DISCUSSION PAPER SERIES

IZA

IZA DP No. 1904

Search Equilibrium, Production Parameters and Social Returns to Education: Theory and Estimation

Christian Holzner
Andrey Launov

December 2005

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Search Equilibrium, Production Parameters and Social Returns to Education: Theory and Estimation

Christian Holzner
ifo Institute for Economic Research

Andrey Launov
University of Würzburg
and IZA Bonn

Discussion Paper No. 1904
December 2005

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
Email: iza@iza.org

IZA Discussion Paper No. 1904
December 2005

ABSTRACT

Search Equilibrium, Production Parameters and Social Returns to Education: Theory and Estimation*

We introduce different skill groups and production functions into the Burdett-Mortensen equilibrium search model. Supermodularity in the production process leads to a positive intrafirm wage correlation between skill groups. Theory implies that increasing returns to scale can lead to a unimodal earnings density with a decreasing right tail even in the absence of productivity dispersion. Our empirical results indicate economy-wide increasing returns to scale. We use the structural estimates of the production parameters to investigate whether private returns to education equal social returns. Our estimates suggest a positive welfare effect from increasing the share of medium-skilled agents in the workforce.

JEL Classification: J21, J23, J64

Keywords: search, wage correlation, social returns to education

Corresponding author:

Christian Holzner
ifo Institute for Economic Research
Poschingerstr. 5
81679 Munich
Germany
Email: holzner@ifo.de

* The authors wish to thank Melvyn Coles, Christian Dustmann, Leo Kaas, Stephan Klasen, Klaus Schmidt, Aico van Vuuren and Klaus Wälde for their useful comments and Joachim Wolff for the major part of the data preparation. The usual disclaimer applies.

1. INTRODUCTION

It is generally agreed that the shape of the wage earnings distribution is determined by the skill distribution of the work force, the production technologies used and the search and matching frictions that govern the allocation of workers to jobs. The aim of this paper is to provide a theoretical and still empirically tractable model that takes all these three factors and their interactions into account. For doing so we extend the search equilibrium model of Burdett and Mortensen (1998) by introducing different skill groups that are linked via a production function which permits decreasing as well as increasing returns to scale.

Since the endogenous wage distribution generated by the original Burdett-Mortensen model has an upward-sloping density, which is at odds with the empirical observation of a flat right tail, there has been a lot of effort to extend the original model in order to generate a more realistically shaped wage distribution. In the present extension we demonstrate that with skill multiplicity and a production function that permits any degree of homogeneity we get a unimodal right-skewed wage offer and earnings densities with a decreasing right tail. Even though we later introduce productivity dispersion our result about the shape of the wage offer and earnings densities is true even for identical employers.

Mortensen (1990) introduces differences in firm productivity and Bowlus et al. (1995) show that this greatly improves the fit to the empirical wage distribution. Bontemps et al. (2000) and Burdett and Mortensen (1998) formulate a closed-form solution for a continuous atomless productivity distribution which translates into a wage earnings density with a decreasing right tail. Bontemps et al. (1999) extend this for both employer and worker heterogeneity. While the structural models with continuous productivity dispersion as suggested by Bontemps et al. (1999, 2000) and Postel-Vinay and Robin (2002) improve the fit to the empirical wage earnings distribution and provide reliable estimates of the labor market transition rates, they are not informative about the production parameters governing the productivity dispersion (see Manning, 2003, p.106f). In this paper different production technologies are explicitly introduced. This allows us to estimate the technology parameters which determine the form of the productivity dispersion.

With the introduction of technology parameters we achieve not only a much more realistic shape of the earnings distribution we also open another dimension in the application of empirical equilibrium search models, making it possible to study the impact that a marginal shift in the skill structure of an economy has on the output as well as on the wage offer and wage earnings distribution. The information contained in technology parameters enables us to evaluate the private and social returns to acquiring a specific skill level and, thus, to investigate whether there is over- or underinvestment into human capital in the economy. In particular, we seek to answer the question whether an increase in output resulting from educating an individual one skill level up would be higher than the private return to the investment in education of the marginal individual. In the empirical part of the paper we estimate the model to answer this question for Germany. We find that a marginal change in the skill structure of the labor force away from low-skilled and towards more medium-skilled workers does indeed generate an increase in output sufficient to overcompensate the society for the additional cost of educating the marginal individual. At the same time, the number of high-skilled workers is found to be close to the socially efficient level.

The underinvestment result can be explained by the work of Acemoglu (1996) and Masters (1998) who show in an undirected search and matching model that individuals will underinvest in skills, since matching frictions and bargaining make it impossible to capture the whole return to the investment. This underinvestment result rests on the assumption that both types of worker search in the same market. If workers of different skills search in segmented markets both over- and underinvestment in education is possible as shown by Saint-Paul (1996). The reason is that a lower unemployment rate among high skilled workers can increase the return to human capital investment to such a degree that the negative effect of search frictions is more than offset.[1] In our model markets are segmented according to skills and hence both over- and underinvestment is possible.

The analysis of different groups of individuals that are segmented into different labor markets is related to the work by Bowlus and Eckstein (2002). Within the simple Burdett-Mortensen model Bowlus and Eckstein (2002) analyze discrimination and skill differences by allowing for different productivity and different transition parameters across races as well as incorporating employers discrimination. Unlike Bowlus and Eckstein (2002), in this paper we rather focus on how the interaction of different skill groups in the production process influences the determination the marginal product and the wage distribution of the each skill group. Furthermore, along with productivity differences, we consider differences in the values of the labor market states across the skill groups.

[1] Acemoglu and Shimer (1999) show that the hold-up problem can be overcome if workers are able to direct their search to potentially different markets.

In the theoretical part of the paper we also demonstrate that whenever skills are complementary in the production process we should observe a positive within-firm correlation between the wages of workers with different skills. Positive intrafirm wage correlation is a well established empirical fact, evidence of which are presented in Katz and Summers (1989) and Barth and Dale-Olsen (2003). The result of a positive wage correlation is the key for the derivation of a closed form solution for the skill-specific wage offer distribution which allows us to structurally estimate the model.

The estimation methodology applied in this paper is based on the one considered in Bowlus et al. (1995, 2001). Skill-multiplicity and the Cobb-Douglas production function used in the econometric model imply identities that allow representing the subset of production parameters as a function of the search frictions parameters and the degree of homogeneity the Cobb-Douglas technology. With the introduction of heterogeneous technologies, skill multiplicity also invokes the identifiability restrictions that link production parameters to the kink points ("cutoff wages") of the wage offer distribution.

The paper proceeds as follows. The theory is presented in Section 2, where we extend the existing Burdett-Mortensen framework, solve for optimal strategies of workers and firms and discuss the properties of the resulting equilibrium wage offer distribution. The empirical implementation of the model is treated in Section 3. We consider the appropriate likelihood function and discuss the relevant estimation method and identifiability issues. Thereafter, in Section 4, we provide a brief description of the data set and in detail discuss the result of the structural estimation of the model and present our results about social and private returns to education. Section 5 concludes.

2. THEORY

In this section we extend the original Burdett-Mortensen model of search equilibrium by introducing different skill groups and different technologies. This allows for the marginal product of a skill group in a particular firm to depend not only on the technology used but also on the skill structure employed in a firm and the size of the firm.

2.1 *Framework*

The model has an infinite horizon, is set in continuous time and concentrates on steady states. Workers are risk neutral and discount at rate r. Before entering the labor market each worker has to decide which skill $i = 1, 2, ..., I$ he wants to acquire. The skill levels are ranked from the lowest $i = 1$ to the highest $i = I$. The cost to acquire a specific skill-level i differs for each worker. By assuming perfect capital markets, workers are able to borrow the cost of education. We assume that the one-off cost $c_{i,a}/r$ can be described by an inverse relationship between a skill specific cost component c_i that is increasing in i and an individual's ability a, i.e. $c_{i,a} = c_i/a$. Ability is distributed according to a continuous distribution function $H(a)$ with support $a \in [\underline{a}, \overline{a}]$ with $\underline{a} > 0$. This assumption is important for the investigation of whether there is over- or underinvestment in the economy, since it guarantees that the cost ranking of all individuals is the same across skill groups. Although ability influences the cost of acquiring a specific skill-level, we assume that it does not influence the productivity at the work place.[2] For the labor market analysis we take the individuals' education decision as given and return to it in section 4.3 where we investigate the question of social returns to education. Given the education choice made, an individual belongs to a skill group $i = 1, 2, ..., I$ whose measures are denoted as q_i, satisfying $\sum q_i = m$. The measure u_i of workers is unemployed and the measure $q_i - u_i$ is employed.

Workers search for a job in the skill-segmented labor markets. Unemployed workers of skill group i encounter a firm that makes them a wage offer corresponding to their education at a Poisson rate λ_i. Employed workers encounter another firm at a Poisson rate λ_e.[3] Then workers decide whether to accept or reject the job offer. A job-worker match is destroyed at an exogenous rate $\delta > 0$. Laid off workers start again as unemployed.

We assume that there exist J distinct production technologies $Y_j(\mathbf{l}(\mathbf{w} \mid \mathbf{w^r}, F(\mathbf{w})))$ indexed by $j = 1, 2, ..., J$, where $\mathbf{l}(\mathbf{w} \mid \mathbf{w^r}, F(\mathbf{w}))$ is the vector of skill groups $l_i(w \mid w_i^r, F_i(w))$ employed by a firm with technology j. The size $l_i(w \mid w_i^r, F_i(w))$ of the skill group depends on the firm's wage offer w_i, the workers' reservation wage w_i^r and the skill specific wage offer distribution $F_i(w)$. We further assume that the production function $Y_j(\mathbf{l}(\mathbf{w} \mid \mathbf{w^r}, F(\mathbf{w})))$ is supermodular in $\mathbf{l}(\mathbf{w} \mid \mathbf{w^r}, F(\mathbf{w}))$, i.e. that labor inputs are complements. Restriction to supermodular production functions is justified later on, when we

[2] This assumption is clearly restrictive. However, as the number of skill groups increases this assumption becomes less and less restrictive, since workers with a higher ability chose a higher skill level, where they are more productive.

[3] λ_e is assumed to be the same across all skill groups, because otherwise we would not be able to derive an explicit solution for the wage offer distribution function. Assuming a constant probability of encountering another firm across different skill groups is equivalent of assuming that the mean employment spell in a job is the same for all skill groups. Fortunately, our data shows only minor differences in the employment spells across skill groups (see Table 1).

use this property to establish Proposition 1.

Definition 1: For any $\mathbf{l} \equiv \mathbf{l}(\mathbf{w} \mid \mathbf{w^r}, F(\mathbf{w}))$ and $\mathbf{l'} \equiv \mathbf{l'}(\mathbf{w} \mid \mathbf{w^r}, F(\mathbf{w}))$, $Y_j(\mathbf{l})$ is *supermodular* in $\mathbf{l}$, if

$$Y_j(\mathbf{l} \wedge \mathbf{l'}) + Y_j(\mathbf{l} \vee \mathbf{l'}) \geq Y_j(\mathbf{l}) + Y_j(\mathbf{l'}),$$

where $\mathbf{l} \vee \mathbf{l'} \equiv (\max(l_1, l'_1), ..., \max(l_I, l'_I))$ and $\mathbf{l} \wedge \mathbf{l'} \equiv (\min(l_1, l'_1), ..., \min(l_I, l'_I))$.

2.2 *Workers' Search Strategy*

As shown by Mortensen and Neumann (1988) the optimal search strategy for a worker of occupation i is characterized by a reservation wage w_i^r, where an unemployed worker is indifferent between accepting or rejecting a wage offer, i.e. $U_i = V_i(w_i^r)$. U_i is the value of being unemployed and $V_i(w_i^r)$ the value of being employed at the reservation wage w_i^r. The flow values of being unemployed and employed are given by

$$rU_i = \lambda_i \int_{w_i^r}^{\bar{w}_i} (V_i(x_i) - U_i)\, dF_i(x_i) - c_{i,a}, \tag{1a}$$

$$rV_i(w_i) = w_i + \lambda_e \int_{w_i}^{\bar{w}_i} (V_i(x_i) - V_i(w_i))\, dF_i(x_i) + \delta\,(U_i - V_i(w_i)) - c_{i,a} \tag{1b}$$

respectively. They can be solved for the reservation wage[4]

$$w_i^r = (\lambda_i - \lambda_e) \int_{w_i^r}^{\bar{w}_i} \left(\frac{1 - F_i(x)}{r + \delta + \lambda_e (1 - F_i(x^-))} \right) dx. \tag{2}$$

The wage offer distribution is given by $F_i(w) = F_i(w^-) + \upsilon_i(w)$, where $\upsilon_i(w)$ is the mass of firms offering wage w to skill group i. Since offering a wage lower than the reservation wage does not attract any worker, we assume with out loss of generality that no firm offers a wage below the reservation wage, i.e. $F_i(w) = 0$ for $w < w_i^r$.

2.3 *Steady State Flows and Skill Group Size*

Equating the flows in and out of unemployment gives the steady state measure of unemployed per skill group, i.e.

$$u_i = \frac{\delta}{\delta + \lambda_i} q_i. \tag{3}$$

Given the assumption of constant Poisson arrival and separation rates Mortensen (1999) has shown that skill group size evolves according to a special Markov-chain known as stochastic birth-death process. The birth rate of a job offered by a firm posting a wage w

[4]The details of the derivation can be found in Mortensen and Neumann (1988).

is given by the average rate at which a job is filled. There are u_i unemployed who leave unemployment at rate λ_i and $(q_i - u_i)$ employed workers who leave their current employer at rate $\lambda_e G_i(w^-)$ to join the firm offering a wage w, where $G_i(w) = G_i(w^-) + \vartheta_i(w)$ denotes the cumulative wage earnings distribution for skill group i. A worker-employer pair splits at rate δ. Moreover, a worker may receive a higher wage offer from another firm, which occurs at rate λ_e, and accepts it, which happens with probability $\overline{F}_i(w) \equiv (1 - F_i(w))$. The death rate of a job is, therefore, given by $\delta + \lambda_e \overline{F}_i(w)$. Mortensen (1999) shows that the skill group size is Poisson distributed with mean

$$E\left[l_i\left(w \mid w_i^r, F_i\left(w\right)\right)\right] = \frac{\lambda_i u_i + \lambda_e G_i(w^-)(q_i - u_i)}{\delta + \lambda_e \overline{F}_i(w)}.$$

Equating the inflow and outflow gives the steady-state measure of employed workers earning a wage less than w

$$G_i(w^-)(q_i - u_i) = \frac{\lambda_i F_i(w^-) u_i}{\delta + \lambda_e \overline{F}_i(w^-)}. \tag{4}$$

Substituting gives

$$E\left[l_i\left(w \mid w_i^r, F_i\left(w\right)\right)\right] = \frac{\delta \lambda_i \left(\delta + \lambda_e\right) / \left(\delta + \lambda_i\right)}{\left[\delta + \lambda_e \overline{F}_i(w)\right]\left[\delta + \lambda_e \overline{F}_i(w^-)\right]} q_i, \tag{5}$$

From (5) it follows that the expected skill group size $E\left[l_i\left(w \mid w_i^r, F_i\left(w\right)\right)\right]$ is (i) increasing in w, if $w \geq w_i^r$, (ii) continuous except where $F_i\left(w\right)$ has a mass point and is (iii) strictly increasing on the support of $F_i\left(w\right)$ and constant on any connected interval off the support of $F_i\left(w\right)$. The intuition behind this result is that on-the-job search implies that the higher the wage offered by a firm is, the more employed workers are attracted from firms offering lower wages and the less workers quit to employers paying higher wages. This leads to a higher steady-state skill group size for firms offering higher wages. For notational simplicity, from now on we use $l_i\left(w\right)$ instead of $l_i\left(w \mid w_i^r, F_i\left(w\right)\right)$.

