trade. Using this and $T = \varrho wu$ from the government budget constraint, we can express the change in rV_U as

$$\Delta rV_U = rV_U - rV_{Ua} = \left(\frac{\gamma - 1}{\gamma} - \varrho u\right) w - \left(\frac{\gamma - 1}{\gamma} - \varrho u_a\right) w_a,$$

where V_{Ua} denotes the value of being unemployed under autarky. As the unemployment rate falls and the wage goes up, it follows that the unemployed unambiguously gain from trade. [35]

The change in utility of workers who continue to be employed by firms with $m \geq m^*$ can most easily be derived by using the firm-level no-shirking condition (2). Holding with equality, the condition implies that the value $V(m)$ to a worker of being employed by a firm with ability m can be expressed as consisting of two parts: the value of being unemployed, V_U, plus a firm specific rent, being equal to the gain from shirking attainable at that firm. Rewriting (2) in terms of m and using (7) yields

$$\Delta V(m) = V(m) - V_a(m) = V_U - V_{Ua} + \frac{1 - m^\gamma}{(r + \delta)\gamma} (w - w_a).$$

As the value of being unemployed and the wage rise after opening to trade, so does the welfare of all workers who stay in employment. The utility change $\Delta V(m)$ is decreasing in m, i.e., the gain from trade is greatest for workers with jobs at surviving firms with the lowest monitoring ability.

As aggregate employment rises, there are also workers who are jobless in autarky but enter employment after the transition to trade. These workers clearly benefit from trade. Their gain consists of the increase in the value of unemployment, $V_U - V_{Ua}$ plus the job rent $(1 - m^\gamma)w/(r + \delta)\gamma$ they enjoy when employed.

What about the workers whose jobs are lost through trade? Affected are all workers who in autarky happened to be employed by firms with $m \in [m_a^*, m^*)$, which are forced to exit. In addition, some of the workers employed in autarky by firms with the next lowest monitoring ability (with $m \in [m^*, m_x^*)$) are displaced from their jobs since these firms contract employment after trade opening, as shown above. Good jobs from the workers' perspective are those requiring little effort, i.e. jobs at firms with low monitoring ability. Thus it is precisely these good jobs which do not survive the opening of trade (including the best jobs in existence). As already mentioned, this result resembles the one obtained by Davis and Harrigan (2011).

All workers displaced from their jobs pass through a period of unemployment before getting re-employed, thereby suffering at least a temporary welfare loss. However, less clear is the impact of job loss on the expected lifetime utility of the workers affected because there are opposing effects. Two cases can be distinguished: the case in which a worker displaced from his job stays unemployed; and the case in which he ends up with a new job at another firm. In either case the worker gains an amount of $V_U - V_{Ua}$ and loses a possibly high rent from the job he has been displaced from. When becoming re-employed, he additionally gains the (possibly

[35] Note that, from $rV_U > 0$, the expressions in brackets are positive.

lower) rent from his new job. As a result, a worker whose job is destroyed may (or may not) experience a utility loss, depending on the increase in the value of unemployment, the job rent lost, and, if re-employed, the rent he has in his new occupation. These findings are summarized in

Proposition 2. *Suppose that there is partitioning of firms by export status and Assumption 1 holds. Then: (i) The unemployed and workers who stay employed gain from trade, as do workers who move from unemployment to employment. (ii) The only workers who may lose are those whose jobs are destroyed; the jobs lost are the ones which in autarky carry relatively high rents to workers.*

6 Trade Liberalization

Two mechanisms of trade liberalization can be considered: a decrease in variable trade costs and a decrease in the fixed costs of exporting. Since both mechanisms induce similar effects, we focus on one of them, a decrease in variable trade costs τ. In particular we are interested in finding out to what extent the effects of a decrease in variable trade costs are in line with those induced by a move from autarky to trade. Thereafter we simulate the effects of variable trade costs on unemployment, wages, and total factor productivity in order to assess their magnitude.

6.1 Decrease in Variable Trade Costs

We begin by considering the effects of a change in τ on the cutoff levels m^* and m_x^*. As shown in Section 4.3, the equilibrium values of the cutoffs are determined by the free entry condition (25) and equation (26), which is an implication of the zero cutoff profit conditions. The determination of the monitoring ability cutoffs is thus perfectly analogous to that of the productivity cutoffs in the Melitz (2003) model. As a consequence, the expressions for $\partial m^*/\partial\tau$ and $\partial m_x^*/\partial\tau$ necessarily restate results already known from Melitz (2003). Differentiating (25) and (26) and using $J'(m) = -(\sigma-1)a(m)/m$ yields

$$\frac{\partial m^*}{\partial\tau} = -\frac{m^*}{\tau}\frac{f_x a(m_x^*)}{fa(m^*) + f_x a(m_x^*)} < 0$$

and

$$\frac{\partial m_x^*}{\partial\tau} = -\frac{m_x^*}{m^*}\frac{fa(m^*)}{f_x a(m_x^*)}\frac{\partial m^*}{\partial\tau} > 0.$$

That means, a decrease in variable trade costs raises the domestic cutoff and lowers the export cutoff. The reason is that lower variable trade costs make foreign intermediates less expensive relative to domestic ones. This raises demand for exporting firms, allowing less able firms to profitably export. At the entry stage a lower export cutoff comes with higher expected profits from sales abroad. And for the free entry condition to be restored, these must be compensated by lower expected profits from domestic sales. This allows only abler firms to enter the domestic market. [36]

