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

August 25, 2005

Abstract

We introduce skill groups and different production technologies into the Burdett-Mortensen model of on the job search. Supermodularity of the different skill groups in the production process leads to a positive intrafirm wage correlation between skill groups. Increasing returns to scale allow the theoretical earnings density to be unimodal with a long right tail even in the absence of productivity dispersion. We perform the structural estimation the model and evaluate the effect that arises from the marginal shift of the skill structure towards larger fraction of high-skilled workers. Our estimates of the production parameters demonstrate economy-wide increasing returns to scale. Furthermore we support the existence of positive welfare effect from increasing the share of high-skilled agents in the workforce.

Keywords: Search, wage correlation, social returns to education.

JEL Classification: J21, J23, J64.

*Authors wish to thank Aico van Vuuren and David Margolis for their useful comments. 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 technology employed by the economy and the search and matching frictions that govern the allocation of workers to jobs. The aim of the paper is to provide a theoretical and still empirically tractable model that takes all three factors and its interactions into account. For doing so we extend the search equilibrium model of Burdett and Mortensen (1998) and derive an explicit functional form for the wage offer and earnings distributions. Our extension explicitly introduces different skill groups that are linked via a production function which permits either constant or increasing returns to scale. The extension to different skill groups allows for the analysis of firms' wage posting behavior, where firms simultaneously compete for workers of different skill groups. As we show this results in a positive correlation between the wages of workers in different skill groups within firms.

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 realistic-shaped wage distribution. 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 right-tailed wage earnings density, depending on the assumed productivity dispersion. Postel-Vinay and Robin (2002) extend this for both employer and worker heterogeneity.

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 offer and earnings densities is true even for identical employers. While the structural models with continuous productivity dispersion as suggested by Bontemps et al. (2000) and Postel-Vinay and Robin (2002) improve the fit to the empirical wage earnings distribution and provide reliable estimates of the labour 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 introduced explicitly. As a result this allows us to estimate the parameters of the production functions even without using firms' data.

In the theoretical part of the paper we demonstrate that whenever skills are complementary in the production process we should observe a positive within-firm correlation between wages of workers with different skills. Positive intrafirm wage correlation is a well established fact, empirical evidence of which are presented in Katz and Summers (1989) and Barth and Dale-Olsen (2003) among many others. Theoretical consideration of the issue is performed by Kremer (1993). In his O-ring theory Kremer (1993) also uses a production function that exhibits complementarity of the working colleagues' abilities not to make a mistake when performing a sequence of tasks in order to complete the final good. One important consequence of the O-ring theory is a positive correlation between wages and the number of tasks and therefore the overall size of the workforce. However, recently Barth and Dale-Olsen (2002) have empirically demonstrated that the employer-size wage effect vanishes once we look at the skill-group size. In view of this result the labour market frictions approach of this paper that predicts a positive correlation between skill-group size and wages may be more favorable then the O-ring theory of Kremer (1993).

We use the estimated parameters of our model to analyze whether there is over- or underinvestment in human capital from a social welfare point of view, i.e. whether the increase in output coming from educating the marginal individual pays off the individual's and the government's investment costs. Underinvestment in (undirected) search or matching models are analyzed by Acemoglu (1996) and Masters (1998). Following Grout (1984) they provide 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. The same mechanism is at work in the present paper. However, underinvestment cannot be attributed to rent sharing solely, in addition it has to be the case that workers of (potentially) different skill have to search in the same market. Allowing for segmented labor markets, where unskilled workers do not search for the same jobs as skilled workers do (and vise versa), makes both over- or underinvestment into education possible. The simple idea is that a lower unemployment rate among high skilled workers can increase the return to human capital investment as shown by Saint-Paul (1996).[1] Given these results in the literature we do not endogenize the matching probabilities in order to show that over- or undereducation can exist. Instead, we assume constant offer arrival rates and investigate empirically whether over- or underinvestment into skills is

[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.

present in the German economy. We find that a marginal change in the skill structure of the labor force towards more high skilled workers does indeed generate an increase in output sufficient to overcompensate the society for the additional cost of education to the marginal individual.