2.4 *Wage Posting*

The foregoing analysis is identical to the Burdett-Mortensen model with an index i for each skill group. The following analysis of the firms' wage posting behavior differs from previous work, since the interdependence of the skill groups in the production process implies that it is no longer optimal to post the wage for one skill group independently from the wages posted for other skill groups.

Each firm posts a wage schedule $\mathbf{w}$ in order to maximize its profit, taking as given the workers' search strategy, i.e. the reservation wage vector $\mathbf{w^r}$, and the other firms' wage posting behavior, i.e. $F(\mathbf{w})$.

$$\pi_j = \max_{\mathbf{w}} E\left[Y_j\left(\mathbf{l}\left(\mathbf{w}\right)\right) - \mathbf{w}^T \mathbf{l}\left(\mathbf{w}\right)\right].$$

Firms form expectations over all possible realizations of the different skill group sizes $l_i\left(w \mid w_i^r, F_i\left(w\right)\right)$ given a firm's choice of the wage schedule and the birth-death process characterized above. Hence, a firm may choose to adjust its wage policy according to the realizations of the different skill group sizes $l_i\left(w \mid w_i^r, F_i\left(w\right)\right)$. Since this problem is intractable, we assume that a firm can specify its wage policy $\mathbf{w}$ only once, which is equivalent of assuming that firms commit to a certain position within the wage offer distribution of each skill group. Firms might for instance commit to paying the same wage to all new recruits, because they face a concern for "fairness" and "equality" on part of the workers.

Given this assumption and the fact that the birth-death process governing the hiring and quitting behavior of workers of one skill group is statistically independent from the birth-death process for another skill group we can write the maximization problem of a type j firm as

$$\pi_j = \max_{\mathbf{w}} \left[Y_j\left(E\left[\mathbf{l}\left(\mathbf{w}\right)\right]\right) - \mathbf{w}^T E\left[\mathbf{l}\left(\mathbf{w}\right)\right]\right], \tag{6}$$

if we take a second order Taylor expansion to approximate the production function.[5]

Denote by $\mathbf{W}_j$ the set of wage offers that maximize equation (6), i.e. $\mathbf{W}_j = \arg\max_{\mathbf{w}} \pi_j$, and the corresponding I-dimensional wage offer distribution for each firm type j by $F_j\left(\mathbf{w}\right) = \left(F_{1j}(w), F_{2j}(w), ..., F_{Ij}(w)\right)$, where $F_{ij}(w)$ denotes the wage offer distribution of type j firms for skill group i.

Definition 2: A *steady state wage posting equilibrium* is a wage offer distribution $F_j\left(\mathbf{w}\right)$

[5]Take the second order Taylor expansion around a vector $\mathbf{r}$ of skill group sizes, i.e.

$$\begin{aligned}
E\left[Y\left(\mathbf{l}\left(\mathbf{w}\right)\right)\right] &= E\left[Y\left(\mathbf{r}\right)\right] + \sum\nolimits_i Y'(\mathbf{r})\, E\left[\left(l_i\left(w\right) - r_i\right)\right] + \sum\nolimits_i \sum\nolimits_l Y'(\mathbf{r})\, E\left[\left(l_i\left(w\right) - r_i\right)\left(l_l\left(w\right) - r_l\right)\right] \\
&= E\left[Y\left(\mathbf{r}\right)\right] + \sum\nolimits_i Y'(\mathbf{r})\left(E\left[l_i\left(w\right)\right] - r_i\right) + \sum\nolimits_i \sum\nolimits_l Y'(\mathbf{r})\left(E\left[l_i\left(w\right)\right] - r_i\right)\left(E\left[l_l\left(w\right)\right] - r_l\right) \\
&= Y\left(E\left[\mathbf{l}\left(\mathbf{w}\right)\right]\right),
\end{aligned}$$

where the second equality follows from the independence of the skill specific hiring and quitting processes.

with $\mathbf{w} \in \mathbf{W}_j$ for each firm type $j \in J$ such that

$$\begin{aligned}
\pi_j &= Y_j \left(E \left[\mathbf{l} \left(\mathbf{w} \right) \right] \right) - \mathbf{w}^T E \left[\mathbf{l} \left(\mathbf{w} \right) \right] \text{ for all } \mathbf{w} \text{ on the support of } F_j \left(\mathbf{w} \right), \\
\pi_j &\geq Y_j \left(E \left[\mathbf{l} \left(\mathbf{w} \right) \right] \right) - \mathbf{w}^T E \left[\mathbf{l} \left(\mathbf{w} \right) \right] \text{ otherwise,}
\end{aligned} \tag{7}$$

given the reservation wage w_i^r for each skill group $i = 1, 2, ..., I$ and a corresponding skill group wage offer distribution $F_i(w)$ such that the reservation wage w_i^r satisfies equation (2) given $F_i(w)$.

2.5 *Properties of the Wage Offer Distribution*

Following Mortensen (1990) we next describe the properties of the aggregate and the skill specific wage offer distributions.

From the supermodularity property of the production function and the fact that the expected skill group size given in equation (5) is increasing in w and upper semi-continuous it follows that profits π_j are supermodular in w_i. Thus, a firm paying higher wages for one skill group also pays higher wages for another skill group.

Proposition 1 *Take a firm of type* $j \in [1, J]$ *offering* $\mathbf{w} \in \mathbf{W}_j$ *and another firm of type* j *offering* $\mathbf{w}' \in \mathbf{W}_j$*, where* $\mathbf{w}$ *and* $\mathbf{w}' \geq \mathbf{w}^\mathbf{r}$*, then either* $\mathbf{w} \geq \mathbf{w}'$ *or* $\mathbf{w} \leq \mathbf{w}'$*.*

Proof. For any $\mathbf{w}$ and $\mathbf{w}' \geq \mathbf{w}^\mathbf{r}$, $\pi_j \left(w_i, \mathbf{w}_{-i} \right)$ (where $-i$ denotes the vector of all skill groups except i) is supermodular, i.e.

$$\pi_j \left(w_i \wedge w_i', \mathbf{w}_{-i} \wedge \mathbf{w}'_{-i} \right) + \pi_j \left(w_i \vee w_i', \mathbf{w}_{-i} \vee \mathbf{w}'_{-i} \right) \geq \pi_j \left(w_i, \mathbf{w}_{-i} \right) + \pi_j \left(w_i', \mathbf{w}'_{-i} \right),$$

because the same inequality holds for output $Y_j \left(E \left[\mathbf{l} \left(w_i, \mathbf{w}_{-i} \right) \right] \right)$ and the wage cost cancel out.

Now, we prove $\mathbf{w} \geq \mathbf{w}'$ by contradiction. For any $\mathbf{w}$ and $\mathbf{w}' \in \mathbf{W}_j$ with $w_i > w_i'$, suppose $\mathbf{w}_{-i} < \mathbf{w}'_{-i}$. The following chain of inequalities results in the desired contradiction.

$$\begin{aligned}
0 &< \pi_j \left(w_i, \mathbf{w}_{-i} \right) - \pi_j \left(w_i \vee w_i', \mathbf{w}_{-i} \vee \mathbf{w}'_{-i} \right) \\
&\leq \pi_j \left(w_i \wedge w_i', \mathbf{w}_{-i} \wedge \mathbf{w}'_{-i} \right) - \pi_j \left(w_i', \mathbf{w}'_{-i} \right) < 0
\end{aligned}$$

The first and the last inequality result from optimality of $\mathbf{w}$ and $\mathbf{w}'$, the second inequality comes from the supermodularity shown above. ■

Evidence for this positive correlation between the wages of workers in different skill groups within firms was found by Katz and Summers (1989), who show evidence that secretaries earn more in firms where average wages are higher. More recently, Barth and Dale-Olsen (2003) find that "high-wage establishments for workers with higher education are high-wage establishments for workers with lower education as well". The explanation provided for this empirical observation in this paper rests on two pillars. Firstly, for each skill group the labor supply curve is upward sloping given the wage offer distribution is dispersed, which can be seen from equation (5). Secondly, the complementarity of skills in the production process guarantees that increasing both labor inputs simultaneously is optimal. The empirical regularity mentioned above justifies our choice of the production function, where labor inputs are complements.

Given that the skill group size is increasing in the wage w_i, it would be suboptimal if the support of the wage offer distributions was not a compact set.

Proposition 2 *The support of each skill specific wage offer distribution $F_i(w)$ is closed and connected, i.e. $supp(F_i) = [w_i^r, \overline{w}_i]$.*

Proof. Suppose not, i.e. no firms offer a wage $w_i \in (w_i^*, w_i^{**}) \subset [w_i^r, \overline{w}_i]$. This cannot be profit maximizing, since the firm offering w_i^{**} can offer $\lim_{\varepsilon \to 0} (w_i^* + \varepsilon)$, have the same skill group size, i.e. $l_i(w_i^{**} \mid w_i^r, F_i(w_i^{**})) = \lim_{\varepsilon \to 0} l_i((w_i^* + \varepsilon) \mid w_i^r, F_i(w_i^* + \varepsilon))$, since $\lim_{\varepsilon \to 0} F_i(w_i^* + \varepsilon) = F_i(w_i^{**})$, and can thus make higher profit. Thus, the support of the wage offer distribution is connected. By the same argument w_i^r is part of the support. The equal profit condition (7) together with the equation for the skill group size (5) implies that the support is also closed at the upper end. ■

Firms with different technologies j make potentially different profits π_j in equilibrium. We index the technologies according to their profitability, i.e. $\pi_j \geq \pi_{j-1} \forall j = 1, 2, ..., J$. The next proposition shows that for any skill group i more profitable firms pay higher wages.

Proposition 3 *Let $F_j : supp(F_j) = \left[\underline{\mathbf{w}}_j, \overline{\mathbf{w}}_j\right]$ and $F_{j-1} : supp(F_{j-1}) = \left[\underline{\mathbf{w}}_{j-1}, \overline{\mathbf{w}}_{j-1}\right]$ be the I-dimensional wage offer distributions of j and $j-1$-type firms respectively. Then, for any wage schedule $\mathbf{w}_j \in [\mathbf{w}^\mathbf{r}, \overline{\mathbf{w}}]$ and $\mathbf{w}_{j-1} \in [\mathbf{w}^\mathbf{r}, \overline{\mathbf{w}}]$ it is true that $\mathbf{w}_j \geq \mathbf{w}_{j-1}$.*

Proof. From the steady state equilibrium condition (7) it follows that:

$$\begin{aligned}
\pi_j &= Y_j\left(E\left[\mathbf{l}\left(\mathbf{w}_j\right)\right]\right) - \mathbf{w}_j^T E\left[\mathbf{l}\left(\mathbf{w}_j\right)\right] \quad \forall \mathbf{w}_j \in supp(F_j) \\
\pi_j &\geq Y_j\left(E\left[\mathbf{l}\left(\mathbf{w}_{j-1}\right)\right]\right) - \mathbf{w}_{j-1}^T E\left[\mathbf{l}\left(\mathbf{w}_{j-1}\right)\right] \quad \forall \mathbf{w}_{j-1} \notin supp(F_j)
\end{aligned}$$

Using the result above we can write

$$\begin{aligned}\pi_j &= Y_j(E[\mathbf{l}(\mathbf{w}_j)]) - \mathbf{w}_j^T E[\mathbf{l}(\mathbf{w}_j)] \geq Y_j(E[\mathbf{l}(\mathbf{w}_{j-1})]) - \mathbf{w}_{j-1}^T E[\mathbf{l}(\mathbf{w}_{j-1})] \\ &\geq Y_{j-1}(E[\mathbf{l}(\mathbf{w}_{j-1})]) - \mathbf{w}_{j-1}^T E[\mathbf{l}(\mathbf{w}_{j-1})] = \pi_{j-1} \geq Y_{j-1}(E[\mathbf{l}(\mathbf{w}_j)]) - \mathbf{w}_j^T E[\mathbf{l}(\mathbf{w}_j)],\end{aligned}$$

where the second inequality results from the fact that $\pi_j \geq \pi_{j-1}$.

The difference of the first and the last terms in this inequality is greater than or equal to the difference of its middle terms, i.e $Y_j(E[\mathbf{l}(\mathbf{w}_j)]) - Y_{j-1}(E[\mathbf{l}(\mathbf{w}_j)]) \geq Y_j(E[\mathbf{l}(\mathbf{w}_{j-1})]) - Y_{j-1}(E[\mathbf{l}(\mathbf{w}_{j-1})])$. Since $\mathbf{l}(\mathbf{w})$ is an increasing function of wages $\mathbf{w}$, the claim follows. ■

To be able to identify a particular technology in the empirical estimation, we assume that technologies strictly dominate each other by profits, i.e. $\pi_j > \pi_{j-1}$. Since Proposition 2 holds true for any wage pair $\mathbf{w}_j, \mathbf{w}_{j-1}$ and thus also for $\underline{\mathbf{w}}_j = \inf(\mathbf{w}_j)$ and $\overline{\mathbf{w}}_{j-1} = \sup(\mathbf{w}_{j-1})$, it follows that $\underline{\mathbf{w}}_j \geq \overline{\mathbf{w}}_{j-1}$. Thus, the more productive firms with technology j pay higher wages for all skill groups.

Furthermore, let γ_j denote the cumulative measure of technology j with $\gamma_j > \gamma_{j-1} > 0$ $\forall j = 1, 2, ..., J$ and $\gamma_J = 1$. Thus, Proposition 3 implies that the fraction of firms with technologies earning profit π_j or less post wages $\overline{\mathbf{w}}_j$ or below. Thus, for every skill group i the wage offer distribution at $\overline{w}_{ij}$is given by γ_j, i.e.

$$F_i(\overline{w}_{ij}) = \gamma_j \tag{8}$$

The next proposition shows under which condition it is not optimal for a type j firm to offer the same wage w_i as a mass of other type j firms does.

Proposition 4 *The wage offer distributions $F_i(w_i)$ of type j firms for skill group i is continuous, if*

$$\begin{aligned}&Y_j[E[l_i(w_i \mid w_i^r, F_i(w_i))], E[\mathbf{l}(\mathbf{w}_{-i})]] - Y_j\left[E\left[l_i\left(w_i \mid w_i^r, F_i\left(w_i^-\right)\right)\right], E[\mathbf{l}(\mathbf{w}_{-i})]\right] \\ > \;& w_i\left(E[l_i(w_i \mid w_i^r, F_i(w_i))] - E\left[l_i\left(w_i \mid w_i^r, F_i\left(w_i^-\right)\right)\right]\right). \end{aligned} \tag{9}$$

If a mass point exists, then it can only exist at the upper bound of the support of $F_i(w_i)$, i.e. $F_i\left(w_i^-\right) = \gamma_j - \upsilon_i(\overline{w}_{ij})$.
If the marginal product at the upper bound of the support of $F_i(w_i)$ exceeds $\overline{w}_{ij}$, i.e.

$$\frac{\partial Y_j[E[\mathbf{l}(\overline{\mathbf{w}})]]}{\partial E[l_i(\overline{w}_{ij} \mid w_i^r, \gamma_j)]} > \overline{w}_{ij}, \tag{10}$$

then a mass point can be ruled out.