How does the change in the cutoffs affect unemployment, wages, and total factor productivity in the countries? According to (31), the unemployment rate is decreasing in ψ_t, and, as (30) shows, the latter is in turn uniquely determined by the cutoffs. The impact of a decrease in τ on unemployment therefore depends on how ψ_t varies with the induced changes in m^* and m_x^*. Three effects can be distinguished. [37] First, as m^* rises and m_x^* falls, revenue is reallocated in favour of exporting firms, which raises the export share in revenue ϖ_x and hence ψ_t. [38] Second, $\psi(m^*)$ rises because some firms at the bottom of the distribution of m are forced to exit. This too raises ψ_t. Third, some firms that served only the domestic market enter the export market, thereby reducing $\psi(m_x^*)$. This makes ψ_t smaller. Due to this third effect opposing the other two, the impact of a decrease in trade costs on ψ_t and hence unemployment in general cannot be signed. As a consequence, even under Assumption 1 the impact on wages cannot be signed either. This contrasts with the sharp results obtained for a transition from autarky to trade. The reason for this divergence of results is that when an economy opens up to trade there is no *marginal* reallocation of market shares from non-exporting to exporting firms, which lowers ψ_t.

To remove the ambiguity of results regarding the effects of a decrease in trade costs we assume that the first two of the above effects dominate the third one. This assumption can be restated by using the fact that the free entry condition (25) implicitly defines m_x^* as a monotonically decreasing function of m^*. Substituting this function for m_x^* into the expression for ψ_t in (30) allows us to express it as a function of the domestic cutoff m^* alone. Assuming that the first

[36] A higher domestic cutoff has an opposing effect on the profitability of exporting: it raises the toughness of competition faced by exporters abroad and tends to reduce their profits. However, this effect is never strong enough to overturn the positive impact effect of lower trade costs on export competitiveness.

[37] These effects parallel those shown up by Helpman and Itskhoki (2010) in their analysis of the response of total factor productivity to a change in labour market frictions.

[38] First, recall that ψ_t in (30) can be rewritten as $\psi_t = \psi(m^*) + \varpi_x[\psi(m_x^*) - \psi(m^*)]$. Second, to see that ϖ_x increases, note that it is given by $\varpi_x = f_x a(m_x^*)/[f a(m^*) + f_x a(m_x^*)]$ and $a(\cdot)$ is a decreasing function ($a'(m) = (1-\sigma)m^{-\sigma} \int_m^1 \xi^{\sigma-1} g(\xi) d\xi - g(m) < 0$). Finally, note that $\psi(\cdot)$ is an increasing function, so that $\psi(m_x^*) > \psi(m^*)$ for $m_x^* > m^*$.

two effects dominate the third one is therefore equivalent to [39]

Assumption 2. *The distribution $G(m)$ is such that ψ_t monotonically increases with m^* for $m_x^* > m^*$.*

Assumption 2 ensures that a lower τ raises ψ_t and therefore reduces a country's unemployment rate. According to (23), aggregate labour demand is decreasing in the domestic cutoff and, under Assumption 1, increasing in the wage. It thus follows that under Assumptions 1 and 2 lower trade costs lead to higher wages. As a result, the unemployed and all workers who stay employed will gain from a decrease in τ, as do initially unemployed workers who are hired by firms that expand employment. Workers who may lose are those who are displaced from their jobs because the firms in which they worked are forced to exit or reduce their employment levels. The jobs lost are again those carrying relatively high rents to workers.

According to (33a), our measure of a country's total factor productivity is given by $TFP = w^{1/\gamma}\phi_t$, where ϕ_t is the overall average monitoring ability of firms. Recall that the latter can be expressed as a revenue-weighted average of $\phi(m^*)$ and $\phi(m_x^*)$, which are measures of the average monitoring ability of firms serving the domestic market and exporting firms, respectively. For reasons similar to those described for the response of ψ_t the impact of lower trade costs on ϕ_t in general cannot be signed. The average monitoring ability of firms serving the domestic market, $\phi(m^*)$, rises. This and the higher export share in revenue raise ϕ_t. On the other hand, the fact that some less able firms start exporting reduces the average monitoring ability of exporters, $\phi(m_x^*)$. This lowers ϕ_t. [40]

As just shown, under Assumptions 1 and 2 a decrease in variable trade costs leads to higher wages, thereby enabling each surviving firm to elicit greater effort from its workers. Under the additional assumption that a larger domestic cutoff results in higher average monitoring ability ϕ_t of firms, a lower τ causes average worker effort and hence total factor productivity to

[39] It is straightforward to derive sufficient conditions for Assumption 2 to hold. They are similar to those provided in Helpman and Itskhoki (2010) for total factor productivity to be rising with the domestic productivity cutoff. Moreover, Helpman and Itskhoki (2010) show that in their model total factor productivity increases with the domestic productivity cutoff when firm productivity follows a Pareto distribution. In our model a similar result with respect to $\psi_t(m^*)$ does not obtain because monitoring ability m has an upper bound of one, implying that the standard Pareto distribution is ruled out. We examined whether Assumption 2 is satisfied when m is drawn from a truncated Pareto distribution with lower bound $m_{\min}$, upper bound 1, and shape parameter $k > 0$. As it turned out, Assumption 2 is satisfied for a variety of plausible parameter sets (e.g., that used in the numerical example presented in Section 6.2). However, it may also happen that Assumption 2 is not satisfied. An example in which ψ_t falls with m^* for a wide range of $m^* < m_x^*$ is generated by the following parameters: $m_{\min} = 0.5$, $k = 2.1$, $\sigma = 2.2$, $\gamma = 12$, $r = 0.04$, $\delta = 0.05$, $f = 1$, $f_x = 1.1$, and $f_E = 4.5$.