Estimation methodology applied in this paper is based on the one considered in Bowlus et al. (1995), (2001). However, skill-multiplicity and Cobb-Douglas production function used in the econometric model impose additional restrictions that must be taken into account when suiting the original method. First, these are the restrictions that allow representing the subset of production parameters as a function of search frictions parameters and the homogeneity degree of the Cobb-Douglas technology. Second, these are the identifiability restrictions that appear with an introduction of employer heterogeneity. Our estimation problem can be also related to that of 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 discrimination of employers. However, unlike in Bowlus and Eckstein (2002), we estimate the parameters of interest by maximum likelihood.

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 formulate 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 the underinvestment into 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 that link the skill groups via the production function.

2.1 *Framework*

The model has an infinite horizon, is set in continuous time and concentrates on steady states. Workers are assumed to be risk neutral and to discount at rate r. Each worker belongs to a skill group $i = 1, 2, ..., I$ whose measures are defined 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. Before choosing a skill-group workers incur a one-off cost c_i for skill-specific education. By assuming perfect capital market workers are able to borrow the cost of education.

Workers search for a job in the skill-segmented labor markets. With probability λ_i unemployed workers of skill group i encounter a firm that makes them a wage offer corresponding to their education, and with probability λ_e employed workers encounter a firm.[2] Then workers decide whether to accept or reject the job offer. 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, 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. has increasing differences in $\mathbf{l}(\mathbf{w} \mid \mathbf{w^r}, F(\mathbf{w}))$ as defined below, and is twice continuously differentiable in $l_i(w \mid w_i^r, F_i(w))$.

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 $l \vee l' \equiv (\max(l_1, l_1'), ..., \max(l_I, l_I'))$ and $l \wedge l' \equiv (\min(l_1, l_1'), ..., \min(l_I, l_I'))$. Supermodularity in l_i implies *increasing differences* in l_i, i.e. for $\mathbf{l} \geq \mathbf{l}'$ it follows that

$$Y_j(l_i, \mathbf{l}_{-i}) + Y_j\left(l_i', \mathbf{l}_{-i}'\right) \geq Y_j\left(l_i, \mathbf{l}_{-i}'\right) + Y_j(l_i', \mathbf{l}_{-i}),$$

where $-i$ denotes the vector of all skill groups except i.

Firms maximize profits by offering a wage schedule $\mathbf{w} = (w_1, w_2, ..., w_I) = (w_i, \mathbf{w}_{-i})$.

2.2 *Workers' Search Strategy*

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. $U_i = V_i(w_i^r)$, where 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. Flow values of being unemployed and

[2] λ_e is not skill group specific, since we would otherwise not be able to derive an explicit wage offer distribution function.

employed

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

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

respectively, can be solved for a reservation wage[3]

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

In order to keep the analysis simple, for the remainder of the paper we assume that $r/\lambda_i \to 0$ as done in the original model by Burdett and Mortensen (1998). 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 assumptions of constant Poisson arrival rates λ_i, λ_e and the constant separation rate δ 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 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 split at rate δ or a worker receives 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

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

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 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*

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. $F\left(\mathbf{w}\right)$.

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

The expectation operator in the equation above is over all possible realizations of the different skill group sizes $l_i\left(w \mid w_i^r, F_i\left(w\right)\right)$ a firm can realize given its choice of the wage schedule and the birth-death process characterized above. Hence, in the steady state a firm might 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. This implies that we can write the maximization problem of a type j firm as

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

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 $F_j(\mathbf{w}) = (F_{1j}(w), F_{2j}(w), ..., F_{Ij}(w))$, 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 $F_j(\mathbf{w})$ with $\mathbf{w} \in \mathbf{W}_j$ for each firm type $j \in J$ such that

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

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.

Given 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 implies 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 $\mathbf{w}' \geq \mathbf{w}^{\mathbf{r}}$, $\pi_j(w_i, \mathbf{w}_{-i})$ is supermodular, i.e.

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

because the same inequality holds for output $Y_j(E[\mathbf{l}(w_i, \mathbf{w}_{-i})])$ 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.

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

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

This positive correlation between the wages of workers in different skill groups within firms is a well established fact. Katz and Summers (1989) show evidence that secretaries earn more in firms where average wages are higher. More recently, Barth and Dale-Olsen (2003) find that "[h]igh-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, labor market frictions lead to an upward sloping labor supply curve for each skill group which can be seen from equation (5). Secondly, we need the complementarity of the skill groups in the production process. This 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.