Proof. Equation (6), and the fact that the cumulative density function $F_i(w_i)$ is right continuous implies

$$\begin{aligned}
& \lim_{\varepsilon \to 0} \pi_j (w_i + \varepsilon, \mathbf{w}_{-i}) + \mathbf{w}_{-i}^T E [\mathbf{l} (\mathbf{w}_{-i})] \\
= \; & Y_j [E [l_i (w_i \mid w_i^r, F_i (w_i))], E [\mathbf{l} (\mathbf{w}_{-i})]] - w_i E [l_i (w_i \mid w_i^r, F_i (w_i))] \qquad (11) \\
> \; & Y_j \left[E \left[l_i \left(w_i \mid w_i^r, F_i \left(w_i^-\right)\right)\right], E [\mathbf{l} (\mathbf{w}_{-i})]\right] - w_i E \left[l_i \left(w_i \mid w_i^r, F_i \left(w_i^-\right)\right)\right] \\
= \; & \pi_j (\mathbf{w}) + \mathbf{w}_{-i}^T E [\mathbf{l} (\mathbf{w}_{-i})]
\end{aligned}$$

since $F_i(w_i) - F_i(w_i^-) = \upsilon_i(w_i) > 0$. If the above inequality holds, no mass point can exist at w_i.

To show that a mass point can only exist at the upper bound of the support of $F_i(w_i)$ suppose that a mass point exists in the interior of the support at $w_i \in \left(\underline{w}_{ij}, \overline{w}_{ij}\right)$. The equal profit condition implies $\pi_j (\mathbf{w}) = \lim_{\varepsilon \to 0} \pi_j (w_i + \varepsilon, \mathbf{w}_{-i})$ and

$$\frac{\Delta \pi_j (\mathbf{w})}{\Delta w_i} \Delta w_i + \frac{\Delta \pi_j (\mathbf{w})}{\Delta E [l_i (w_i)]} \Delta E [l_i (w_i)] = 0,$$

where $\Delta w_i = \varepsilon$ and $\Delta E [l_i (w_i)] = E [l_i (w_i)] - E \left[l_i \left(w_i^-\right)\right]$. Rearranging implies

$$\frac{\Delta E [l_i (w_i)]}{\Delta w_i} = \frac{E [l_i (w_i)]}{\dfrac{Y_j [E [l_i (w_i)], E [\mathbf{l} (\mathbf{w}_{-i})]] - Y_j \left[E \left[l_i \left(w_i^-\right)\right], E [\mathbf{l} (\mathbf{w}_{-i})]\right]}{E [l_i (w_i)] - E \left[l_i \left(w_i^-\right)\right]} - w_i},$$

where $E \left[l_i \left(w_i^-\right)\right] = E \left[l_i \left(w_i \mid w_i^r, F_i \left(w_i^-\right)\right)\right]$. Since equation (5) together with Proposition 2 implies that $E [l_i (w_i)]$ is strictly increasing in w_i on its support $\left[\underline{w}_{ij}, \overline{w}_{ij}\right]$, i.e. $\Delta E [l_i (w_i)] / \Delta w_i > 0$, this expression is positive if and only if inequality (11) holds, i.e. only if no mass point exists. Thus, a mass point cannot exist in the interior of the support of $F_i (w_i)$ but only at the upper bound, i.e. $F_i \left(w_i^-\right) = \gamma_j - \upsilon_i (\overline{w}_{ij})$.

Rewriting inequality (11) and using the fact that $F_i \left(w_i^-\right) = \gamma_j - \upsilon_i (\overline{w}_{ij})$ gives

$$\frac{Y_j [E [l_i (w_i)], E [\mathbf{l} (\mathbf{w}_{-i})]] - Y_j \left[E \left[l_i \left(w_i^-\right)\right], E [\mathbf{l} (\mathbf{w}_{-i})]\right]}{E [l_i (w_i)] - E \left[l_i \left(w_i^-\right)\right]} > \overline{w}_{ij}.$$

A necessary condition for the absence of a mass point can be obtained by letting $\upsilon_i (\overline{w}_{ij}) \to 0$, i.e.

$$\lim_{\upsilon_i (\overline{w}_{ij}) \to 0} \frac{Y_j [E [l_i (w_i)], E [\mathbf{l} (\mathbf{w}_{-i})]] - Y_j \left[E \left[l_i \left(w_i^-\right)\right], E [\mathbf{l} (\mathbf{w}_{-i})]\right]}{E [l_i (w_i)] - E \left[l_i \left(w_i^-\right)\right]} = \frac{\partial Y_j [E [\mathbf{l} (\overline{\mathbf{w}})]]}{\partial E [l_i (\overline{w}_{ij} \mid w_i^r, \gamma_j)]}.$$

■

The basic argument as to why the wage offer distributions can be continuous is given by Burdett and Mortensen (1998). If all firms offer the same wage for one skill group, then individual firms could attract a significantly larger expected skill group size by offering a slightly higher wage. This wage increase can be arbitrarily small, whereas the resulting increase in the skill group size is significant, since all workers currently working for the "mass-point" wage will change to the new employer as soon as they get this higher wage offer. The deviation from a mass point is, thus, profitable if the increase in total output induced by a slight wage increase is higher than the increase in total wage cost. This is stated by the condition (9) in Proposition 4.

In order to be able to derive an explicit solution for the wage offer distribution, we continue under the assumption that no mass points exist. If all wage offer distributions are continuous, then an immediate result of Proposition 1 is that a firm occupies the same position in the wage offer distribution of every skill group. To formalize this, let us introduce an index k which orders the firms of type j as they increase their wage offer for skill group i. Then Proposition 1 implies that for all $\mathbf{w} \in \mathbf{W}_j$

$$F_{ij}^{k}(w) = F_{lj}^{k}(w) \text{ for all } i, l = 1, 2, ..., I. \tag{12}$$

To be able to use the above property, let us define

$$E\left[l_i\left(w \mid w_i^r, F_i\left(w_i\right)\right)\right] \equiv r_{ij} h_j\left(w\right),$$

where

$$h_j(w) = \frac{\left[\delta + \lambda_e\left(1 - \gamma_{j-1}\right)\right]^2}{\left[\delta + \lambda_e \overline{F}_j(w)\right]\left[\delta + \lambda_e \overline{F}_j\left(w^-\right)\right]}, \qquad r_{ij} = \frac{\delta\left(\delta + \lambda_e\right)\lambda_i / \left(\delta + \lambda_i\right)}{\left[\delta + \lambda_e\left(1 - \gamma_{j-1}\right)\right]^2} q_i.$$

The fact that $h_j(w)$ depends only on the position the firm takes in the wage offer distribution, i.e. on $F_j(w)$, implies that $h_j(w)$ does not depend on any skill specific parameter. Additionally we approximate the production technology j by using a second order Taylor expansion around the minimum wage $\underline{w}_{ij}$ that firms with technology j post. Given a technology $Y_j(\mathbf{r}_j)$ homogeneous of degree ξ_j, the Taylor Expansion is given by

$$Y_j\left(\mathbf{l}\left(\mathbf{w}_j\right)\right) = Y_j\left(\mathbf{r}_j\right) + \sum\nolimits_i Y_j'\left(\mathbf{r}_j\right)\left[r_{ij} h_j(w) - r_{ij}\right] + \frac{1}{2}\sum\nolimits_i \sigma_{ij}\left[h_j(w) - 1\right]^2,$$

where

$$Y_j'\left(\mathbf{r}_j\right) = \frac{\partial Y_j\left(\mathbf{r}_j\right)}{\partial l_i} \quad \text{and} \quad \sigma_{ij} = \sum\nolimits_l \frac{\partial^2 Y_j\left(\mathbf{r}_j\right)}{\partial l_i \partial l_l} r_{lj} r_{ij} = \left(\xi_j - 1\right) Y_j'\left(\mathbf{r}_j\right) r_{ij}.$$

We use the results of Propositions 1-4, invoke the equal profit condition $\pi_j = \pi_j^r$, apply the Taylor Expansion and use the first order condition to derive the skill-specific wage offer distribution. Proposition 5 provides the solution for $F_i(w_i)$ as a function of w_i.

Proposition 5 *Assume that the production functions $Y_j\left(E\left[\mathbf{l}\left(\mathbf{w}\right)\right]\right)\ \forall j = 1, 2, ..., J$ are supermodular and that no mass point exists. Then a unique equilibrium wage offer distribution $F_{ij}(w_i)$ for each skill group $i = 1, 2, ..., I$ exists and has the following form*
(i) for $\xi_j = 1$

$$F_{ij}(w_i) = \frac{\delta + \lambda_e}{\lambda_e} - \frac{\delta + \lambda_e(1-\gamma_{j-1})}{\lambda_e}\sqrt{\frac{Y_j'\left(\mathbf{r}_j\right) - w_i}{Y_j'\left(\mathbf{r}_j\right) - \underline{w}_{ij}}}, \tag{13}$$

(ii) for $\xi_j \neq 1$

$$F_{ij}(w_i) = \frac{\delta + \lambda_e}{\lambda_e} - \frac{\delta + \lambda_e\left(1-\gamma_{j-1}\right)}{\lambda_e\sqrt{\frac{\left(Y_j'(\mathbf{r}_j)-w_i\right)r_{ij}-\sigma_{ij}-\sqrt{\left(\left(Y_j'(\mathbf{r}_j)-w_i\right)r_{ij}-\sigma_{ij}\right)^2+4(\sigma_{ij}-\mu_{ij})\left(\left(Y_j'(\mathbf{r}_j)-\underline{w}_{ij}\right)r_{ij}-\mu_{ij}\right)}}{-2(\sigma_{ij}-\mu_{ij})}}} \tag{14}$$

for any $w_i \in [\underline{w}_{ij}, \overline{w}_{ij}]$, where

$$\mu_{ij} = \frac{r_{ij}}{\sum_i r_{ij}}\frac{1}{2}\sum\nolimits_i \sigma_{ij}.$$ [6]

A necessary condition for an upward sloping wage offer density $f_{ij}(w_i)$ is

$$\left(2-\xi_j\right)\frac{\partial Y_j\left(\mathbf{r}_j\right)}{\partial r_{ij}} - w_i > 0. \tag{15}$$

Proof. See Appendix. ■

The aggregate wage offer distribution is then given by

$$F(w) = \sum_{i=1}^{I}\frac{q_i}{m}F_i(w_i) = \sum_{i=1}^{I}\frac{q_i}{m}\sum_{j=1}^{J}\left(\gamma_j - \gamma_{j-1}\right)F_{ij}(w_i).$$

The comparative statics results of the original Burdett-Mortensen model are valid for the above aggregate wage offer distribution function as well. If the arrival rate of on-the-job offers, i.e. λ_e, goes to zero, then the wage offer distribution $F_i(w)$ collapses to

[6] A special case for $F_{ij}(w_i)$ when $\left(Y_j'\left(\mathbf{r}_j\right) - \underline{w}_{ij}\right)r_{ij} = \mu_{ij}$ is shown in the proof of Proposition 5. Since it implies artificial restrictions on ξ_j considering this case here is neither interesting nor useful.

a mass point at the reservation wage w_i^r, which equals the Diamond (1971) monopsony solution. If moving from one job to another becomes very easy, i.e. λ_e goes to infinity, the competition among firms drives wages up and the wage earnings distribution $G_i(w)$ converges to a mass point at the marginal product of the skill group.

For a production function with homogeneity of degree one, the explicit wage offer distribution resembles the distribution derived in Burdett and Mortensen (1998) and has its typical increasing density. As an upward-sloping earnings density is at odds with the empirically observed decreasing right tail, Mortensen (1990) introduces differences in firm productivity by allowing for different productivity levels in order to improve the fit to the empirical wage earnings distribution. Bowlus et al. (1995) demonstrate that this greatly improves the fit to the empirical earnings distribution. Bontemps et al. (2000) and Burdett and Mortensen (1998) formulate a closed-form solution for a continuous atomless productivity distribution, which translates into a right-tailed wage earnings distribution.[7]

The novelty of this paper is that the wage offer distribution given in Proposition 5 can have an increasing and a decreasing density for a given production technology. Although we allow for the possibility that heterogeneous production technologies are used, we do not need any technology dispersion to get a hump-shaped density. As stated in condition (15) only technologies with homogeneity of degree $2 > \xi_j$ can have an increasing density. Notice further that as the wage w increases, condition (15) is more likely to be violated implying that the wage offer density can have an upward sloping part for small wages and an downward sloping part for large wages.

The reason why increasing returns to scale can bend the wage offer density in such a way that it depicts a decreasing right tail, is the equal profit condition. Let us focus on the case with a homogenous production function with increasing returns to scale, i.e. $\xi_j > 1$ and compare it to an economy with constant returns to scale, where the marginal product of firms offering the reservation wage schedule are equivalent in both environments. Hence equilibrium profits are the same in both economies. First note that the skill group size and thus output is determined solely by the firm's position in the wage offer distribution. Consider now two firms sitting at the same position of the wage offer distribution, one with constant returns to scale the other with increasing returns to scale. Since the output of the firm with increasing returns is higher than the output of the firm with constant returns, the firm with increasing returns to scale has to pay higher wages due to the

[7]However, tail behavior of the productivity density, hence offer and earnings densities, in this case is subject to additional restrictions (see Bontemps et al., 2000; Proposition 8).

equal profit condition. Thus, the larger the returns to scale are, the larger is the wage difference paid by "neighboring" firms at the upper end of the wage offer distribution. Or, in other words, in an economy with increasing returns to scale the relative mass of firms sitting on a fixed interval decreases the closer we get to the upper bound. This mechanism eventually leads to a downward-sloping wage offer density in an economy with high enough returns to scale.

Remarkable enough, Mortensen (2000) also implicitly restricts his analysis to production functions with increasing returns to scale when deriving endogenously the employer heterogeneity based on match specific capital. He assumes that the production technology has constant returns with respect to labor but increasing economies of scale due to the capital k employed by the firm, i.e. $Y(l(w)) = k^{\alpha} l(w)$, where $\alpha > 0$. By simulation he shows that for positive α the wage offer distribution has a flat right tail.

Finally, consider the equilibrium earnings density $g_{ij}(w_i)$. From (15) follows that $\xi_j > 2$ is a sufficient condition for $f_{ij}(w_i)$ to have a decreasing right tail. The tail of the density function defined on $[\underline{w}_{i1}, \overline{w}_{iJ}]$ converges to zero at the fastest possible rate (see Bontemps et al., 2000, proof of Proposition 8). However letting $\overline{w}_{iJ}$ go to infinity we get the following result for the behavior of the earnings density function.