[40] The latter effect is again absent in a move from autarky to trade. And this explains why a transition to trade unambiguously raises the average monitoring ability across all firms.

rise. [41] These results are summarized in the following proposition.

Proposition 3. *Let Assumptions 1 and 2 hold. Then: (i) A decrease in variable trade costs reduces unemployment and raises wages. (ii) The unemployed, workers who stay employed, and initially unemployed workers who enter employment gain from lower variable trade costs; the only workers who may lose are those whose jobs (providing relatively high rents) are destroyed. (iii) If average monitoring ability increases with the domestic cutoff, a decrease in variable trade costs raises total factor productivity.*

6.2 Numerical Example

We present in this section a numerical example to shed some light on the magnitude of the effects of changes in variable trade costs. An assumption commonly made in Melitz-type models of heterogeneous firms is that productivity draws follow a Pareto distribution. As we know, the standard Pareto form cannot be adopted here because the structure of the model requires a distribution function with upper bound 1 of monitoring ability m. That is why we assume that m is drawn from a truncated Pareto distribution with lower bound $m_{\min}$, upper bound 1, and shape parameter $k > 0$ given by

$$G(m) = \frac{1}{1 - m_{\min}^k}\left[1 - \left(\frac{m_{\min}}{m}\right)^k\right], \quad 0 < m_{\min} \leq m \leq 1.$$

We set $k = 3.4$ and $m_{\min} = 0.5$.

The remaining parameters are: $\sigma = 3.8$, $\gamma = 12$, $r = 0.04$, $\delta = 0.05$, $f = 1$, $f_x = 1.3$, $f_E = 10$, and $L = 1$; in order to illustrate the quantitative impact of unemployment benefits, we simulate our model for both a high ($\varrho = 0.5$) and a low ($\varrho = 0.3$) value of the benefit replacement rate.

The values of k and σ are those calibrated by Bernard, Eaton, Jensen, and Kortum (2003) for a standard Pareto distribution of productivity to fit plant data for the US manufacturing sector. These values are also used by Ghironi and Melitz (2005), Davis and Harrigan (2011), and Felbermayr, Prat, and Schmerer (2011). [42] As already noted, with work effort viewed as a dimension of labour supply, we may consider the elasticity of disutility of effort, γ, as being related to the intertemporal elasticity of substitution in labour supply. The choice of $\gamma = 12$ then implies an intertemporal substitution elasticity of $1/(\gamma - 1) \approx 0.09$, which is within the -0.07 to 0.45 range of estimates reported by Pencavel (1986). The values of σ and γ satisfy the

---

[41] In all of the simulations we performed using a truncated Pareto distribution (including the examples in which ψ_t decreases with m^* for some $m^* < m_x^*$) the average monitoring ability ϕ_t turned out to be monotonically increasing with m^* for $m^* < m_x^*$. However, we have been unable to prove that this in general holds true when m follows a truncated Pareto distribution.

[42] Davis and Harrigan (2011), in a second approach, set $k = \sigma = 2$.

requirement $\sigma > (2\gamma - 1)/(\gamma - 1)$ (Assumption 1). [43]

The equilibrium values of the monitoring ability cutoffs depend on the ratios f/f_E and f_x/f_E. We therefore normalize the fixed cost of production f to 1. The choice of a fixed cost of exporting $f_x = 1.3$ and an entry cost $f_E = 10$ then implies that for $\tau = 1.3$ (a benchmark introduced by Ghironi and Melitz, 2005) the share of exporting firms is 20.5%. This number is roughly in line with that reported in Bernard et al.(2003) and thereafter commonly used in the literature.

The choice of $\varrho = 0.5$ is in line with the simple average of 0.497, calculated from the 1999 data for 20 OECD countries reported in Nickell, Nunziata, and Ochel (2005). Setting the labour force $L = 1$ does not matter for the equilibrium cutoff levels and the unemployment rate. However, as final goods production exhibits external economies of scale, it does affect variables such as wages and total factor productivity.

Assumption 2 according to which ψ_t increases with m^* is satisfied in our example. Moreover, average monitoring ability ϕ_t increases with m^*. As a consequence, the simulation necessarily corroborates what we know from Proposition 3 about the qualitative effects of trade liberalization.

The curves u_H and u_L in Figure 4 depict the response of the unemployment rate to variation in τ for the cases of a high and a low benefit replacement rate, respectively. According to (31), a higher benefit replacement rate always comes with a higher unemployment rate. When variable trade costs decline, unemployment falls but the simulation suggests that the effect is rather small. A reduction in τ from 1.5 to 1.1 lowers the unemployment rate from 8.80% to 8.67% when $\varrho = 0.5$ and from 6.12% to 6.03% when $\varrho = 0.3$. [44]

By contrast, trade liberalization has quite substantial effects on both wages and total factor productivity. The curves w_H and w_L in Figure 5 show the variation in wages for a high and a low benefit replacement rate, respectively. At a given τ a higher ϱ leads to a higher unemployment rate and therefore a lower wage, which implies that the curve w_H lies below the curve w_L. A decrease in τ from 1.5 to 1.1 raises wages by 10.44% when $\varrho = 0.5$ and by 10.35% when $\varrho = 0.3$.