Note, that Proposition 1 does not guarantee that a firm occupies the same position in the wage offer distribution of all skill groups, because it is possible that there is a mass point in the wage offer distribution of skill group i but not in the wage offer distribution in the other $-i$ skill groups.

Given that the skill group size is increasing in the wage w_i, it would be a waist of money, 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 connected and closed from below, 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, compare equation (7). 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

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

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 $Y_j(E\left[\mathbf{l}\left(\mathbf{w}_j\right)\right]) - Y_{j-1}(E\left[\mathbf{l}\left(\mathbf{w}_j\right)\right]) \geq Y_j(E\left[\mathbf{l}\left(\mathbf{w}_{j-1}\right)\right]) - Y_{j-1}(E\left[\mathbf{l}\left(\mathbf{w}_{j-1}\right)\right])$. Since $\mathbf{l}\left(\mathbf{w}\right)$ is an increasing function of wages $\mathbf{w}$, the claim follows. ■

In order 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.

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

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 $F_i\left(w_i\right)$ of type j firms for skill group i is continuous, if*

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

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* $\overline{w}_{ij}$, *then mass points can be ruled out, i.e.*

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

Proof. Suppose a mass point exists at $w_i \in \left[\underline{w}_{ij}, \overline{w}_{ij}\right]$. Equation (6), and the fact that the cdf $F_i(w_i)$ is right continuous implies

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

since $F_i(w_i) - F_i(w_i^-) = \upsilon_i(w_i) > 0$. If the above inequality holds, when a mass point exists at w_i.
To show that mass points can only exist at the upper bound of the support of $F_i(w_i)$ note that equation (5) together with Proposition 2 implies that $E\left[l_i\left(w_i \mid w_i^r, F_i(w_i)\right)\right]$ is strictly increasing in w_i on its support $\left[\underline{w}_{ij}, \overline{w}_{ij}\right]$, i.e. $\Delta E\left[l_i\left(w_i \mid w_i^r, F_i(w_i)\right)\right]/\Delta w_i > 0$. Using the equal profit condition (7) implies

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

where $E\left[l_i\left(w_i^-\right)\right] = E\left[l_i\left(w_i \mid w_i^r, F_i\left(w_i^-\right)\right)\right]$. This expression is only 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

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

A necessary condition for no mass point to exist can be derived by letting $\upsilon_i(\overline{w}_{ij}) \rightarrow 0$, i.e.

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

■

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 is higher than the increase in total wage cost induced by a slight wage increase. 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 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 1 (i.e. firm $k = 1$ offers $\underline{w}_{1j}$), then Proposition 1 implies that for all $\mathbf{w} \in \mathbf{W}_j$

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

In order to be able to us the above property we introduce the following separation of a skill group size, where we rewrite the skill group size as

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

where

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

The fact that $h_j(w)$ depends only on the position the firm takes in the wage offer distribution, i.e. $F_j(w)$, implies that $h_j(w)$ does not depend on any skill specific parameter. Since we want to derive an explicit functional form for the wage offer distribution for each skill group i we additionally have to 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 $Y_j(\mathbf{r}_j)$ is homogeneous of degree ξ_j the Taylor Expansion is given by

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

where

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

Using the results of Propositions 1-3 we invoke the equal profit condition $\pi_j = \pi_j^r$ and apply the Taylor Expansion and 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 *Given that production functions $Y_j(E[\mathbf{l}(\mathbf{w})])$ $\forall j = 1, 2, ..., J$ are supermodular and given 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 that has the following form*
(i) for $\xi_j = 1$

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

(ii) for $\xi_j \neq 1$

$$F_{ij}(w_i) = \frac{\delta + \lambda_e}{\lambda_e} \tag{14}$$

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

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

$$\mu_{ij} = \frac{r_{ij}}{\sum_i r_{ij}}\frac{1}{2}\sum\nolimits_i \sigma_{ij},$$

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

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

Proof. See Appendix. ■

The aggregate wage offer distribution is given by

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

A special case for $F_{ij}(w_i)$ when $\left(Y_j'(\mathbf{r}_j) - \underline{w}_{ij}\right) r_{ij} = \mu_{ij}$ 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.

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. Since an upward-sloping earnings density is at odds with the empirical observation of a flat 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, depending on the assumed productivity dispersion.[4]

The novelty 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. A production technology with decreasing returns to scale would result in a negative wage offer density for at least one skill group, hence violate the first order condition and result in a violation of the continuity condition.