Proposition 6 *Let $\overline{w}_{iJ} \to \infty$. Under the sufficient condition for a decreasing right tail of $f_{iJ}(w_i)$ the right tail of the equilibrium earnings density $g_{iJ}(w_i)$ converges at a rate faster than w^{-2}. Speed of convergence is a power law that positively depends on the degree of homogeneity of the production function.*

Proof. Using (4) and (14), the closed form solution for the earnings density is

$$\begin{aligned} g_{iJ}(w_i) &= \frac{(\delta+\lambda_e) r_{iJ}}{2\lambda_e(\delta+\lambda_e(1-\gamma_{J-1}))} \\ &\times \frac{\sqrt{-\frac{\left(Y_J'(\mathbf{r}_J)-w_i\right)r_{iJ}-\sigma_{iJ}}{2(\sigma_{iJ}-\mu_{iJ})} + \frac{\sqrt{\left(\left(Y_J'(\mathbf{r}_J)-w_i\right)r_{iJ}-\sigma_{iJ}\right)^2+4(\sigma_{iJ}-\mu_{iJ})\left(\left(Y_J'(\mathbf{r}_J)-\underline{w}_{iJ}\right)r_{iJ}-\mu_{iJ}\right)}}{2(\sigma_{iJ}-\mu_{iJ})}}}{\sqrt{\left(\left(Y_J'\left(\mathbf{r}_J\right)-w_i\right)r_{iJ}-\sigma_{iJ}\right)^2+4\left(\sigma_{iJ}-\mu_{iJ}\right)\left(\left(Y_J'\left(\mathbf{r}_J\right)-\underline{w}_{iJ}\right)r_{iJ}-\mu_{iJ}\right)}}. \end{aligned}$$

Define

$$\begin{aligned} A(w_i) &\equiv \frac{\left(Y_J'(\mathbf{r}_J)-w_i\right)r_{iJ}-\sigma_{iJ}-\sqrt{\left(\left(Y_J'(\mathbf{r}_J)-w_i\right)r_{iJ}-\sigma_{iJ}\right)^2+4(\sigma_{iJ}-\mu_{iJ})\left(\left(Y_J'(\mathbf{r}_J)-\underline{w}_{iJ}\right)r_{iJ}-\mu_{iJ}\right)}}{-2(\sigma_{iJ}-\mu_{iJ})} \text{ and} \\ B(w_i) &\equiv \left(\left(Y_J'\left(\mathbf{r}_J\right)-w_i\right)r_{iJ}-\sigma_{iJ}\right)^2+4\left(\sigma_{iJ}-\mu_{iJ}\right)\left(\left(Y_J'\left(\mathbf{r}_J\right)-\underline{w}_{iJ}\right)r_{iJ}-\mu_{iJ}\right). \end{aligned}$$

Then the first derivative of $g_{iJ}(w_i)$ can be written as

$$g'_{iJ}(w_i) = -\frac{(\delta+\lambda_e)r_{iJ}^2}{2\lambda_e(\delta+\lambda_e(1-\gamma_{J-1}))}\left(\left[\frac{A(w_i)}{B(w_i)}\right]^{\frac{3}{2}} - \frac{3}{2}\left[\frac{A(w_i)}{B^2(w_i)}\right]^{\frac{1}{2}}\right).$$

For $\underline{w}_{iJ} \to \infty$ and $\overline{w}_{iJ} \to \infty$, we get that $A(w_i)$ is $O(1)$ and $B(w_i)$ is $O\left(w_i^{2(\xi_J-1)}\right)$, since σ_{iJ} is $O\left(w_i^{\xi_J-1}\right)$. This leads to

$$g'_{iJ}(w_i) \in O\left(w_i^{-2(\xi_J-1)}\right).$$

Finally, representing the sufficient condition for the decreasing right tail of the $f_{iJ}(w_i)$ as $\xi_j = 2 + \epsilon/2$, $\forall\, \epsilon > 0$, we see that $g'_{iJ}(w_i)$ is $O\left(w_i^{-2-\epsilon}\right)$. ■

The result of Proposition 6 implies that the right tail of the equilibrium earnings density encompasses the families of Pareto and Singh-Maddala distributions, which are acknowledged to have the best fit to the observed high-earnings data.[8] This result, as in Bontemps et al. (2000), also excludes the distributions with exponential speed of convergence, e.g. lognormal, from the set of possible candidates for the equilibrium earnings distribution. Finally, allowing for the increasing returns of the production function, we extend the result of Proposition 8 in Bontemps et al. (2000), demonstrating that the earnings density can converge both slower and faster then w^{-3}.

3. ECONOMETRIC MODEL

Now we consider the structural econometric model based on the theory presented above. For the empirical implementation we assume a Cobb-Douglas production technology

$$Y_j(l(\mathbf{w}_j)) = p_j \prod_{i=1}^{I} l_i(w_j)^{\alpha_{ij}}. \tag{16}$$

with homogeneity of degree $\xi_j = \sum_i \alpha_{ij}$, $\alpha_{ij} > 0$. The model is estimated by maximum likelihood using the methodology that builds on Bowlus et al. (1995, 2001).

3.1 *The Likelihood Function*

The likelihood function is constructed along the lines of van den Berg and Ridder (1998). For a Poisson process with rate θ, the joint distribution of the elapsed (t_e) and residual (t_r)

[8]See Singh and Maddala (1976). Additionally, McDonald (1984) shows that Singh-Maddala distribution outperforms the majority of the generalizations of the conventional earnings distributions.

duration of time spent by an individual in a certain state of the labor market is $f(t_e, t_r) = \theta^2 e^{-\theta(t_e+t_r)}$. For an individual that belongs to the i-th skill group the appropriate Poisson rates are λ_i if the person is unemployed and $\delta + \lambda_e [1 - F_i(w_i)]$ if the person is employed at wage w_i. Furthermore:

- *For the unemployed:* The equilibrium probability of sampling an unemployed agent who belongs to i-th skill group is $m^{-1} q_i \delta / (\delta + \lambda_i)$. In case the subsequent job transition is observed, we know the offered wage and can record the value of the wage offer density $f_i(w_i)$.
- *For the employed:* The equilibrium probability of sampling an agent who belongs to i-th skill group and earns wage w_i is $m^{-1} q_i g_i(w_i) \lambda_i / (\delta + \lambda_i)$. In case the transition to the next state is observed, we record the destination state. The probabilities of exit to unemployment and to next job are $\rho_{j\to u} = \delta / \left(\delta + \lambda_e \overline{F}_i(w_i)\right)$ and $\rho_{j\to j} = \lambda_e \overline{F}_i(w_i) / \left(\delta + \lambda_e \overline{F}_i(w_i)\right)$, respectively.

For convenience of the estimation we define $\kappa_i = \lambda_i/\delta$, $\kappa_e = \lambda_e/\delta$. Then the likelihood contributions of unemployed ($\ell_{(u)\,i}$) and employed ($\ell_{(e)\,i}$) individuals affiliated with the i-th skill group are

$$\ell_{(u)\,i} = \frac{q_i}{m(1+\kappa_i)} [\delta\kappa_i]^{2-d_r-d_l}\, e^{-\delta\kappa_i[t_e+t_r]} \, [f_i(w_i)]^{1-d_r}\,, \tag{17}$$

$$\begin{aligned}\ell_{(e)\,i} = g_i(w_i)\frac{q_i}{m}\frac{\kappa_i}{1+\kappa_i}\left[\delta\left(1+\kappa_e\overline{F}_i(w_i)\right)\right]^{1-d_l} e^{-\delta\left(1+\kappa_e\overline{F}_i(w)\right)[t_e+t_r]} \\ \times \left[\left[\delta\kappa_e\overline{F}_i(w_i)\right]^{d_t}\delta^{1-d_t}\right]^{1-d_r}.\end{aligned} \tag{18}$$

In (17) and (18) $d_l = 1$ if a spell is left-censored, 0 otherwise; $d_r = 1$ if a spell is right-censored, 0 otherwise; $d_t = 1$ if there is a job-to-job transition, 0 otherwise. Substitution of the appropriate $g_i(w_i)$, $f_i(w_i)$ and $F_i(w_i)$ into (17) and (18), where $g_i(w_i)$ is obtained from $F_i(w_i)$ using (4), completes the formulation of the likelihood function.

Notice that the individual contributions (17) and (18) are the same as in Bowlus et al. (1995, 2001) except of the probability terms $m^{-1}q_i/(1+\kappa_i)$ and $m^{-1}q_i\kappa_i/(1+\kappa_i)$. Though, the main differences of our model are driven by the functional forms of the offer and earnings distributions.

3.2 *Homogeneous Firms*

It is instructive to start with the model with no productivity dispersion, since the theory allows obtaining an earnings density with a decreasing right tail even with homogeneous employers. This density will have $I-1$ jumps at infimum wages and $I-1$ spikes at supremum wages of each skill group.

Consider the unknowns of the econometric model. The skill measures $\{q_i\}_{i=1}^{I}$ are known from the data and given by the sample sizes of each skill group. Furthermore, to avoid bounds of the likelihood function depending on the parameters, Kiefer and Neumann (1993) justify using the extreme order statistics $\{\min(w_i), \max(w_i)\}$ as the consistent estimates for $\underline{w}_i$ and $\overline{w}_i$ respectively. Under employer homogeneity the assumed production function modifies to $Y(l(\mathbf{w})) = p\prod_{l=1}^{I} l_l(w)^{\alpha_l}$. The functional form of the wage offer distribution with homogeneous employers is $F(w) = \sum_{i=1}^{I}\frac{q_i}{m}F_i(w_i)$, where $F_i(w_i)$ is given in Proposition 5 with $J=1$ and $\kappa_{i,e} = \lambda_{i,e}/\delta$. Recognizing that $F_i(\overline{w}_i) = 1$ and using $Y(l(\mathbf{w}))$ we get the following solution for the common productivity parameter [9]

$$p = \frac{r_i}{\prod_{i=1}^{I} r_i^{\alpha_i}}\left[\alpha_i - \frac{\xi-1}{\eta}\left(\frac{\xi\,(1+\eta)\,r_i}{2\sum_i r_i} - \alpha_i\right)\right]^{-1}\left(\frac{\overline{w}_i - \eta\underline{w}_i}{1-\eta}\right), \tag{19}$$

where $\eta \equiv (1+\kappa_e)^{-2}$.

Since (19) holds for any i one can represent any α_i as a function of ξ and the rest of structural parameters. Namely, for any $i, l = 1,..,I$ the following holds

$$\alpha_i\,\frac{(\overline{w}_l - \eta\underline{w}_l)\,r_l}{(\overline{w}_i - \eta\underline{w}_i)\,r_i} - \alpha_l = \frac{\xi\,(\xi-1)\,(1+\eta)\,r_l}{2\,(\xi+\eta-1)\sum_{k=1}^{I} r_k}\left[\frac{\overline{w}_l - \eta\underline{w}_l}{\overline{w}_i - \eta\underline{w}_i} - 1\right],$$

Without loss of generality setting $i = 1$, $l = 2,...,I$ and recognizing that $\alpha_1 = \xi - \sum_{k=2}^{I}\alpha_k$, we get a system of $I-1$ linear equations, which gives a unique solution for α in terms of $\left\{\{\kappa_i\}_{i=1}^{I}, \kappa_e, \delta, \xi\right\}$.[10] Since the frictions parameters $\left\{\{\kappa_i\}_{i=1}^{I}, \kappa_e, \delta\right\}$ are uniquely identified from the duration data irrespective of the functional form of the offer distribution (e.g. Koning et al., 1995), it follows that the production size ξ is uniquely identified from the labor costs data.

3.3 *Heterogeneous Firms*

Production functions for heterogeneous employers are given in (16). The relevant occupation-specific wage offer distribution $F_i(w)$ is provided in Proposition 5. Rewritten in $\kappa_{i,e}$ terms,

[9]Use equation (A.3) in the appendix with $F_i(\overline{w}_i) = 1$, $\sigma_i = \alpha_i\,(\xi-1)\,Y\,(\mathbf{r})$ and $Y'(\mathbf{r})\,/r_i = \alpha_i Y\,(\mathbf{r})$ for the derivation.

[10]To see this it is sufficient to rewrite the system in the matrix form. The matrix to be inverted will have a particular structure that never allows one row to be a linear combination of the others because $\frac{\overline{w}_l-\eta\underline{w}_l}{\overline{w}_i-\eta\underline{w}_i} > 0\ \forall\, i, l$.

it becomes

$$
\begin{aligned}
F_i(w_i) = {} & \frac{1+\kappa_e}{\kappa_e} \\
& - \frac{1+\kappa_e\left(1-\gamma_{j-1}\right)}{\kappa_e\sqrt{\frac{\big(Y_j'(\mathbf{r}_j)-w_i\big)r_{ij}-\sigma_{ij}-\sqrt{\big(\big(Y_j'(\mathbf{r}_j)-w_i\big)r_{ij}-\sigma_{ij}\big)^2+4(\sigma_{ij}-\mu_{ij})\big(\big(Y_j'(\mathbf{r}_j)-\underline{w}_{ij}\big)r_{ij}-\mu_{ij}\big)}}{-2(\sigma_{ij}-\mu_{ij})}}}, \qquad (20)
\end{aligned}
$$

where

$$
\begin{aligned}
r_{ij} &= \frac{\kappa_i/\left(1+\kappa_i\right)\left(1+\kappa_e\right)}{\left[1+\kappa_e\left(1-\gamma_{j-1}\right)\right]^2}q_i, \qquad Y_j'\left(\mathbf{r}_j\right) = \frac{\alpha_{ij}}{r_{ij}}p_j\prod\nolimits_{i=1}^{I} r_{ij}^{\alpha_{ij}}, \\
\sigma_{ij} &= \alpha_{ij}\left(\xi_j-1\right)Y_j\left(\mathbf{r}_j\right), \quad \text{and} \quad \mu_{ij} = \frac{r_{ij}}{\sum_i r_{ij}}\frac{1}{2}\sum\nolimits_i \sigma_{ij}
\end{aligned}
$$

for all $w_i \in [\underline{w}_{ij}, \overline{w}_{ij}]$, $i = 1, ..., I$ and $j = 1, ..., J$.

Remembering that $\gamma_j = F_i(\overline{w}_{ij})$, we can use (16) and (20) to derive the productivity level of the firm

$$
p_j = \frac{r_{ij}}{\prod_{i=1}^{I} r_{ij}^{\alpha_{ij}}}\left[\alpha_{ij} - \frac{\xi_j-1}{\eta_j}\left(\frac{\xi_j\left(1+\eta_j\right)r_{ij}}{2\sum_i r_{ij}} - \alpha_{ij}\right)\right]^{-1}\left(\frac{\overline{w}_{ij}-\eta_j\underline{w}_{ij}}{1-\eta_j}\right), \qquad (21)
$$

where $\eta_j = \left[\left(1+\kappa_e[1-\gamma_j]\right)/\left(1+\kappa_e[1-\gamma_{j-1}]\right)\right]^2$.

Consider the unknowns of the econometric model with heterogeneous firms. As before, skill group size and group-specific bounds for the offer distributions are available from the data. At the same time there appears an additional set of unknown cutoff wages $\{\overline{w}_{ij}\}_{i,j=1}^{I,J-1}$ for the firm-specific wage offer. Unlike in the homogeneous model, the existence of the unknown cutoff wages does not allow us to use equation (21) to write down α_{ij} as a function of exclusively ξ_j and frictions parameters. However, knowing that $\overline{w}_{ij} = \underline{w}_{ij-1}$ provides us with additional cross-restrictions on p_{j-1} and p_j. Using these cross-restrictions together with the fact that (21) is the same for all i and noticing that the parameter subsets $\{\alpha_{ij}\}_{i,j=1}^{I-1,J}$ and $\{\overline{w}_{ij}\}_{i,j=1}^{I,J-1}$ are completely determined by (21), two representations of the model are possible:

1. Cutoff wages $\{\overline{w}_{ij}\}_{i,j=1}^{I,J-1}$ can be expressed as a function of production parameters $\{\alpha_{ij}\}_{i,j=1}^{I-1,J}$, search friction parameters $\left\{\{\kappa_i\}_{i=1}^{I}, \delta, \kappa_e\right\}$ and the returns to scale parameters $\{\xi_j\}_{j=1}^{J}$,

2. Production parameters $\{\alpha_{ij}\}_{i,j=1}^{I-1,J}$ can be expressed as a function of cutoff wages $\{\overline{w}_{ij}\}_{i,j=1}^{I,J-1}$, search friction parameters $\left\{\{\kappa_i\}_{i=1}^{I}, \delta, \kappa_e\right\}$ and the returns to scale parameters $\{\xi_j\}_{j=1}^{J}$.

Irrespective of the choice of the parameter subset to be substituted out, (21) implies that there exist $J(I-1)$ independent equations that completely determine cutoff wages and production parameters, because neither $\{\overline{w}_{ij}\}_{i,j=1}^{I,J-1}$ nor $\{\alpha_{ij}\}_{i,j=1}^{I-1,J}$ appear outside the system of these equations. Moreover, for I skill groups there exist $(J-1)I$ unknown production parameters and $J(I-1)$ unknown cutoff wages. Since both representations must be equivalent to each other, we conclude that the parameters cannot be identified whenever $J(I-1) \neq (J-1)I$. From this follows that $I = J$ symmetry is a necessary condition for identification of the model with employer heterogeneity.