The curves TFP_H and TFP_L in Figure 6 depict the variation in total factor productivity for the high and low replacement rate cases. Since total factor productivity is increasing in the wage and at a given τ a higher replacement rate comes with a lower wage, TFP_H lies below TFP_L. Changing τ from 1.5 to 1.1 raises total factor productivity in both cases by about 5.3% (with the percentage differing in the second digit). Interestingly, though the variation in the benefit

---

[43] Using a truncated Pareto distribution requires some further parameter restrictions to be met. First, the functions $\widetilde{m}(\cdot)$ and $a(\cdot)$ are defined for $k \neq \sigma - 1$. Second, for the functions $\psi(\cdot)$ and $\phi(\cdot)$ to be defined, the additional restrictions $\gamma + \sigma - 1 \neq k$ and $k \neq \sigma$, respectively, must be satisfied. All of these requirements are met.

[44] Under autarky the unemployment rates are 9.00% when $\varrho = 0.5$ and 6.26% when $\varrho = 0.3$.

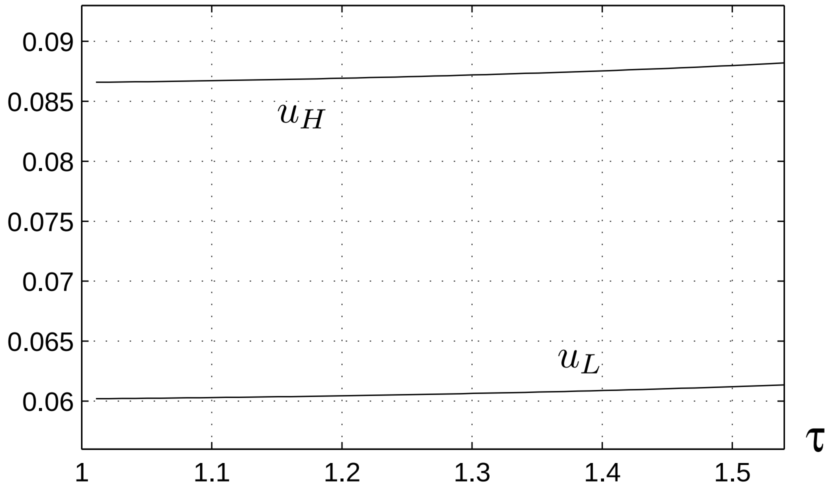

FIGURE 4. Response of unemployment to τ

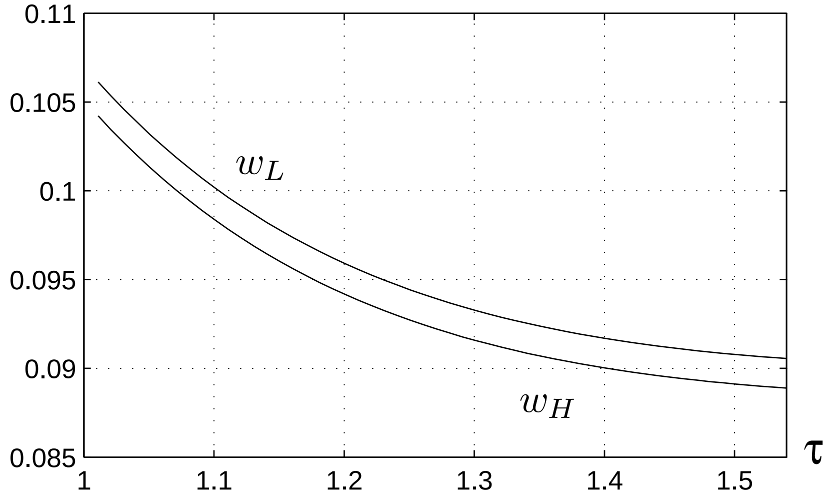

FIGURE 5. Response of wages to τ

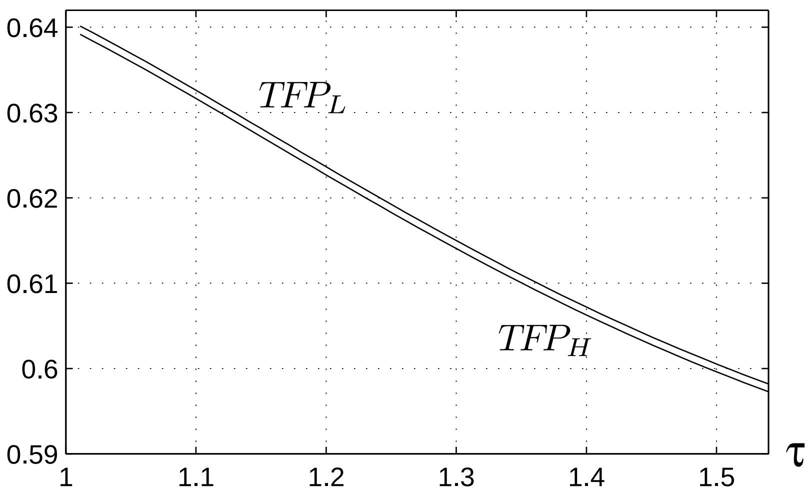

FIGURE 6. Response of total factor productivity to τ

replacement rate has a significant impact on unemployment, its effects on both wages and total factor productivity are relatively small. The reason is that in our example the response of wages to changes in aggregate employment is rather weak.

7 Conclusion

This paper has studied the impact of trade in a symmetric two-country model that combines shirking-based efficiency wages with Melitz-style heterogeneity of firms. Unlike earlier work on the subject that has treated productivity as a fixed parameter of the firm, our model features a simple mechanism, driven by the optimizing behaviour of firms, by which firm productivity becomes an endogenous variable. Key to this mechanism is that firms are free to offer their workers a contract specifying a wage rate and any positive level of effort. Firm heterogeneity comes into play by allowing firms to differ in their efficacy to monitor workers. In equilibrium, firms abler in monitoring ask greater work effort and are therefore more productive than less able ones while all of them pay the same wage. A further implication is that homogeneous workers employed by firms with different monitoring abilities have different utility levels. Good jobs from a worker's perspective are those at less able firms associated with low effort costs, and these coexist with bad jobs at abler firms where effort costs to workers are high.