The reason for why increasing returns to scale can bend the wage offer density in such a way that is depicts a long right tail has its cause in the equal profit condition. Let us focus on the case with a homogenous production function with increasing returns to scale

[4] 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).

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. First note that the skill group size is determined solely by the firm's position in the wage offer distribution. Thus, the shape of the wage offer distribution does not matter for the output generated. Due to increasing returns to scale the output of firms at the top of the wage offer distribution increases more than compared to an economy with constant returns to scale. In order for firms on the top of the wage offer distribution to make the same profits as firms at the lower end, the firms in an environment with increasing returns to scale have to pay higher wage in order to satisfy the equal profit condition as compared to firms in an environment with constant returns to scale who are at the same position of the wage offer distribution (except of course the firm offering the reservation wage schedule). Thus, the wage offer distribution in an economy with increasing returns to scale is more dispersed. If the returns to scale are large enough, the wage difference paid by "neighboring" firms at the upper end of the wage offer distribution increases generating a decreasing wage offer density.

Mortensen (2000) makes implicitly a similar restriction 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 on increasing economies of scale due to the capital k employed by the firm, i.e. $Y(l(w)) = k^{\alpha} l(w)$. By simulation he shows that for positive α the wage offer distribution has a flat right tail.

Decreasing tail of the offer density implies the same for the earnings density. Consider the the latter in more detail. 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 at the highest possible rate. However letting $\{\underline{w}_{iJ}, \overline{w}_{iJ}\}$ go to infinity we get the following result.

Proposition 6 *Let $\underline{w}_{iJ} \rightarrow \infty$ and $\overline{w}_{iJ} \rightarrow \infty$. 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 the rate faster then 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) one obtains the closed form solution for the equilibrium earnings density

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

Define

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

$$B(w_i) \equiv \left(\left(Y_J'\left(\mathbf{r}_J\right)-w_i\right)r_{iJ}-\sigma_{iJ}\right)^2+4\left(\sigma_{iJ}-\mu_{iJ}\right)\left(\left(Y_J'\left(\mathbf{r}_J\right)-\underline{w}_{iJ}\right)r_{iJ}-\mu_{iJ}\right).$$

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

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

For $\underline{w}_{iJ} \to \infty$ and $\overline{w}_{iJ} \to \infty$ $A(w_i) = O(1)$ and $B(w_i) = O\left(w_i^{2(\xi_J-1)}\right)$, which leads to

$$g_{iJ}'(w_i) = O\left(w_i^{-2(\xi_J-1)}\right).$$

Finally, under the sufficient condition for the decreasing right tail of the $f_{iJ}(w_i)$ we get $g_{iJ}'(w_i) = O\left(w_i^{-2-\delta}\right)$, where $\delta > 0$. ■

The result of Proposition 6 tells us that the equilibrium earnings density of Proposition 5 encompasses the family of Pareto and Singh-Maddala densities, right tail of which is acknowledged to have the best fit to the observed high-earnings data (see Singh and Maddala, 1976). Similarly to the equilibrium densities of Bontemps et al. (2000), tail behaviour of $g_{iJ}(w_i)$ excludes the distributions with the exponential speed of convergence (e.g. lognormal) form the set of possible functional form candidates for the equilibrium earnings distribution. Furthermore, increasing returns of the production function extend the result of Proposition 8 in Bontemps et al. (2000) allowing earnings density to converge both slower and faster then w^{-3}.

Finally, the comparative statics results of the original Burdett-Mortensen model are still valid for the general wage offer distribution function. 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 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.

3. ECONOMETRIC MODEL

Here we consider in detail the structural econometric model based on the theory presented above. We assume a Cobb-Douglas production technology which allows for constant and increasing returns to scale, i.e.

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

with $\sum_i \alpha_{ij} = \xi_j \geq 1$, $\alpha_{ij} > 0$.

In general, we build upon the model developed by Bowlus et al. (1995), (2001). In the discussion to follow we put special emphasis on such new features as parameter identification and related modification of the estimation procedure.