Although both specifications are equally possible, expressing cutoff wages as a function of the rest of the parameters is a strictly dominated one because cutoff wages are the discontinuity points of the likelihood function. Thus, substituting them with known functions of the rest of the parameters means that no gradient-based methods can be used to estimate the model. Even though derivative-free methods are available, a serious problem may appear when the assumption of no mass points in the offer distribution stated in Proposition 4 becomes a binding restriction. We therefore choose the second way to represent the model. Equation (21) then implies that for any $i, l = 1, .., I$ the following identity holds, i.e.

$$\alpha_{ij} \frac{\left(\overline{w}_{lj} - \eta_j \underline{w}_{lj}\right) r_{lj}}{\left(\overline{w}_{ij} - \eta_j \underline{w}_{ij}\right) r_{ij}} - \alpha_{lj} = \frac{\xi_j \left(\xi_j - 1\right)\left(1 + \eta_j\right) r_{lj}}{2\left(\xi_j + \eta_j - 1\right) \sum_{k=1}^{I} r_{kj}} \left[\frac{\overline{w}_{lj} - \eta_j \underline{w}_{lj}}{\overline{w}_{ij} - \eta_j \underline{w}_{ij}} - 1\right].$$

This gives rise to a system of $J(I-1)$ linear equations with $J(I-1)$ unknown cutoff wages. It is also easy to see that for $J = 1$ the above identity reduces to the one described in the previous subsection. Rewriting the implied system in a matrix form, one can find that the matrix to be inverted is block-diagonal. Each and every block in it has the same structure as the matrix of a corresponding problem in section 3.2, out of which invertability follows.

The unique solution for $\{\alpha_{ij}\}_{i,j=1}^{I-1,J}$ reduces the parameter space to the subset of the location parameters of the discontinuity points of the likelihood function $\{\overline{w}_{ij}\}_{i,j=1}^{I,J-1}$ and the subset of shape parameters $\theta \equiv \left\{\{\kappa_i\}_{i=1}^{I}, \delta, \kappa_e, \{\xi_j\}_{j=1}^{J}\right\}$. Chernozhukov and Hong (2004) demonstrate that in the considered class of models shape and location parameters are independent of each other. Thus conditional identifiability will imply joint identifiability of the both. Within the subset of shape parameters search frictions are uniquely identified using the duration data. From this follows that production sizes are uniquely identified from the labor costs data.

The above representation of the model fits into a convenient stepwise estimation strategy developed by Bowlus et al. (1995, 2001). At the first step, given the starting values for the structural parameters, cutoff wages are estimated by simulated annealing. At the second step, given the estimates of the cutoff wages, the likelihood function is maximized with respect to θ. The second step is a "smooth" optimization and can be efficiently executed using gradient-based methods. Substituting the estimates from both steps into (4) and (8) we calculate the new point mass values γ_j

$$\gamma_j = 1 - \sum_{i=1}^{I} \frac{q_i}{m} \frac{1 - \hat{G}_i(\overline{w}_{ij})}{1 + \kappa_e \hat{G}_i(\overline{w}_{ij})}, \tag{22}$$

where $\hat{G}_i$ is a nonparametric estimate of the skill-specific earnings distribution, and the cycle repeats.

Provided that the maximum likelihood estimates satisfy the condition stated in Proposition 4, we can apply the result of Chernozhukov and Hong (2004) who show that the asymptotic distribution of the subset of shape parameters is $N(0, \mathbf{I}^{-1})$, where

$$\mathbf{I} = n^{-1} \sum_{l=1}^{n} \frac{\partial \ell_l(\theta)}{\partial \theta} \frac{\partial \ell_l(\theta')}{\partial \theta}. \tag{23}$$

Furthermore, when the restriction of Proposition 4 is not binding, Chernozhukov and Hong (2004) show that the bootstrap also consistently estimates the asymptotic covariance matrix above.

3.4 *Specification Check*

We have derived the wage offer distribution (14) under the assumption that all skill specific wage offer distributions $F_i(w_i)$ are continuous. Consider an arbitrary skill group l. Proposition 4 implies that the distribution function $F_l(w_l)$ is continuous if condition (10) is satisfied, i.e.

$$\alpha_{lj} \frac{p_j \prod_{i=1}^{I} l_i(\overline{w}_{ij})^{\alpha_{ij}}}{l_l(\overline{w}_{lj})} > \overline{w}_{lj}. \tag{24}$$

The estimated parameters are consistent only when the model is properly specified, i.e. when (24) holds. In case (24) is violated at the unconstrained maximum, constrained MLE must be calculated.[11]

[11] One can also notice that with no skill differentiation, constant returns and identical employers, (24) reduces to $1 > \overline{w}/p$ implying continuous offer distribution in the original Burdett-Mortensen model.

Furthermore, the estimated parameters must be consistent with the assumption that profits of the firms with different technologies are ranked, i.e.

$$0 \leq \pi_{j-1} < \pi_j. \tag{25}$$

In conclusion, we also like to point out that whenever any of the above restrictions is binding at the maximum the asymptotic covariance matrix of the ML estimator is no longer given by (23) and the exact form of it is unknown. Moreover even in simpler models with binding inequality constraints it is shown that bootstrap fails to consistently estimate the covariance matrix of the true parameters (see Andrews, 2000, for a discussion). Given that in the literature a consistent covariance matrix estimator for the cases where the inequality restrictions are binding at the maximum (even for smooth likelihood functions) has so far not been derived, we will present the confidence intervals using (23) to provide the reader with at least rough information about the size of standard errors. Though, when interpreting these confidence intervals, caution is necessary.

4. EMPIRICAL APPLICATION

4.1 *The Data*

We use data from the German Socio-Economic Panel – a longitudinal survey of German households which was started at 1984 and conducted on the annual basis ever since. Our sample contains information from the waves of 1984 to 2001. The analysis is restricted to the working age population of native West Germans and major groups of foreigners living in West Germany.

According to the theoretical model, we have only two states, namely "full time employment" and "unemployment". Since utility maximizing behavior of the representatives of the other groups, such as part-time employed, self-employed or non-participants can be different from behavior of the individuals considered by the model, we exclude all the agents who are neither full time employed nor unemployed from the sample (as in Koning et al., 1995, and Bontemps et al., 2000).

To estimate the model we need information on both duration and wages. We get duration information by choosing a reference year and sampling all employed and unemployed individuals at this year. After doing so, for each observation we track the individual history backwards and forwards to restore the elapsed and residual duration of his/her staying in the current state of the market. Whenever a residual spell is complete, we also record information about the exit state. Retrieving the duration lengths proceeds as described in Bontemps et al. (2000). The reference year is set to 1995.

Unlike in the rest of empirical equilibrium search models, when collecting the wage data we differentiate between net wage received by the worker and labor costs to the firm. In the theoretical model we have two sets of parameters, namely workers' search intensities and production parameters. Since the theory states that reservation wage and labor size depend on just the position of the firm in the wage offer distribution, frictional parameters can be estimated using any of these two types of earnings data, since the ordering of the firms does not change by taking labor costs instead of net wages. For identification of the production parameters, to the contrary, labor costs are crucial because the magnitude of the costs of production influences the size of estimated factor elasticities. Therefore the labour costs and not net wages are used for the estimation.

GSOEP provides the data on both net and gross wages. Individuals who are employed at their interview provide the earnings information of one month prior to the interview. For the unemployed we use the first reported wage after the end of unemployment, provided that the transition to the job is observed. All wages are deflated by the West German consumer price index at prices of 1998. Labor costs are defined as the sum of gross wage and firms' contributions to the employees' social security payments. Information on the latter is available, for instance, form the publications of the Federation of German Pension Insurance Institutes ("Verband Deutscher Rentenversicherungsträger"; see VDR, 2004, p.243, 245).

In our application we estimate the model with three different productivity levels and three different skill groups. Skill stratification of the sample is performed on the basis of the International Standard Classification of Education (ISCED) of 1997.

We define as "low- skilled" all individuals who have inadequate or general elementary training, i.e. codes "1" and "2". Individuals with middle vocational training, i.e. code "3", represent the "medium-skilled" group. Finally, as "high-skilled" we qualify all those with higher vocational training, university education etc, i.e. codes "4" to "6".

A summary of duration and wage data is presented in Tables 1 and 2 both for full sample and distinct skill groups. Skill differentiation reflects such basic facts about less skilled in comparison to higher skilled as higher level of unemployment, higher rate of job loss, longer unemployment duration. Additionally net wages and labor costs are summarized by kernel density plots (see Figures A.1-2 in the Appendix). As expected,

Table 1: Descriptive Statistics of Event History Data *

	Skills			Full Sample
	Low	Medium	High	
Number of individuals	898	1931	1062	3891
Employed	746	1786	1025	3557
Unemployed	152	145	37	334
Employed Agents				
Uncensored observations with				
job → job transition	49	187	178	414
job → unemployment transition	98	126	41	256
Mean time spell between two states [job duration]	129.639	109.815	89.566	107.576
(std. deviation)	(114.92)	(102.14)	(85.42)	(101.01)
Censored observations				
a) Left-censored durations only				
with job → job transition	3	12	6	21
with job → unemployment transition	1	13	1	15
b) Right-censored durations only	575	1407	781	2763
c) Both left- and right-censored durations	20	41	18	79
Mean time spell [both uncensored and censored]	163.637	153.259	154.096	155.677
(std. deviation)	(116.23)	(118.84)	(120.30)	(118.76)
Unemployed Agents				
Uncensored observations (u → j transition)	37	49	13	99
Mean time spell between two states [job duration]	19.595	22.429	10.538	19.808
(std. deviation)	(14.35)	(26.72)	(12.22)	(21.41)
Censored observations				
a) Left-censored durations (u → j transition) only	1	2	-	3
b) Right-censored durations only:	106	89	24	219
c) Both left- and right-censored durations	8	5	-	13
Mean time spell [both uncensored and censored]	40.974	32.310	24.270	35.362
(std. deviation)	(36.37)	(31.90)	(23.07)	(33.61)

* *Duration data in Months*

Table 2: Descriptive Statistics of Earnings Data

	Skills			Full Sample
	Low	Medium	High	
Labour Costs				
Sample Minimum	734.	1038.	1646.	734.
Mean Cost	4431 (1417)	5245 (1903)	6950 (2642)	5554 (2258)
Sample Maximum	12057.	17348.	20523.	20523.
Net Wages				
Sample Minimum	604.	635.	952.	604.
Mean Wage	2472 (809)	2880 (1083)	3967 (1667)	3101 (1356)
Sample Maximum	6878.	9524.	11534.	11534.

density of both net wages and labor costs of the low-skilled are more peaked at its' leftmost part of the support than those of the higher skills. Also mean net wage of high-skilled workers amounts to DM 3967 which exceeds that of medium-skilled by 27% and of low-skilled by more then 37%. Labor costs are roughly the same across the skills and almost double the net wage.

Finally, comparing the duration statistics for the full sample with that of Bontemps et al. (2000) we can see that both West German and French data are of about the same magnitude.

4.2 *Estimation Results: Fit of the Model*

First we estimate the model with identical employers setting off with the constant returns assumption (see Table A.1 in the Appendix). When doing so, we also fit the original Burdett-Mortensen model with no productivity dispersion to compare it with the results provided by our extension.[12] It turns out that the structural parameters estimated with both original model and our extension with constant-returns specification do not sig-

[12] For the sake of brevity here and henceforward we do not report the estimates from the original Burdett-Mortensen model.

nificantly differ from each other. This implies that from the empirical perspective the sole introduction of skill differences does not improve the estimates of search frictions. Predicted theoretical offer and labor costs densities (Figures A.3-4 respectively) for the extended theoretical model with constant returns production function have two jumps at the reservation wages of the medium- and high-skilled workers and two spikes at the maximum wages of the low- and medium-skilled workers. This generates a "quasi"-falling right tail of the aggregate density despite that skill-specific ones are strictly increasing. However, even with large I the model with constant returns has limited potential of fitting the data.

The fit of the model improves when we relax the assumption of a constant returns production technology (the second column in Table A.1). Along with statistically significant increasing returns to scale we find that, when inserted into the unemployment equation (3), the estimates of κ_i and δ match the observed skill-specific unemployment rates closer. Though the most interesting result is displayed in Figures A.3-4 where we see that increasing returns imply the offer and labor costs densities with strictly decreasing right tails even in absence of productivity dispersion. Even though the predicted labor costs density is still too flat pointing towards existence of employer heterogeneity in the data, this result alone is already remarkable.

The initial unrestricted estimates of the model with variable returns to scale and identical employers do not meet the "no mass point condition" of Proposition 4. Therefore the results reported in the second column of Table A.1 are those obtained by maximizing the likelihood function subject to (24). Furthermore we restrict profits to be non-negative. It turns out that at the constrained maximum the condition in (24) is not binding. However, the non-negativity of profits is violated on the upper end of the offer distribution. As a consequence the non-negativity constraint on the firms profit is binding at the maximum.[13]

Next we estimate the model with employer heterogeneity and constant returns technology (Table A.2, column one). As before, we also fit the original Burdett-Mortensen model with $J = 3$. Again, the parameters we get from the original Burdett-Mortensen model and from our extension with constant returns technologies hardly differ. Even though skill multiplicity eventually provides a better fit of the predicted labour costs density, convex spikes and locally increasing right tail still remain the negative feature of the

[13] From this also follows that the asymptotic covariance matrix of the estimated parameters is unknown. We report confidence intervals based on (23). However, since the true parameters lie on the boundary of the parameter space, (23) does not provide correct values (see also Section 3.4).

constant-returns specification (see Figure A.6). Furthermore, with the constant returns to scale technologies increasing the number of skill/productivity types leads to steeper spikes in the predicted theoretical densities.

Allowing for increasing returns once again improves the fit of the model considerably. Though, as in the case with identical firms, the unrestricted MLEs still violate the profit ranking. Therefore we perform the estimation of the model given (24) and (25). Remarkable enough, at the restricted maximum the "no mass point condition" of Proposition 4 is again inactive which provides empirical support for the k-percent rule (12). However, the ranking constraints $\pi(\overline{w}_{ij-1}) < \pi(\underline{w}_{ij})$ turn out to be binding. On the one hand, this might simply be a consequence of the insufficient heterogeneity of the production side. On the other hand, this can also be interpreted as an empirical indication of the restrictiveness of the equal-profit condition among firms of the same productivity type. For instance, it may be the case that firms differ in the size of the capital stock which implies different profit levels even though the technology they use could be the same.[14]

The estimates of the model with increasing returns and three-point productivity dispersion are presented in the second column of Table A.2. Comparing them to the estimates from the specification with identical firms and increasing returns technology two important improvements can be noticed. First, we manage to obtain a better fit for the magnitude of returns to scale in the whole economy. According to our estimates the degree of homogeneity is 1.04 for the "low-productive" technology, 1.40 for the "medium-productive" technology and 4.92 for the "high-productive" one. Given the estimated fraction of each technology $[\gamma_j - \gamma_{j-1}]$ in the economy, these estimates imply the economy-wide returns to scale at the level of 1.20. This is in line with numerous evidences from the literature on the estimation of the returns to scale using different types of production functions. Typical estimates in this literature support the increasing returns hypothesis and range from about 1.1 to about 1.35 (see Färe at al., 1985, Kim, 1992, and Zellner and Ryu, 1998, and references therein). Second, and even more important, productivity dispersion with increasing returns technologies leads to much better fitting offer and labor costs densities. In Figures A.5-6 one can easily see the dominance of the increasing over constant returns specification in terms of both shape of the right tail and smoothed out spikes around the mean.

Finally, Figures A.7-8 present the skill-specific components of the aggregate offer and

[14] A similar possibility of violating the productivity ranking in the original Burdett-Mortensen model with employer heterogeneity is discussed by Bowlus et al. (1995), p.S127.

labour costs densities. As expected, for every higher skill level, they mirror the rightward shift of the probability of getting a better offer.