The model yields some sharp predictions about the effects of a transition from autarky to trade. Trade opening unambiguously reduces unemployment and - under a plausible condition - raises wages and aggregate factor productivity. Moreover, under that condition the unemployed and all workers who stay in employment gain from trade. However, there are workers who may lose because their jobs are destroyed. As for the question of what sort these jobs can be expected to be, our model again offers an unambiguous answer. The opening to trade forces firms with low monitoring ability to exit or contract. That means, only good jobs from a worker's perspective are lost through trade, including the best (least onerous) ones in existence. All of these results have been derived without imposing restrictions on the distribution of monitoring abilities of firms, besides the general ones entailed by the model's structure.

The paper has also examined the impact of a decrease in variable trade costs. The mechanisms at work are similar to those in the case of a move from autarky to trade. The difference in Melitz-type models as the present one is that lower variable trade costs induce additional firms to enter the export market (that initially served only the domestic market). That is why the effects of a decrease in trade costs on unemployment, wages, and aggregate factor productivity are in general ambiguous. Conditions are provided under which a decrease in variable trade costs induces effects being qualitatively identical to those triggered by a move from autarky to trade. To study the quantitative impact of trade liberalization, we have simulated the model. We have found significant gains in terms of aggregate productivity and real wages but rather small effects on the unemployment rate.

As for the result that good jobs from a worker's point of view are threatened by trade, this paper is in line with the findings obtained by Davis and Harrigan (2011). As in ours, in their model high-rent jobs are those at firms with low monitoring ability. However, since in theirs worker effort is fixed by assumption, less able firms have to pay higher wages than abler ones to induce their workers not to shirk. Good jobs are thus those that pay relatively high wages. Moreover, in the Davis-Harrigan model firms differ in both monitoring ability and inherent productivity. As a consequence, there is "a conditional threat to good jobs": conditional on some productivity level, only good jobs are lost through trade. By contrast, in our model the threat through trade to good jobs is an unconditional one.

Labour market characteristics such as unemployment benefits vary greatly across countries. This suggests that the thing to do should be to develop models in which the countries differ with respect to such characteristics, e.g., in which they have different benefit replacement rates. However, studying trade between asymmetric countries in models like the present one seems to be complicated. [45] For one thing, in equilibrium the cutoff levels will differ across countries, as will the aggregate variables such as the price index and aggregate income. In addition, there might arise cross-country differences in wages and hence in the producer prices charged by firms with a given monitoring ability or productivity.

[45] Recently, Felbermayr, Larch, and Lechthaler (2011) have developed an asymmetric two-country, one-sector model with heterogeneous firms and search and matching frictions to study how a country's optimal level of the benefit replacement rate varies with country size and trade openness. However, their model cannot be solved analytically so that they have to resort to a numerical analysis.

Appendix

A Derivation of Equation (15)

Using the definition of $\overline{e^\gamma}$ together with the definition of $\widetilde{m}$ in (10), equations (7), (9), and $(1-u)L = Ml(\widetilde{m})$ yields

$$\begin{aligned}
\overline{e^\gamma} &= \frac{1}{(1-u)L}\int_{m^*}^{1} e(m)^\gamma l(m) M\mu(m)dm \\
&= \frac{w}{l(\widetilde{m})}\int_{m^*}^{1} m^\gamma l(m)\mu(m)dm \\
&= w\widetilde{m}^{1-\sigma}\int_{m^*}^{1} m^{\gamma+\sigma-1}\mu(m)dm \\
&= \frac{w\int_{m^*}^{1} m^{\gamma+\sigma-1}g(m)dm}{\int_{m^*}^{1} m^{\sigma-1}g(m)dm} \\
&= w\psi(m^*)
\end{aligned}$$

with $\psi(m^*)$ defined as in (15).

B Derivation of Equations (28)

Using the definitions of $\widetilde{m}$ in (10) and $\widetilde{m}_t$ in (27) together with $p(m) = (\widetilde{m}_t/m)p(\widetilde{m}_t)$ yields

$$\begin{aligned}
P &= \left[\int_{m^*}^{1} p(m)^{1-\sigma} M\mu(m)dm + \int_{m_x^*}^{1}[\tau p(m)]^{1-\sigma}M_x\mu_x(m)dm\right]^{\frac{1}{1-\sigma}} \\
&= p(\widetilde{m}_t)\widetilde{m}_t\left[M\int_{m^*}^{1} m^{\sigma-1}\mu(m)dm \ + M_x\tau^{1-\sigma}\int_{m_x^*}^{1} m^{\sigma-1}\mu_x(m)dm\right]^{\frac{1}{1-\sigma}} \\
&= p(\widetilde{m}_t)\widetilde{m}_t\left[M\widetilde{m}^{\sigma-1} + M_x\left(\tau^{-1}\widetilde{m}_x\right)^{\sigma-1}\right]^{\frac{1}{1-\sigma}} \\
&= M_t^{\frac{1}{1-\sigma}}p(\widetilde{m}_t).
\end{aligned}$$