3.1 *The Likelihood Function*

Let us start from the formulation of the likelihood function. For Poisson process with rate θ the joint distribution of the elapsed (t_e) and residual (t_r) duration of time spent by an individual in a certain state of the labour market is $f(t_e, t_r) = \theta^2 e^{-\theta(t_e+t_r)}$. For an individual that belongs to i-th skill group the appropriate Poisson rates are λ_i if the person is unemployed and $\delta + \lambda_e [1 - F_i(w)]$ if the person is employed at wage w. Furthermore:

- *For Unemployed:* 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 $f_i(w)$.
- *For Employed:* Equilibrium probability of sampling an agent who belongs to i-th skill group and earns wage w is $m^{-1} q_i g_i(w) \lambda_i / (\delta + \lambda_i)$. 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 $\rho_{j\to u} = \delta / \left(\delta + \lambda_e \overline{F}_i(w)\right)$ and $\rho_{j\to j} = \lambda_e \overline{F}_i(w) / \left(\delta + \lambda_e \overline{F}_i(w)\right)$ respectively.

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

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

$$\mathcal{L}_{(i)\,e} = g_i(w)\frac{q_i}{m}\frac{\kappa_i}{1+\kappa_i}\left[\delta\left(1+\kappa_e\overline{F}_i\,(w)\right)\right]^{1-d_l}e^{-\delta\left(1+\kappa_e\overline{F}_i(w)\right)[t_e+t_r]}$$

$$\times\left[\left[\delta\kappa_e\overline{F}_i\,(w)\right]^{d_t}\delta^{1-d_t}\right]^{1-d_r}. \tag{18}$$

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 $g_i\,(w)$, $f_i(w)$ and $F_i(w)$ into (17) and (18) completes the formulation of the likelihood function, where $g_i(w)$ is obtained from $F_i(w)$ using (4).

Notice that except of probability terms $m^{-1}q_i/(1+\kappa_i)$ and $m^{-1}q_i\kappa_i/(1+\kappa_i)$ (17) and (18) are the same as in Kiefer and Neumann (1993) or Bowlus et al. (1995). The main differences are rather 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 $I-1$ spike at supremum wages of each skill group.

Under employer homogeneity the assumed production function modifies to $Y(l(\mathbf{w})) = p\prod_{l=1}^{I} l_l(w)^{\alpha_l}$. Functional form of the wage offer distribution with homogeneous employers is $F(w) = \sum_{i=1}^{I}\frac{q_i}{m}F_i(w)$, where $F_i(w)$ is given in Proposition 5 with $J=1$. Rewritten in terms of $\kappa_{i,e}$ the skill-specific offer distribution becomes

$$F_i(w_i) = \frac{1+\kappa_e}{\kappa_e} - \frac{1+\kappa_e}{\kappa_e\sqrt{\frac{\left(Y_i'(\mathbf{r})-w\right)r_i-\sigma_i-\sqrt{\left(\left(Y_i'(\mathbf{r})-w\right)r_i-\sigma_i\right)^2+4(\sigma_i-\mu_i)\left(\left(Y_i'(\mathbf{r})-\underline{w}_i\right)r_i-\mu_i\right)}}{-2(\sigma_i-\mu_i)}}}, \tag{19}$$

where

$$\begin{aligned} r_i &= \frac{\kappa_i}{(1+\kappa_e)\,(1+\kappa_i)}q_i, \qquad Y_i'\,(\mathbf{r}) = \frac{\alpha_i}{r_i}p\prod\nolimits_{i=1}^{I} r_i^{\alpha_i}, \\ \sigma_i &= \alpha_i\,(\xi-1)\,Y\,(\mathbf{r})\,, \quad \text{and} \quad \mu_i = \frac{r_i}{\sum_i r_i}\frac{1}{2}\sum\nolimits_i \sigma_i. \end{aligned}$$

Recognizing that $F_i(\overline{w}_i) = 1$ we use $Y(l(\mathbf{w}))$ to get the following solution for the common productivity parameter

$$p = \frac{r_i}{\prod_{i=1}^{I} r_i^{\alpha_i}} \left[\alpha_i - \frac{\xi - 1}{\eta} \left(\frac{\xi (1+\eta) r_i}{2 \sum_i r_i} - \alpha_i \right) \right]^{-1} \left(\frac{\overline{w}_i - \eta \underline{w}_i}{1 - \eta} \right). \tag{20}$$