4.3 *Estimation Results: Social Returns to Education*

We use our estimation results to investigate whether the education level in the economy is efficient, i.e. whether the social return to education measured by the increase in output resulting from educating the marginal individual (that is indifferent between acquiring the skill levels $i-1$ and i) equals the private return of the marginal individual.

Following Grout (1984), who discusses the hold-up problem as a potential source of underinvestment, Acemoglu (1996) and Masters (1998) develop models where underinvestment results from the fact that search or matching frictions make it impossible for workers to capture the whole return on their investment. However, there can also be overinvestment in our model, because the assumption of skill-segmented labor markets makes it possible that a lower unemployment rate among high skilled workers can increase the return to human capital investment to such a degree that workers overinvest in skills.[15]

To be able to investigate the question of whether there is over- or underinvestment, we first analyze the social planner's problem who has to allocate each individual to a specific skill level. Since we assume that workers are risk neutral, the distribution of income does not matter for the aggregate welfare function. Thus, the social planner maximizes total output produced by all firms minus the aggregate cost of education.

Aggregate output is obtained by integrating from the firm offering the reservation wage schedule, i.e. $F_{i1}\left(w_i^r\right) = 0$, to the firm offering the maximum wage to all skill groups, i.e. $F_{iJ}\left(\overline{w}_i\right) = 1$. Since our theoretical model predicts that each firm's labor input is uniquely identified by the firm's position in the wage offer distribution F, we can write aggregate output by

$$E(Y) = \int_0^1 Y_j(\mathbf{l}(F))dF.$$

Given that the individual cost $c_{i,a}$ of acquiring skill level i is inversely related to ability a, i.e. $c_{i,a} = c_i/a$, and that the skill specific component c_i is increasing in the skill level, i.e. $c_i > c_{i-1}$, the social planer will ask high ability workers to acquire a high skill level and low ability workers to acquire a low skill level. Given the ability distribution $H(a)$ among individuals on the support $a \in [\underline{a}, \overline{a}]$ with $\underline{a} > 0$ the social planer will divide the population into separated ability segments such that the measure of workers with skill

[15] This is due to the assumption of segmented labor markets for all skill groups. If we assumed a constant arrival rate across all unemployed workers, the theoretical model would predict underinvestment.

i is given by $q_i = m\left[H\left(a_{i+1}\right) - H\left(a_i\right)\right]$, where a_i equals the lowest ability type in skill group i with $a_1 = \underline{a}$ and $a_{I+1} = \overline{a}$. Thus, choosing the ability type worker a_i is identical to choosing the measure q_i of workers with skill i. The average cost of education incurred by the individuals that the social planner asks to acquire skill level i is given by

$$E\left[c_{i,a}|a_{i+1} \geq a \geq a_i\right] = \int_{a_i}^{a_{i+1}} \frac{c_i}{a} dH\left(a\right).$$

Thus, assuming I skill levels the social planner's problem is given by

$$\begin{gathered}\left\{a_i^S\right\}_{i=2}^I = \arg\max_{\left\{a_i\right\}_{i=2}^I} \left[\int_0^1 Y_j(\mathbf{l}(F))dF - m\sum_{i=1}^I \int_{a_i}^{a_{i+1}} \frac{c_i}{a} dH\left(a\right)\right] \\ \text{s.t. } q_i = m\left[H\left(a_{i+1}\right) - H\left(a_i\right)\right] \;\; \forall \;\; i \in I, \\ \sum_{i=1}^I q_i = m,\, a_1 = \underline{a},\, a_{I+1} = \overline{a}\end{gathered}$$

It follows that the socially efficient skill structure is characterized by

$$\int_0^1 \frac{\partial Y(\mathbf{l}(F))}{\partial q_i} dF \bigg|_{\sum_{i=1}^I q_i = m} = \left(c_i - c_{i-1}\right)/a_i^S \;\; \forall \;\; i \in I,$$

i.e. the social welfare is maximized if the cost the marginal individual incurs equals the output-increase generated by all firms.[16]

Denote the measure of any adjacent skill groups by n_i so that $n_i = q_i + q_{i-1}$. It is easy to show that for a j-type firm the marginal change in output due to educating a marginal i-skilled worker is

$$\begin{aligned}\frac{\partial Y_j(\mathbf{l}(F))}{\partial q_i}\bigg|_{n_i = q_i + q_{i-1}} &= Y_j(\mathbf{l}(F))\left[\sum_{i=1}^I \frac{\alpha_{ij}}{l_i(F)} \frac{\partial l_i(F)}{\partial q_i}\right] \\ &= Y_j(\mathbf{l}(F))\left[\frac{\alpha_{ij}}{q_i} - \frac{\alpha_{i-1j}}{n_i - q_i} + \frac{2\kappa_e\left(\alpha_{ij} + \alpha_{i-1j}\right)}{1 + \kappa_e\left[1 - F\right]}\left(\frac{\partial F}{\partial q_i}\right)\right],\end{aligned}$$

[16] The first order condition is given by:

$$\begin{aligned}\int_0^1 \frac{\partial Y(\mathbf{l}(F))}{\partial q_i} dF\bigg|_{\sum_{i=1}^I q_i = m} \times \frac{\partial m\left[H\left(a_{i+1}\right) - H\left(a_i\right)\right]}{\partial a_i} &= \frac{\partial m\left[\int_{a_i}^{a_{i+1}} \frac{c_i}{a} dH\left(a\right) + \int_{a_i - 1}^{a_i} \frac{c_{i-1}}{a} dH\left(a\right)\right]}{\partial a_i} \\ \int_0^1 \frac{\partial Y(\mathbf{l}(F))}{\partial q_i} dF\bigg|_{\sum_{i=1}^I q_i = m} \times \left(-mh\left(a_i^S\right)\right) &= mh\left(a_i^S\right)\frac{-c_i + c_{i-1}}{a_i^S}\end{aligned}$$

Furthermore, we assume that the skill-specific components c_i are such that a solution to the social planner's problem exists.

which implies an expected change in the total output

$$E(\Delta Y) = \int_0^1 \left. \frac{\partial Y_j(\mathbf{l}(F))}{\partial q_i} \right|_{n_i = q_i + q_{i-1}} dF = \sum_{j=1}^{J} \int_{\gamma_{j-1}}^{\gamma_j} \left. \frac{\partial Y_j(\mathbf{l}(F))}{\partial q_i} \right|_{n_i = q_i + q_{i-1}} dF. \tag{26}$$

In order to see whether the social returns from educating an individual to a higher skill level exceed the private returns of doing so, we proceed comparing the marginal increase in output caused by a change in the skill structure with the private return the marginal individual gets from acquiring this skill level.

In equilibrium it has to be true that the marginal worker is exactly indifferent between the two skill groups, i.e. $U_i = U_{i-1}$. Thus, using (1a), the private return to educating oneself from the "low" to the "high" level can be written as

$$\begin{aligned} rU_i - rU_{i-1} &= \kappa_i \int_{w_i^r}^{\bar{w}_i} \frac{\bar{F}_i(w)}{1 + r/\delta + \kappa_e \bar{F}_i(w)} dw - \kappa_{i-1} \int_{w_{i-1}^r}^{\bar{w}_{i-1}} \frac{\bar{F}_{i-1}(w)}{1 + r/\delta + \kappa_e \bar{F}_{i-1}(w)} dw \\ &= \left(c_i - c_{i-1}\right) / a_i^I. \end{aligned} \tag{27}$$

Note, that (27) refers to the optimal decision of unemployed *individual*, which implies that the net wages w_i^r and $\overline{w}_i$ – not the wage costs – are the bounds of the distribution of the net offer. Therefore in order to compute the correct private returns we have to translate the estimated cutoff wages expressed in terms of labor costs into the cutoff wages expressed in terms of net wages. Finally, drawing on the OECD statistics, the average real interest rate over the considered period of 1984-2001 is equal to 3.6%.[17]

We use the estimates of the structural parameters to evaluate (26)-(27) and see whether the present skill structure is efficient. In doing so, we consider two cases, namely:

1. Marginal shift from Medium to High skills (the fraction of low-skilled is constant),
2. Marginal shift from Low to Medium skills (the fraction of high-skilled is constant).

Taking the first case, the marginal increase of the fraction of high-skilled workers by educating the marginal medium-skilled worker induces an output increase of DM 2269.88. At the same time, the period private return of the investment into high skills amounts to DM 2277.60. Thus, the fraction of high-skilled workers almost precisely matches its' socially optimal level.

For the next case, however, the result is different. The output effect of the marginal change of the skill structure towards increasing the share of medium-skilled workers in

[17]Source: OECD Economic Outlook, No.77. Price base for the calculation is set to 1998, as that of the earnings data.

the economy is again positive and, although somewhat smaller in its value, amounts to DM 2057.27. But the private return of investing into medium skills lies at the level of DM 821.67, which is less than half of the social return. Thus we obtain strong evidence of underinvestment in skills at the low-to-medium level and conclude that subsidizing the education of the low-skilled must be welfare improving from the social prospective. Going back to the definition of skills this means that it would be socially optimal to reduce the fraction of workers with inadequate or general elementary training and increase the fraction of those with middle-vocational training.

Although, we are able to provide new insights of whether there is over- or underinvestment in an economy, our framework does not allow us to determine the source of the inefficiency. The detected underinvestment could either be caused by the hold-up problem that workers face when making their investment decision or by a positive human capital externality due to an education spillover.

5. CONCLUSION

This paper extends the search equilibrium model of Burdett and Mortensen (1998) by introducing different skill groups and linking them via a production function which permits any degree of homogeneity. With increasing returns to scale we are able to generate a decreasing wage offer density. Allowing for heterogeneity leads to further improvement of the shape of the wage offer and earnings distributions predicted by the model. Another important result of the extended model is that local monopsony power of firms and complementarity of skills in the production function imply that firms occupy the same position in the wage offer distribution for each skill group. This fact makes our theory consistent with the empirical findings that wages of workers of different skill groups employed at the same firm are positively correlated.

We apply our model to learn whether there is over- or underinvestment in human capital in Germany. Our results show that the private return of the investment of a low skilled worker to become medium skilled is only half of the social return of such a marginal change in the skill structure. This suggests that social returns to education exceed private returns and that a policy designed to promote education at lower levels would be welfare improving. At the same time the number of high-skilled workers is found to be close to the socially efficient level.

REFERENCES

Acemoglu, D., "A Microfoundation for Social Increasing Returns to Human Capital Accumulation", *Quarterly Journal of Economics* 111 (1996).

Acemoglu, D., and R., Schimer, "Holdups and Efficiency with Search Frictions", *International Economic Review* 40 (1999), 827-849.

Andrews, D., "Inconsistency of the Bootstrap when a Parameter is on the Boundary of the Parameter Space", *Econometrica* 68 (2000), 399-405.

Barth, E., and H., Dale Olsen, "Skill-Group Size and Wages", *New Zealand Economic Papers* 36 (2002), 83-84.

Barth, E., and H., Dale Olsen, "Assortative Matching in the Labor Market", mimeo, Institute for Social Research, Norway (2003).

Bontemps, C., J.-M., Robin, and G.J., van den Berg, "An Empirical Equilibrium Job Search Model with Search on the Job and Heterogenous Workers and Firms", *International Economic Review* 40 (1999), 1039-1075.

Bontemps, C., J.-M., Robin, and G.J., van den Berg, "Equilibrium Search with Productivity Dispersion: Theory and Estimation", *International Economic Review* 41 (2000), 305-358.

Bowlus, A., Kiefer, N., and G., Neumann, "Estimation of Equilibrium Wage Distributions with Heterogeneity", *Journal of Applied Econometrics* 10 (1995), S119-S131.

Bowlus, A., Kiefer, N., and G., Neumann, "Equilibrium Search Models and the Transition from School to Work", *International Economic Review* 42 (2001), 317-343.

Bowlus, A., and Z., Eckstein, "Discrimination and Skill Differences in an Equilibrium Search Model", *International Economic Review* 43 (2002), 1309-1346.

Burdett, K., and D.T., Mortensen, "Wage Differentials, Employer Size and Unemployment", *International Economic Review* 39 (1998), 257-273.

CHERNOZHUKOV, V., AND H., HONG, "Likelihood Estimation and Inference in a Class of Nonregular Econometric Models", *Econometrica* 72 (2004), 1445-1480.

DIAMOND, P., "A Model of Price Adjustment", *Journal of Economic Theory* 3 (1971), 156-168.

FÄRE, R., JANSSON, L., AND K., LOVELL "Modelling Scale Economies with Ray-Homothetic Production Functions", *Review of Economics and Statistics* 67 (1985), 624-629.

GROUT, P., "Investment and Wages in the Absence of Binding Contracts: A Nash Bargaining Approach", *Econometrica* 52 (1984), 449-460.

KATZ, L., AND L., SUMMERS, "Industry Rents: Evidence and Implications," *Brookings Papers on Economic Activity: Microeconomics* (1989), 209-275.

KIEFER, N.M., AND G.R., NEUMANN, "Wage Dispersion with Homogeneity: The Empirical Equilibrium Search Model", in H. Bunzel et al., eds., *Panel Data and Labor Market Analysis*, (Amsterdam: North Holland, 1993).

KIM, Y., "Translog Production Functions and Variable Returns to Scale", *Review of Economics and Statistics* 74 (1992), 546-552.

KONING, P., G., RIDDER, AND G.J., VAN DEN BERG, "Structural and Frictional Unemployment in an Equilibrium Search Model with Heterogeneous Agents", *Journal of Applied Econometrics* 10 (1995), 133-151.

MANNING, A., "Monopsony in Motion", (Princeton University Press, 2003).

MASTERS, A., "Efficiency of Investment in Human and Physical Capital in a Model of Bilateral Search and Bargaining", *International Economic Review* 39 (1998), 477-494.

MCDONALD, J., "Some Generalized Functions for the Size Distribution of Income", *Econometrica* 52 (1984), 647-663.

MORTENSEN, D.T., "Equilibrium Wage Distribution: a Synthesis", in J. Hartog et al., eds., *Panel Data and Labor Market Studies*, (Amsterdam: North Holland 1990).

MORTENSEN, D.T., "Modelling Matched Job-Worker Flows", mimeo (1999).

MORTENSEN, D.T., "Equilibrium Unemployment with Wage Posting: Burdett-Mortensen Meet Pissarides", in H., Bunzel et al., eds., *Panel Data and Structural labor Market Models* (Amsterdam: Elsevier, 2000).

MORTENSEN, D.T., AND G.R., NEUMAN, "Estimating Structural Models of Unemployment and Job Duration", in W.A., Barnett et al., eds., *Dynamic Econometric Modelling, Proceedings of the Third International Symposium in Economic Theory and Econometrics* (Cambridge: Cambridge University Press, 1988).

POSTEL-VINAY, F., AND J.-M., ROBIN, "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity", *Econometrica* 70 (2002), 2295-2350.

SAINT-PAUL, G., "Unemployment and Increasing Private Returns to Human Capital", *Jouranl of Public Economics* 61 (1996), 1-20.

SINGH, S., AND G., MADDALA, "A Function for Size Distribution of Income", *Econometrica* 44 (1976), 963-970.

VAN DEN BERG, G.J., AND G., RIDDER, "An Empirical Equilibrium Search Model of the Labor Market", *Econometrica* 66 (1998), 1183-1221.

VDR, "Rentenversicherung in Zeitreihen", DRV-Schrift, Band 22, Frankfurt am Main (2004).

ZELLNER, A., AND H., RYU, "Alternative Functional Forms for Production, Cost and Returns to Scale Functions", *Journal of Applied Econometrics* 13 (1998), 101-127.