Let $\mu_x(m)$ be the density conditional on exporting with $\mu_x(m) = g(m)/[1-G(m_x^*)]$ if $m \geq m_x^*$ and zero otherwise. Using (9), (10), (27), and $x_x(m) = \tau^{-\sigma}x_d(m)$,

$$\begin{aligned}
Y &= \left[\int_{m^*}^{1} x_d(m)^{\frac{\sigma-1}{\sigma}} M\mu(m)dm + \int_{m_x^*}^{1} x_x(m)^{\frac{\sigma-1}{\sigma}} M_x\mu_x(m)dm\right]^{\frac{\sigma}{\sigma-1}} \\
&= \widetilde{m}_t^{-\sigma}\left[x_d(\widetilde{m}_t)^{\frac{\sigma-1}{\sigma}} M\int_{m^*}^{1} m^{\sigma-1}\mu(m)dm \ + x_x(\widetilde{m}_t)^{\frac{\sigma-1}{\sigma}}M_x\int_{m_x^*}^{1} m^{\sigma-1}\mu_x(m)dm\right]^{\frac{\sigma}{\sigma-1}} \\
&= \widetilde{m}_t^{-\sigma}\left[x_d(\widetilde{m}_t)^{\frac{\sigma-1}{\sigma}} M\widetilde{m}^{\sigma-1} + x_x(\widetilde{m}_t)^{\frac{\sigma-1}{\sigma}}M_x\widetilde{m}_x^{\sigma-1}\right]^{\frac{\sigma}{\sigma-1}} \\
&= x_d(\widetilde{m}_t)\widetilde{m}_t^{-\sigma}\left[M\widetilde{m}^{\sigma-1} + M_x\left(\tau^{-1}\widetilde{m}_x\right)^{\sigma-1}\right]^{\frac{\sigma}{\sigma-1}}
\end{aligned}$$

$$= \quad M_t^{\frac{\sigma}{\sigma-1}} x_d(\widetilde{m}_t).$$

Using (9), (10), (27), and $R_x(m) = \tau^{1-\sigma} R_d(m)$,

$$\begin{aligned}
TR &= \int_{m^*}^1 R_d(m) M\mu(m)dm + \int_{m_x^*}^1 R_x(m) M_x \mu_x(m)dm \\
&= \widetilde{m}_t^{1-\sigma}\left[R_d(\widetilde{m}_t) M \int_{m^*}^1 m^{\sigma-1}\mu(m)dm + R_x(\widetilde{m}_t) M_x \int_{m_x^*}^1 m^{\sigma-1}\mu_x(m)dm\right] \\
&= R_d(\widetilde{m}_t)\widetilde{m}_t^{1-\sigma}\left[M\widetilde{m}^{\sigma-1} + M_x\left(\tau^{-1}\widetilde{m}_x\right)^{\sigma-1}\right] \\
&= M_t R_d(\widetilde{m}_t).
\end{aligned}$$

Using $p(\widetilde{m}_t) = M_t^{1/(\sigma-1)}$ (from $P = 1$),

$$\begin{aligned}
TR &= M_t R_d(\widetilde{m}_t) \\
&= M_t p(\widetilde{m}_t) x_d(\widetilde{m}_t) \\
&= M_t^{\frac{\sigma}{\sigma-1}} x_d(\widetilde{m}_t) = Y.
\end{aligned}$$

From (29), $M(\overline{\pi} + f + \chi f_x) = (1-\theta)Y$, hence $(1-u)Lw = Y - M(\overline{\pi} + f + \chi f_x) = \theta Y$ and

$$(1-u)L = \frac{\theta Y}{w} = \frac{\theta}{w} M_t R_d(\widetilde{m}_t).$$

Using $x(m) = e(m)l(m)$ and (8),

$$\begin{aligned}
R_d(\widetilde{m}_t) &= p(\widetilde{m}_t)x_d(\widetilde{m}_t) = p(\widetilde{m}_t)e(\widetilde{m}_t)l_d(\widetilde{m}_t) \\
&= \frac{w}{\theta} l_d(\widetilde{m}_t).
\end{aligned}$$

Hence, $(1-u)L = M_t l_d(\widetilde{m}_t)$.

C Derivation of Equation (30)

Using the definition of $\overline{e^\gamma}$ together with the definition of $\psi(\cdot)$ in (15), equations (7), (9), and $(1-u)L = M_t l_d(\widetilde{m}_t)$ yields

$$\begin{aligned}
\overline{e^\gamma} &= \frac{1}{(1-u)L}\left[\int_{m^*}^1 e(m)^\gamma l_d(m) M\mu(m)dm + \int_{m_x^*}^1 e(m)^\gamma l_x(m) M_x \mu(m)dm\right] \\
&= \frac{w}{M_t l_d(\widetilde{m}_t)}\left[\int_{m^*}^1 m^\gamma l_d(m) M\mu(m)dm + \int_{m_x^*}^1 m^\gamma l_x(m) M_x \mu_x(m)dm\right] \\
&= \frac{w}{M_t l_d(\widetilde{m}_t)}\left[M l_d(\widetilde{m})\widetilde{m}^{1-\sigma}\int_{m^*}^1 m^{\gamma+\sigma-1}\mu(m)dm \; + \right. \\
&\qquad \left. M_x l_x(\widetilde{m}_x)\widetilde{m}_x^{1-\sigma}\int_{m_x^*}^1 m^{\gamma+\sigma-1}\mu_x(m)dm\right] \\
&= w\psi_t,
\end{aligned}$$

where

$$\psi_t = \frac{1}{M_t l_d(\widetilde{m}_t)} \left[M l_d(\widetilde{m}) \psi(m^*) + M_x l_x(\widetilde{m}_x) \psi(m_x^*) \right].$$