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

Consider the unknowns of the econometric model. The skill measures $\{q_i\}_{i=1}^{I}$ are known from the data and they are nothing else but sample sizes of each skill group. Furthermore, to avoid bounds of the likelihood function depending on the parameters, Kiefer and Neumann (1993) suggest extreme order statistics $\{\min(w_i), \max(w_i)\}$ as the consistent estimates for $\underline{w}_i$ and $\overline{w}_i$ respectively. Finally, from the fact that (20) holds for any i one can represent any α_i as a function of ξ and the rest of structural parameters. Namely (20) implies that for any $i, l = 1, .., I$ there 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 that is easily verified to provide a unique solution for $\boldsymbol{\alpha}$ in terms of $\left\{ \{\kappa_i\}_{i=1}^{I}, \kappa_e, \delta, \xi \right\}$.[5]

To demonstrate that the model with the parameter space that eventually reduces to ξ and search frictions is identifiable it is enough to notice that 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). From this follows that production size ξ is uniquely identified from the labour costs data.

3.3 *Heterogeneous Firms*

For heterogeneous employers the production functions are given in (16). The relevant occupation-specific wage offer distribution $F_i(w)$ is provided in Proposition 5. Rewritten

[5] 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 since $\frac{\overline{w}_l - \eta \underline{w}_l}{\overline{w}_i - \eta \underline{w}_i} > 0 \; \forall i, l$.

in $\kappa_{i,e}$ terms it becomes

$$F_i(w_i) = \frac{1+\kappa_e}{\kappa_e} - \frac{1+\kappa_e\left(1-\gamma_{j-1}\right)}{\kappa_e\sqrt{\frac{\left(Y_j'(\mathbf{r}_j)-w_i\right)r_{ij}-\sigma_{ij}-\sqrt{\left(\left(Y_j'(\mathbf{r}_j)-w_i\right)r_{ij}-\sigma_{ij}\right)^2+4\left(\sigma_{ij}-\mu_{ij}\right)\left(\left(Y_j'(\mathbf{r}_j)-\underline{w}_{ij}\right)r_{ij}-\mu_{ij}\right)}}{-2\left(\sigma_{ij}-\mu_{ij}\right)}}}, \quad (21)$$

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 $j = 1, ..., J$. Additionally we assume that for any i and j none of α_{ij} is equal to each other.

Remembering that $\gamma_j = F_i(\overline{w}_{ij})$ we use (16) and (21) to derive the productivity level of the firm

$$p_j = \frac{r_{ij}}{\prod_{i=1}^{I} r_{ij}^{\alpha_{ij}}}\left[\alpha_{ij}-\frac{\xi_j-1}{\eta_j}\left(\frac{\xi_j\left(1+\eta_j\right)r_{ij}}{2\sum_i r_{ij}}-\alpha_{ij}\right)\right]^{-1}\left(\frac{\overline{w}_{ij}-\eta_j\underline{w}_{ij}}{1-\eta_j}\right), \quad (22)$$

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, existence of the unknown cutoff wages does not allow using (22) to write down α_{ij} as a function of exclusively ξ_j and frictional 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 (22) 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 (22) two representations of the model are possible:

1. cutoff wages $\{\overline{w}_{ij}\}_{i,j=1}^{I,J-1}$ are expressed as a function of production parameters $\{\alpha_{ij}\}_{i,j=1}^{I-1,J}$, search frictions and $\boldsymbol{\xi}$,

2. production parameters $\{\alpha_{ij}\}_{i,j=1}^{I-1,J}$ are expressed as a function of cutoff wages $\{\overline{w}_{ij}\}_{i,j=1}^{I,J-1}$, search frictions and $\boldsymbol{\xi}$.

First of all, irrespective of the choice of the parameter subset to be substituted out, (22) implies that there exist $J(I-1)$ independent equations that completely determine cutoff wages and production parameters.[6] Moreover, for I skill groups there exist $(J-1)I$ unknown production parameters and $J(I-1)$ unknown cutoff wages. Since both above 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.

Next, we notice that despite both specifications are equally possible, expressing cutoff wages as a function of the rest of the parameters, is the strictly dominated one. The reason is that cutoff wages are the discontinuity points of the likelihood function, so substituting them with known functions of the rest of the parameters means that no gradient-based methods can be used when estimating 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 becomes violated at the solution. This case will imply constrained derivative-free optimization subject to the no mass point condition (for detailed discussion see Proposition 4 and p.22 later on), which is already a very difficult task.

Choosing the second way to represent the model one can show that (22) implies that for any $i, l = 1, .., I$ there holds an identity

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

which 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.