APPENDIX

Figure A.1: “Kernel Estimates of Net Earnings Densities”

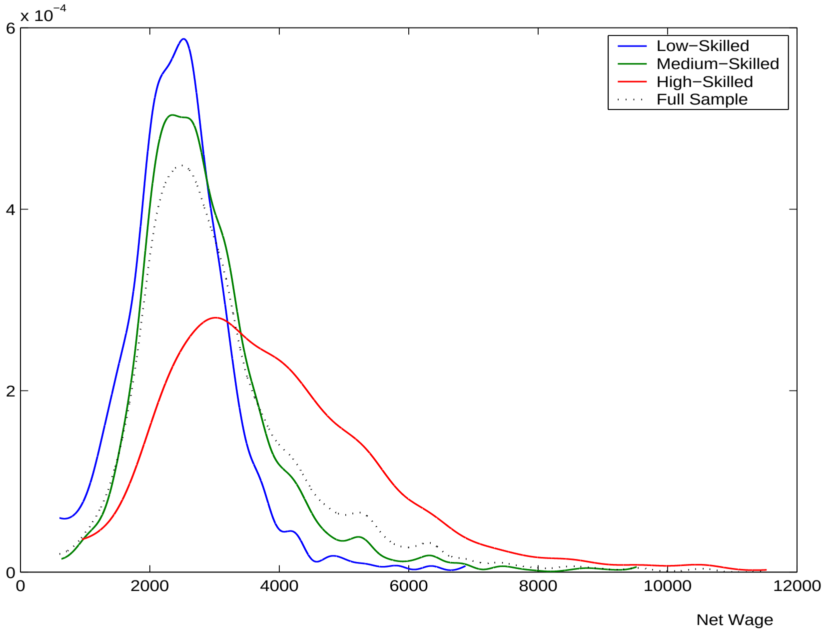


Figure A.2: “Kernel Estimates of Labour Cost Densities”

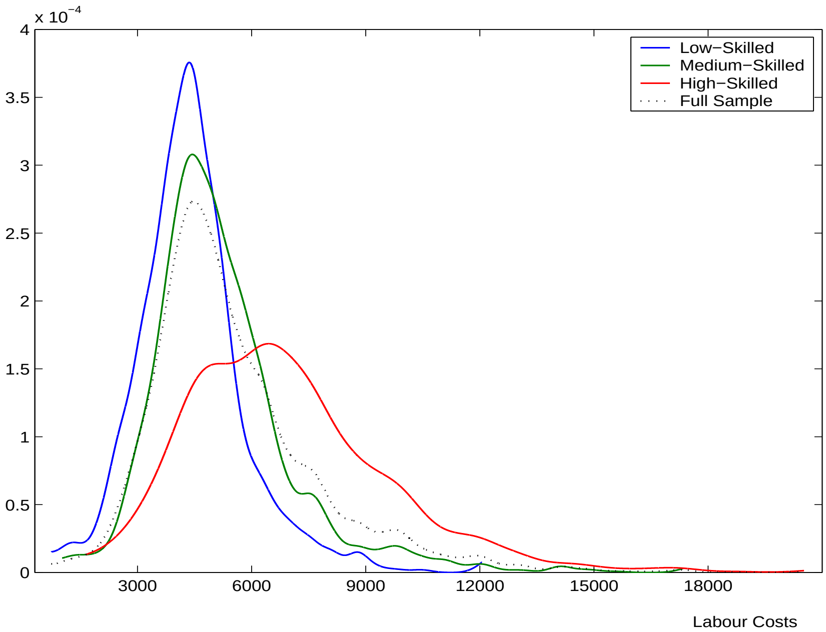


Table A.1: "Estimation Results: Homogeneous Firms"

	Specification			
	Constant Returns*		Increasing Returns	
κ_{u1}	4.6182	[4.1640, 5.0725]	5.9115	[5.2372, 6.5858]
κ_{u2}	8.2312	[7.6093, 8.8531]	10.4875	[9.5566, 11.4183]
κ_{u3}	14.1192	[12.5421, 15.6963]	17.8712	[15.4814, 20.2611]
κ_e	0.1605	[0.1421, 0.1789]	2.0963	[1.7342, 2.4585]
δ	0.0066	[0.063, 0.0068]	0.0043	[0.0041, 0.0045]
ξ			2.0000	[1.7945, 2.2053]
α_1	0.1513		0.3704	
α_2	0.5080		1.0044	
$\ln(\mathcal{L})$	-68248.06		-66758.10	

* *Here and henceforward 95% confidence intervals in square brackets*

Figure A.3: "Aggregate Wage Offer Densities: Homogeneous Firms"

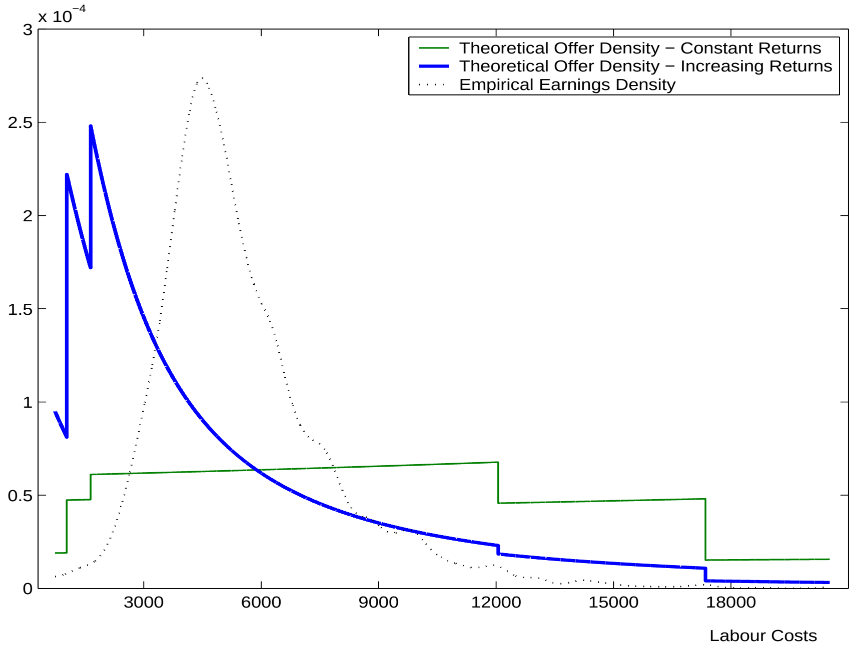


Figure A.4: "Aggregate Labour Costs Densities: Homogeneous Firms"

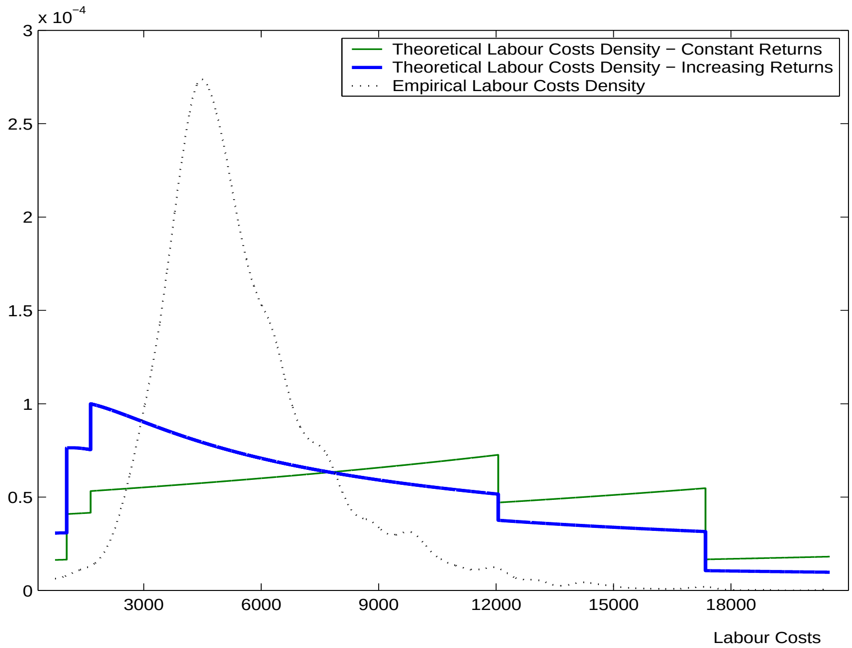


Table A.2: "Estimation Results: 3-Point Employer Heterogeneity"

	Specification				
	Constant Returns			Increasing Returns	
κ_{u1}	5.6156	[4.9973, 6.2339]	κ_{u1}	5.9612	[5.2742, 6.6481]
κ_{u2}	9.9702	[9.1169, 10.8234]	κ_{u2}	10.6176	[9.6662, 11.5691]
κ_{u3}	17.0121	[14.8258, 19.1985]	κ_{u3}	18.0656	[15.6320, 20.4991]
κ_e	2.1277	[1.9869, 2.2684]	κ_e	3.6432	[3.3926, 3.8939]
δ	0.0047	[0.0045, 0.0049]	δ	0.0042	[0.0040, 0.0044]
			ξ_1	1.0381	[1.0324, 1.0437]
			ξ_2	1.3961	[1.2977, 1.4945]
			ξ_3	4.9201	[3.1342, 6.7060]

$\{\alpha_{ij}\}$	$j=1$	$j=2$	$j=3$	$\{\alpha_{ij}\}$	$j=1$	$j=2$	$j=3$
$i=1$	0.1772	0.1449	0.1499	$i=1$	0.1896	0.2466	0.9822
$i=2$	0.4622	0.4939	0.5212	$i=2$	0.4850	0.6586	2.4929

$\{\overline{w}_{ij}\}$	$j=1$	$j=2$	$\{\overline{w}_{ij}\}$	$j=1$	$j=2$
$i=1$	4431	5698	$i=1$	4431	5698
$i=2$	5065	7597	$i=2$	5065	6964
$i=3$	6964	9992	$i=3$	6964	9992

	$j=1$	$j=2$		$j=1$	$j=2$
γ_j	0.7905	0.9610	γ_j	0.8485	0.9685

$\ln(\mathcal{L})$	-65059.96	$\ln(\mathcal{L})$	-64843.50

Figure A.5: "Aggregate Wage Offer Densities: 3-Point Employer Heterogeneity"

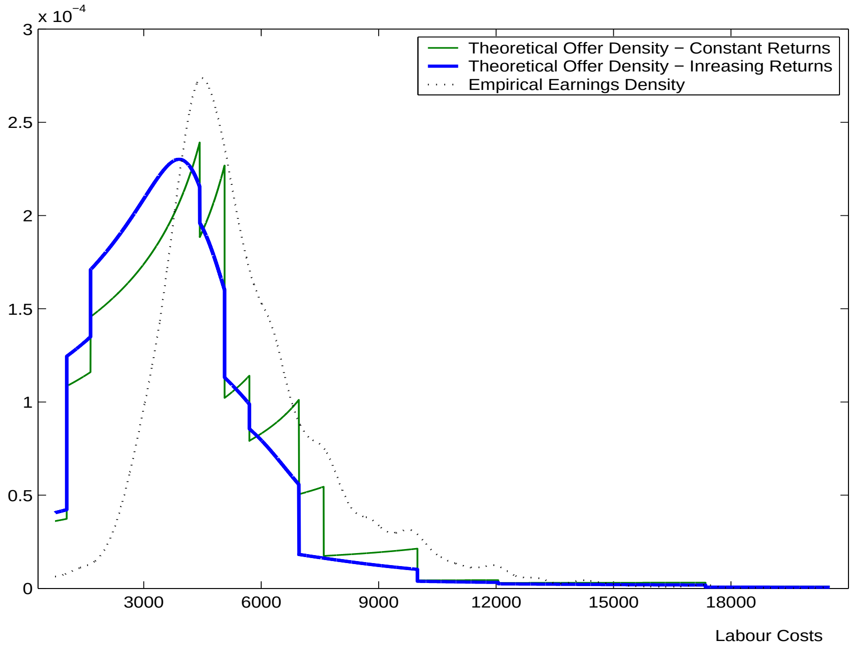


Figure A.6: "Aggregate Labour Costs Densities: 3-Point Employer Heterogeneity"

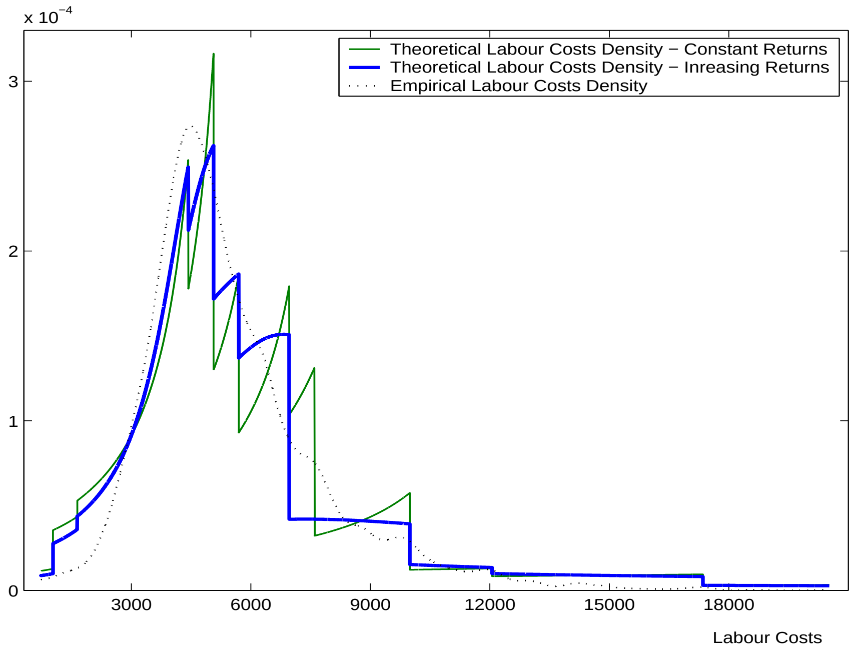


Figure A.7: “3-Point Employer Heterogeneity: Skill-Specific Theoretical Offer Densities”

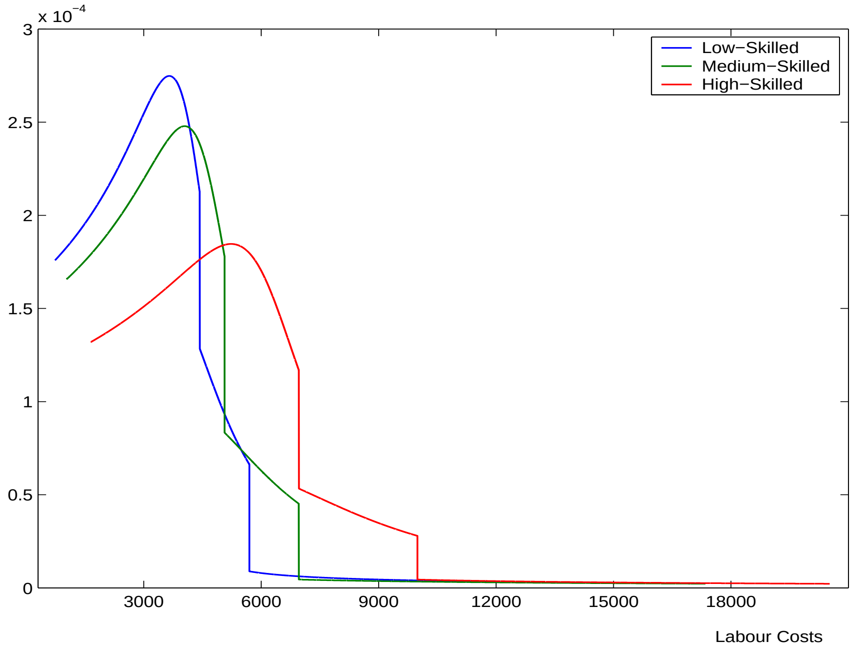
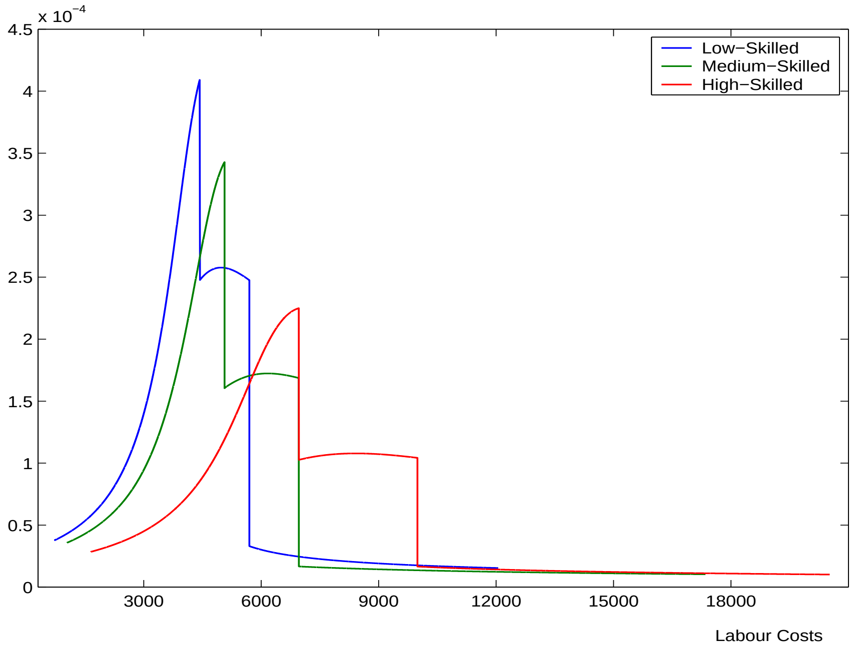


Figure A.8: “3-Point Employer Heterogeneity: Skill-Specific Theoretical Labour Costs Densities”



Proof of Proposition 5.