Noting that $M_t l_d(\widetilde{m}_t) = M l_d(\widetilde{m}) + M_x l_x(\widetilde{m}_x)$, we find that ψ_t is a weighted average of $\psi(m^*)$ and $\psi(m_x^*)$, where the weights are the shares of domestic sales and exports in aggregate employment. From (9),

$$\frac{l_d(\widetilde{m})}{l_d(\widetilde{m}_t)} = \frac{R_d(\widetilde{m})}{R_d(\widetilde{m}_t)} = \left(\frac{\widetilde{m}}{\widetilde{m}_t}\right)^{\sigma-1}.$$

Noting that $M_t R_d(\widetilde{m}_t) = M R_d(\widetilde{m}) + M_x R_x(\widetilde{m}_x)$, it follows that the shares of domestic sales and exports in employment equal their shares in revenue.

It remains to be shown that the weights can be rewritten in terms of $a(\cdot)$. From the zero cutoff profit conditions we have $R_d(m^*) = \sigma f$ and $R_x(m_x^*) = \sigma f_x$. Using this and, from (9), $R_d(\widetilde{m}) = (\widetilde{m}/m^*)^{\sigma-1} R_d(m^*)$ and $R_x(\widetilde{m}_x) = (\widetilde{m}_x/m_x^*)^{\sigma-1} R_x(m_x^*)$ gives

$$\begin{aligned} M R_d(\widetilde{m}) &= M \left(\frac{\widetilde{m}}{m^*}\right)^{\sigma-1} \sigma f, \\ M_x R_x(\widetilde{m}_x) &= M_x \left(\frac{\widetilde{m}_x}{m_x^*}\right)^{\sigma-1} \sigma f_x = \chi M \left(\frac{\widetilde{m}_x}{m_x^*}\right)^{\sigma-1} \sigma f_x. \end{aligned}$$

Using this together with (24), (29), and $Y = TR$ yields

$$\begin{aligned} TR &= \sigma M (\overline{\pi} + f + \chi f_x) \\ &= \sigma M \left[f \left(\frac{\widetilde{m}}{m^*}\right)^{\sigma-1} + \chi f_x \left(\frac{\widetilde{m}_x}{m_x^*}\right)^{\sigma-1} \right] \end{aligned}$$

and

$$\varpi_d = \frac{M R_d(\widetilde{m})}{TR} = \frac{f (\widetilde{m}/m^*)^{\sigma-1}}{f (\widetilde{m}/m^*)^{\sigma-1} + \chi f_x (\widetilde{m}_x/m_x^*)^{\sigma-1}}.$$

Multiplying the numerator and the denominator by $1 - G(m^*)$ gives

$$\varpi_d = \frac{f a(m^*)}{f a(m^*) + f_x a(m_x^*)},$$

where $a(m) \equiv m^{1-\sigma} \int_m^1 \xi^{\sigma-1} g(\xi) d\xi$.

D Derivation of Equations (33) and (33a)

From (7), $e_a(m) = m w_a^{1/\gamma}$, and from (9), $l_a(m) = (m/\widetilde{m}_a)^{\sigma-1} l_a(\widetilde{m}_a)$, where $\widetilde{m}_a \equiv \widetilde{m}(m_a^*)$. Using this together with $(1 - u_a)L = M_a l_a(\widetilde{m}_a)$ and the definition of ϕ in

$$TFP_a = \frac{1}{(1 - u_a)L} \int_{m_a^*}^1 e_a(m) l_a(m) M_a \mu(m) dm$$

yields

$$
\begin{aligned}
TFP_a &= \frac{w_a^{1/\gamma}}{(1-u_a)L}\int_{m_a^*}^1 m l_a(m) M_a \mu(m) dm \\
&= \frac{w_a^{1/\gamma}}{(1-u_a)L} M_a l_a(\widetilde{m}_a)\widetilde{m}_a^{1-\sigma}\int_{m_a^*}^1 m^\sigma \mu(m) dm \\
&= \frac{w_a^{1/\gamma}}{M_a l_a(\widetilde{m}_a)} M_a l_a(\widetilde{m}_a)\phi(m_a^*) \\
&= w_a^{1/\gamma}\phi(m_a^*).
\end{aligned}
$$

Equation (33a) is derived in a similar way. Using (7), (9), and $(1-u)L = M_t l_d(\widetilde{m}_t)$ in conjunction with the definition of ϕ yields

$$
\begin{aligned}
TFP &= \frac{w^{1/\gamma}}{(1-u)L}\left[\int_{m^*}^1 m l_d(m) M \mu(m) dm + \int_{m_x^*}^1 m l_x(m) M_x \mu_x(m) dm\right] \\
&= \frac{w^{1/\gamma}}{(1-u)L}\left[M l_d(\widetilde{m})\widetilde{m}^{1-\sigma}\int_{m^*}^1 m^\sigma \mu(m) dm + M_x l_x(\widetilde{m}_x)\widetilde{m}_x^{1-\sigma}\int_{m_x^*}^1 m^\sigma \mu(m) dm\right] \\
&= \frac{w^{1/\gamma}}{M_t l_d(\widetilde{m}_t)}[M l_d(\widetilde{m})\phi(m^*) + M_x l_x(\widetilde{m}_x)\phi(m_x^*)].
\end{aligned}
$$

Accounting for $M_t l_d(\widetilde{m}_t) = M l_d(\widetilde{m}) + M_x l_x(\widetilde{m}_x)$, the employment weights can be rewritten in terms of $a(\cdot)$, as shown in the derivation of (30) above. This leads to equation (33a).