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 $\left\{\{\kappa_i\}_{i=1}^{I}, \delta, \kappa_e, \left\{\xi_j\right\}_{j=1}^{J}\right\}$. Chernozhukov and Hong (2004)

[6] i.e. 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.

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 labour costs data.

The above representation of the model fits into a convenient stepwise estimation strategy developed by Bowlus et al. (1995). 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 $\boldsymbol{\theta}$. The second step is a "smooth" optimization and can be efficiently executed using the gradient-based methods. Given the estimates from both steps into (4) and (8) we calculate the new point mass values γ_j

$$\gamma_j = 1 - \sum_{i=1}^{I} \frac{q_i}{m} \frac{1 - \hat{G}_i(\overline{w}_{ij})}{1 + \kappa_e \hat{G}_i(\overline{w}_{ij})}, \tag{23}$$

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

$$\mathbf{I} = n^{-1} \sum_{i=1}^{n} \frac{\partial}{\partial \boldsymbol{\theta}_1} \mathcal{L}_i(\boldsymbol{\theta}) \frac{\partial}{\partial \boldsymbol{\theta}_1} \mathcal{L}_i(\boldsymbol{\theta}'). \tag{24}$$

Furthermore Chernozhukov and Hong (2004) validate bootstrap methods for the estimation of 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. As argued in Proposition 4 a mass point can only exist, if increasing the wage further would imply that the additional wage cost outweighs the additional output produced with the additionally recruited workers. Consider an arbitrary skill group h. Proposition 4 implies that the distribution function $F_h(w_h)$ is continuous, if condition (10) is satisfied, i.e.

$$\alpha_{hj} \frac{p_j \prod_{i=1}^{I} l_i(\overline{w}_{ij})^{\alpha_{ij}}}{l_h(\overline{w}_{hj})} > \overline{w}_{hj}. \tag{25}$$

The estimated parameters are consistent only when the model is properly specified, i.e. when (25) holds.

It is also easy to see that in a special case with no skill differentiation, constant returns and unique productivity type firms, which is the original Burdett-Mortensen model, (25) gives us $1 > \overline{w}/p$, which is always true, implying continuous offer distribution in the basic BM model.

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

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

In terms of the Burdett-Mortensen model with discrete employer heterogeneity the above condition will imply the ranking of productivity levels. Possibility of violation of productivity ranking in applied work is discussed by Bowlus et al. (1995), p.S127.

One should also keep in mind that whenever any of the above restrictions is bindig at the maximum the asymptotic covariance matrix of the ML estimator is no longer given by (24) and the exact form of it is unknown. Moreover even in the simpler models with inequality constraints it is shown that bootstrap fails to consistently estimate the covariance matrix when the true parameter is on the boundary of the parameter space (see Andrews, 2000, for discussion).

Finally we notice that in the extended model with distinct productivity types another (weaker) way to see whether (12) holds is to consider $\hat{G}_i\left(\overline{w}_{ij}|\arg\max_{\{\boldsymbol{\theta},\gamma_j\}}(\mathcal{L})\right)$. Both (12) and (4) imply that $\hat{G}_i = \hat{G}_l\ \forall i, l \in [2, I]$. At the same time (23) does not restrict $\hat{G}_i$ to be equal to each other. Thus, if $\{\boldsymbol{\theta},\gamma_j\}\ \forall j \in [2, J-1]$ is a consistent estimate of the true parameters the values of the empirical earnings distribution at the skill-specific cutoff wages must not be significantly different from each other.

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 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 (see Koning et al., 1995, van den Berg and Ridder, 1998).

To estimate the model we need have 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. Both elapsed and residual spells can be incomplete due to overshooting the starting and terminal dates of the observation period while the spell is still in progress. To minimize the number of incomplete spells and at the same time provide the most recent information about the length of total unemployment or job duration we choose 1995 as a reference year. Whenever residual spell is complete we also record information about the exit state (one should keep in mind that in the setting of the model, job-to-job changes are also considered as “change of state”).

Unemployment duration is calculated from the retrospective labour force status calendarium of the GSOEP, in which respondents have to provide their labour force status for every month of the previous calendar year.