Define

$$\begin{aligned} h_j(w) &= \frac{[\delta+\lambda_e(1-\gamma_{j-1})]^2}{\left[\delta+\lambda_e\overline{F}_j(w)\right]^2},\ r_{ij} = \frac{\delta\lambda_i(\delta+\lambda_e)}{(\delta+\lambda_i)[\delta+\lambda_e(1-\gamma_{j-1})]^2}q_i \\ Y_j'(\mathbf{r}_j) &= \frac{\partial Y_j(\mathbf{r}_j)}{\partial l_i},\ \text{and } \sigma_{ij} = \sum\nolimits_l \frac{\partial^2 Y_j(\mathbf{r}_j)}{\partial l_i \partial l_l} r_{lj} r_{ij}. \end{aligned}$$

The second order Taylor expansion of the production function around r_j is given by

$$Y_j(\mathbf{l}(\mathbf{w}_j)) = Y_j(\mathbf{r}_j) + \sum\nolimits_i Y_j'(\mathbf{r}_j)[r_{ij}h_j(w) - r_{ij}] + \frac{1}{2}\sum\nolimits_i \sigma_{ij}[h_j(w)-1]^2.$$

Note, that $h_j(w)$ is independent of the skill group i, because of equation (12). Using the equal profit condition for the equilibrium, i.e. $\pi_j(\mathbf{w}_j) = \pi_j\left(\underline{\mathbf{w}}_j\right)$, and substituting gives

$$\begin{aligned} D = \sum\nolimits_i \left(Y_j'(\mathbf{r}_j) - w_i\right) r_{ij}h_j(w) + \frac{1}{2}\sum\nolimits_i \sigma_{ij}(h_j(w)-1)^2 \\ -\sum\nolimits_i \left(Y_j'(\mathbf{r}_j) - \underline{w}_{ij}\right) r_{ij} = 0 \end{aligned} \tag{A.1}$$

The first order condition for wage w_i satisfies,

$$\left(\frac{\partial Y_j(\mathbf{l}(\mathbf{w}))}{\partial l_i(w_i)} - w_i\right) l_i(w_i) = l_i(w_i)^2 \left[\frac{dl_i(w_i)}{dw_i}\right]^{-1}, \tag{A.2}$$

where rhs can be written as

$$l_i(w_i)^2 \left[\frac{dl_i(w_i)}{dw_i}\right]^{-1} = [r_{ij}h_j(w)]^2 \left[r_{ij}\frac{dh_j(w)}{dw_i}\right]^{-1}$$

According to the result that all firms occupy the same position in all wage offer distribution, changing the wage for one skill group implies a change of all other wages in the same direction, i.e. according to equation (A.1)

$$\begin{aligned} &[r_{ij}h_j(w)]^2 \left[r_{ij}\frac{dh_j(w)}{dw_i}\right]^{-1} = r_{ij}h_j(w)^2 \left(\frac{-\partial D/\partial h_j(w)}{-\sum_i \partial D/\partial w_i}\right) \\ &= \frac{r_{ij}}{\sum_i r_{ij}} \left(\sum\nolimits_i \left(Y_j'(\mathbf{r}_j) - w_i\right) r_{ij}h_j(w) + \sum\nolimits_i \sigma_{ij}\left(h_j(w)^2 - h_j(w)\right)\right). \end{aligned}$$

Using a Taylor-Expansion for the first derivative of the production function and substituting $l_l(w_l)$ out gives

$$Y_j'(\mathbf{l}(\mathbf{w})) = Y_j'(\mathbf{r}_j) + \sum\nolimits_l \frac{\partial^2 Y_j(\mathbf{r}_j)}{\partial l_i \partial l_l}(r_{lj}h_j(w) - r_{lj}).$$

The first order condition can therefore be written as

$$\begin{aligned} & \left(Y_j'(\mathbf{r}_j) - w_i\right) r_{ij} h_j(w) + \sigma_{ij}\left(h_j(w)^2 - h_j(w)\right) \\ = \; & \frac{r_{ij}}{\sum_i r_{ij}} \left(\sum\nolimits_i \left(Y_j'(\mathbf{r}_j) - w_i\right) r_{ij} h_j(w) + \sum\nolimits_i \sigma_{ij}\left(h_j(w)^2 - h_j(w)\right)\right). \end{aligned}$$

Substituting $\sum_i \left(Y_j'(\mathbf{r}_j) - w_i\right) r_{ij} h_j(w)$ from equation (A.1) gives

$$\begin{aligned} & \left(Y_j'(\mathbf{r}_j) - w_i\right) r_{ij} h_j(w) + \sigma_{ij}\left(h_j(w)^2 - h_j(w)\right) \\ = \; & \frac{r_{ij}}{\sum_i r_{ij}} \sum\nolimits_i \left(Y_j'(\mathbf{r}_j) - \underline{w}_{ij}\right) r_{ij} + \frac{r_{ij}}{\sum_i r_{ij}} \frac{1}{2} \sum\nolimits_i \sigma_{ij}\left[h_j(w)^2 - 1\right]. \end{aligned}$$

Evaluating this equation at $\underline{w}_{ij}$ and substituting $\frac{r_{ij}}{\sum_i r_{ij}} \sum_i \left(Y_j'(\mathbf{r}_j) - \underline{w}_{ij}\right) r_{ij}$ gives

$$\begin{aligned} & \left(Y_j'(\mathbf{r}_j) - w_i\right) r_{ij} h_j(w) + \sigma_{ij}\left(h_j(w)^2 - h_j(w)\right) \\ = \; & \left(Y_j'(\mathbf{r}_j) - \underline{w}_{ij}\right) r_{ij} + \frac{r_{ij}}{\sum_i r_{ij}} \frac{1}{2} \sum\nolimits_i \sigma_{ij}\left[h_j(w)^2 - 1\right]. \end{aligned}$$

Rearranging gives

$$\left(\sigma_{ij} - \mu_{ij}\right) h_j(w)^2 + \left(\left(Y_j'(\mathbf{r}_j) - w_i\right) r_{ij} - \sigma_{ij}\right) h_j(w) = \left(Y_j'(\mathbf{r}_j) - \underline{w}_{ij}\right) r_{ij} - \mu_{ij}, \quad \text{(A.3)}$$

where $\mu_{ij} = \frac{r_{ij}}{\sum_i r_{ij}} \frac{1}{2} \sum_i \sigma_{ij}$.

For a production function with homogeneity of degree one $\sigma_{ij} = 0$ for all i. So we get

$$F_{ij}(w_i) = \frac{\delta + \lambda_e}{\lambda_e} - \frac{\delta + \lambda_e(1 - \gamma_{j-1})}{\lambda_e} \sqrt{\frac{Y_j'(\mathbf{r}_j) - w_i}{Y_j'(\mathbf{r}_j) - \underline{w}_{ij}}}.$$

Apart from this, a special case appears if $\left(Y_j'(\mathbf{r}_j) - \underline{w}_{ij}\right) r_{ij} - \mu_{ij} = 0$. In this case we get

$$F_{ij}(w_i) = \frac{\delta + \lambda_e}{\lambda_e} - \frac{\delta + \lambda_e(1 - \gamma_{j-1})}{\lambda_e} \sqrt{\frac{\left(Y_j'(\mathbf{r}_j) - \underline{w}_{ij}\right) r_{ij} - \sigma_{ij}}{\left(Y_j'(\mathbf{r}_j) - w\right) r_{ij} - \sigma_{ij}}}.$$

This solution, however, implies artificial restrictions on ξ_j, so its consideration is neither interesting nor useful.

Otherwise, we get the following solution for the quadratic function

$$\begin{aligned} h_j(w) = & -\frac{\left(Y_j'(\mathbf{r}_j) - w_i\right) r_{ij} - \sigma_{ij}}{2\left(\sigma_{ij} - \mu_{ij}\right)} \\ & \pm \frac{\sqrt{\left(\left(Y_j'(\mathbf{r}_j) - w_i\right) r_{ij} - \sigma_{ij}\right)^2 + 4\left(\sigma_{ij} - \mu_{ij}\right)\left(\left(Y_j'(\mathbf{r}_j) - \underline{w}_{ij}\right) r_{ij} - \mu_{ij}\right)}}{2\left(\sigma_{ij} - \mu_{ij}\right)}. \quad \text{(A.4)} \end{aligned}$$

The wage offer density implied by the quadratic function (A.3) has to be positive, i.e.

$$\frac{dF_{ij}(w)}{dw_i} = -\frac{-r_{ij}h_j(w)}{\left(2(\sigma_{ij}-\mu_{ij})h_j(w) + \left(\left(Y_j'(\mathbf{r}_j) - w_i\right) r_{ij} - \sigma_{ij}\right)\right) \frac{\partial h_j(w)}{\partial F_{ij}(w)}} > 0.$$

Since $\frac{\partial h_j(w)}{\partial F_{ij}(w)} > 0$, it follows that $2(\sigma_{ij}-\mu_{ij})h_j(w) + \left(\left(Y_j'(\mathbf{r}_j) - w_i\right) r_{ij} - \sigma_{ij}\right)$ has to be greater than zero. Rewriting equation (A.4) implies that only the positive solution is valid, i.e.

$$\begin{aligned} &+\sqrt{\left(\left(Y_j'(\mathbf{r}_j) - w_i\right) r_{ij} - \sigma_{ij}\right)^2 + 4(\sigma_{ij}-\mu_{ij})\left(\left(Y_j'(\mathbf{r}_j) - \underline{w}_{ij}\right) r_{ij} - \mu_{ij}\right)} \\ &= 2(\sigma_{ij}-\mu_{ij})h_j(w) + \left(Y_j'(\mathbf{r}_j) - w_i\right) r_{ij} - \sigma_{ij} > 0. \end{aligned} \tag{A.5}$$

Hence the cumulative wage offer distribution is given by

$$F_{ij}(w_i) = \frac{\delta+\lambda_e}{\lambda_e} - \frac{\delta+\lambda_e(1-\gamma_{j-1})}{\lambda_e\sqrt{\frac{\left(Y_j'(\mathbf{r}_j)-w_i\right)r_{ij}-\sigma_{ij}-\sqrt{\left(\left(Y_j'(\mathbf{r}_j)-w_i\right)r_{ij}-\sigma_{ij}\right)^2+4(\sigma_{ij}-\mu_{ij})\left(\left(Y_j'(\mathbf{r}_j)-\underline{w}_{ij}\right)r_{ij}-\mu_{ij}\right)}}{-2(\sigma_{ij}-\mu_{ij})}}}.$$

In order to see that the wage offer density can be increasing and decreasing consider the explicit solution to the wage offer density

$$\begin{aligned} f_{ij}(w_i) &= \frac{(\delta+\lambda_e(1-\gamma_{j-1}))r_{ij}}{2\lambda_e\sqrt{\left(\left(Y_j'(\mathbf{r}_j)-w_i\right)r_{ij}-\sigma_{ij}\right)^2+4(\sigma_{ij}-\mu_{ij})\left(\left(Y_j'(\mathbf{r}_j)-\underline{w}_{ij}\right)r_{ij}-\mu_{ij}\right)}} \\ &\quad\times\frac{1}{\sqrt{\frac{\left(Y_j'(\mathbf{r}_j)-w_i\right)r_{ij}-\sigma_{ij}-\sqrt{\left(\left(Y_j'(\mathbf{r}_j)-w_i\right)r_{ij}-\sigma_{ij}\right)^2+4(\sigma_{ij}-\mu_{ij})\left(\left(Y_j'(\mathbf{r}_j)-\underline{w}_{ij}\right)r_{ij}-\mu_{ij}\right)}}{-2(\sigma_{ij}-\mu_{ij})}}}. \end{aligned}$$

The slope of the wage offer density is given by

$$\begin{aligned} \frac{\partial f_{ij}(w)}{\partial w} &= -\frac{\left(\left(Y_j'(\mathbf{r}_j)-w_i\right)r_{ij}-\sigma_{ij}\right)^2+4(\sigma_{ij}-\mu_{ij})\left(\left(Y_j'(\mathbf{r}_j)-\underline{w}_{ij}\right)r_{ij}-\mu_{ij}\right)-2r_{ij}\left(\left(Y_j'(\mathbf{r}_j)-w_i\right)r_{ij}-\sigma_{ij}\right)}{\left(\left(Y_j'(\mathbf{r}_j)-w_i\right)r_{ij}-\sigma_{ij}\right)^2+4(\sigma_{ij}-\mu_{ij})\left(\left(Y_j'(\mathbf{r}_j)-\underline{w}_{ij}\right)r_{ij}-\mu_{ij}\right)} \\ &\quad\times\frac{\frac{(\delta+\lambda_e(1-\gamma_{j-1}))r_{ij}^2}{4\lambda_e\sqrt{\left(\left(Y_j'(\mathbf{r}_j)-w_i\right)r_{ij}-\sigma_{ij}\right)^2+4(\sigma_{ij}-\mu_{ij})\left(\left(Y_j'(\mathbf{r}_j)-\underline{w}_{ij}\right)r_{ij}-\mu_{ij}\right)}}}{\sqrt{\frac{\left(Y_j'(\mathbf{r}_j)-w_i\right)r_{ij}-\sigma_{ij}-\sqrt{\left(\left(Y_j'(\mathbf{r}_j)-w_i\right)r_{ij}-\sigma_{ij}\right)^2+4(\sigma_{ij}-\mu_{ij})\left(\left(Y_j'(\mathbf{r}_j)-\underline{w}_{ij}\right)r_{ij}-\mu_{ij}\right)}}{-2(\sigma_{ij}-\mu_{ij})}}}. \end{aligned}$$

Thus, a necessary condition for the wage offer density to be upward sloping is that $\left(Y_j'(\mathbf{r}_j) - w_i\right) r_{ij} - \sigma_{ij} > 0$. Substituting σ_{ij}, and using the Euler Theorem gives the stated condition. ∎