References

Agell, J., Lundborg, P., 1995. Fair wages in the open economy. Economica 62, 335–351.

Alexopoulos, M., 2003. Growth and unemployment in a shirking efficiency wage model. Canadian Journal of Economics 36, 728–746.

Allgulin, M., Ellingsen, T., 2002. Monitoring and pay. Journal of Labor Economics 20, 201–216.

Altenburg, L., Brenken, B., 2008. Effort, trade, and unemployment. Canadian Journal of Economics 41, 864–893.

Amiti, M., Davis, D.R., 2011. Trade, firms, and wages: theory and evidence. Review of Economic Studies 78, Advance Access published August 22, 2011.

Bernard, A.B., Eaton, J., Jensen, J.B., Kortum, S., 2003. Plants and productivity in international trade. American Economic Review 93, 1268–1290.

Bernard, A.B., Jensen, J.B., Redding, S.J., Schott, P.K., 2007. Firms in international trade. Journal of Economic Perspectives 21, 105–130.

Blanchard, O., Giavazzi, F., 2003. Macroeconomic effects of regulation and deregulation in goods and labor markets. Quarterly Journal of Economics 118, 879–907.

Brecher, R.A., 1974. Minimum wage rates and the pure theory of international trade. Quarterly Journal of Economics 88, 98–116.

Brecher, R.A., 1992. An efficiency-wage model with explicit monitoring: unemployment and welfare in an open economy. Journal of International Economics 32, 179–191.

Copeland, B.R., 1989. Efficiency wages in a Ricardian model of international trade. Journal of International Economics 27, 221–244.

Davidson, C., Martin, L., Matusz, S., 1988. The structure of simple general equilibrium models with frictional unemployment. Journal of Political Economy 96, 1267–1293.

Davidson, C., Martin, L., Matusz, S., 1999. Trade and search generated unemployment. Journal of International Economics 48, 271–299.

Davis, D.R., 1998. Does European unemployment prop up American wages? National labor markets and global trade. American Economic Review 88, 478–494.

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

Demidova, S., 2008. Productivity improvements and falling trade costs: boon or bane? International Economic Review 49, 1437–1462.

Egger, H., Kreickemeier, U., 2008. Fairness, trade, and inequality. CESifo Working Paper, No 2344.

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

Ethier, W.J. (1982) National and international returns to scale in the modern theory of international trade. American Economic Review 72, 389–405.

Felbermayr, G., Larch, M., Lechthaler, W., 2011. Endogenous labor market institutions in an open economy. Ifo Working Paper, No 108.

Felbermayr, G., Prat, J., Schmerer, H.-J., 2011. Globalization and labor market outcomes: wage bargaining, search frictions, and firm heterogeneity. Journal of Economic Theory 146, 39–73.

Ghironi, F., Melitz, M.J., 2005. International trade and macroeconomic dynamics with heterogeneous firms. Quarterly Journal of Economics 120, 865–915.

Green, F., 2006. Demanding Work. The Paradox of Job Quality in the Affluent Economy. Princeton University Press, Princeton and Oxford.

Green, F., McIntosh, S., 2001. The intensification of work in Europe. Labour Economics 8, 291–308.

Helpman, E., Itskhoki, O., 2010. Labour market rigidities, trade and unemployment. Review of Economic Studies 77, 1100–1137.

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

Hoon, H.T., 2001. Adjustment of wages and equilibrium unemployment in a Ricardian global economy. Journal of International Economics 54, 193–209.

Kreickemeier, U., Nelson, D., 2006. Fair wages, unemployment and technological change in a global economy. Journal of International Economics 70, 451–469.

Matusz, S.J., 1996. International trade, the division of labor, and unemployment. International Economic Review 37, 71–84.

Mehta, S.R., 1998. The law of one price and a theory of the firm: a Ricardian perspective on interindustry wages. Rand Journal of Economics 29, 137–156.

Melitz, M.J., 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. Econometrica 71, 1695–1725.

Moselle, B., 1996. Efficiency wages and the hours/unemployment trade-off. Northwestern University, Center for Mathematical Studies in Economics and Management Science Discussion Paper, No 1153.

Nickell, S., Nunziata, L., Ochel, W., 2005. Unemployment in the OECD since the 1960s. What do we know? Economic Journal 115, 1–27.

Pencavel, J., 1986. Labor supply of men: a survey, in: Ashenfelter, O., Layard, R. (Eds.), Handbook of Labor Economics, Vol. 1. North-Holland, Amsterdam, pp. 3–102.

Rasmussen, B.S., 2002. Efficiency wages and the long-run incidence of progressive taxation. Journal of Economics 76, 155–175.

Romer, D., 2006. Advanced Macroeconomics, 3rd edn. McGraw-Hill/Irwin, Boston.

Scheve, K., Slaughter, M.J., 2006. Public opinion, international economic integration, and the welfare state, in: Bardhan, P., Bowles, S., Wallerstein, M. (Eds.), Globalization and Egalitarian Redistribution. Princeton University Press, Princeton and Oxford, pp. 217–259.

Shapiro, C., Stiglitz, J.E., 1984. Equilibrium unemployment as a worker discipline device. American Economic Review 74, 433–444.

Tybout, J.R., 2003. Plant- and firm-level evidence on "new" trade theories, in: Choi, E.K., Harrigan, J. (Eds.), Handbook of International Trade. Blackwell, Oxford, pp. 388–415.

Walsh, F., 1999. A multisector model of efficiency wages. Journal of Labor Economics 17, 351–376.