Retrieving job duration requires a bit more elaboration. First of all every currently employed individual provides information about the calendar month and year of the job start. Though for those who have undergone a job change we need to check additionally the date and the type of this job change. Apart from job changes to a new employers or within firm job changes with wage promotion, which classifies as change of state, this can also be company takeover, return to work etc. Thus only simultaneous application of both sources of information allows us to find the correct starting date. Similarly we find the endpoint of the job spell. The calendar end of job spells is set to the first reported job end in subsequent waves or to the first reported job start with new employer or within the same firm.

We also need to comment on incomplete spells. Those incomplete from the left can be seldom observed in the data. In our data set, the main reason for a spell being incomplete from the left is that it is not always possible to determine its exact calendar month (sometimes even year), because the respondent was simply not interviewed prior to the start of the spell. There are much more spells incomplete from the right. This happens because of the two reasons. First of all, the spell can still be in progress by the end of the available observation period. Secondly, spells that terminate by exit to non-participation are treated as right-censored.

The final bit of information necessary for the estimation of the model is earnings. Here we differentiate between net wage received by the worker and labour 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 labour size depend on just the position of the firm in the wage offer distribution, frictional parameters can be estimated using any of the two types of earnings data, provided that the ordering of the firms does not change when we pass from net wages to labour costs. For identification of the production parameters, to the contrary, labour costs are crucial because they enter the employers' profit maximization problem explicitly.

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 sample of job spells we use wage information provided by respondents at the year for which the sample is drawn. For the sample of unemployment spells we use the first reported wage after the end of unemployment, given that the transition to the job is observed. All wages are deflated by the West German consumer price index at prices of 1998. Labour costs are defined as a sum of gross wage and firms' contributions to the employees' social security payments. Information on the latter is available form the Social Security Office.

In our application we estimate the model with two different productivity levels and two different skill groups. Skill stratification of the sample is performed on the basis of the "International Standard Classification of Education (ISCED)" code. We identify as "low-skilled" all individuals who have inadequate, general elementary or middle vocational training. As "high-skilled" we qualify all the rest, i.e. those with higher vocational training, university education etc.

Summary of duration and wage data is presented in Table 1 and Table 2 respectively. Along with the information about the full sample we present summary statistics for the two skill groups. The data on skills reflect such basic facts about less skilled in comparison to high-skilled as higher level of unemployment, higher rate of job loss and longer

Table 1: Descriptive Statistics of Event History Data *

	Full Sample	Low-Skilled	High-Skilled
Number of individuals:	3893	2831	1062
Employed:	3558	2533	1025
Unemployed:	335	298	37
Employed Agents:			
Uncensored observations with:			
job → job transition:	414	236	178
job → unemployment transition:	265	224	41
mean time spell between two states [job duration]:	107.576	116.150	89.566
(std. deviation):	(101.01)	(106.67)	(85.42)
Censored observations			
a) left-censored durations only			
with job → job transition:	21	15	6
with job → unemployment transition:	15	14	1
b) right-censored durations only:	2764	1983	781
c) both left- and right-censored durations:	79	61	18
Mean time spell [both uncensored and censored]:	155.685	156.328	154.096
(std. deviation):	(118.74)	(118.13)	(120.30)
Unemployed Agents:			
Uncensored observations (u → j transition):	100	87	13
mean time spell between two states [job duration]:	19.850	21.241	10.538
(std. deviation):	(21.31)	(22.06)	(12.22)
Censored observations			
a) left-censored durations (u → j transition) only:	3	3	-
b) right-censored durations only:	219	195	24
c) both left- and right-censored durations:	13	13	-
Mean time spell [both uncensored and censored]:	35.328	36.701	24.270
(std. deviation):	(33.56)	(34.73)	(23.07)

* *Duration data in Months*

Table 2: Descriptive Statistics of Earnings Data

	Full Sample	Low-Skilled	High-Skilled
Labour Costs:			
Sample Minimum:	733.	733.	1645.
Mean Cost:	5551 (2260)	4998 (1813)	6950 (2642)
Sample Maximum:	20524.	17349.	20524.
Net Wages:			
Sample Minimum:	603.	603.	952.
Mean Wage:	3095 (1356)	2757 (1034)	3962 (1663)
Sample Maximum:	11535.	9524.	11535.

unemployment duration. Additionally net wages and labour costs are summarized by kernel density plots (see Figures A.1-2 in the Appendix). As expected, densities of both net earnings and labour costs of the low-skilled are more skewed to the right than those of the high-skilled. Also mean net wage of high-skilled workers amounts to DM 3962 which exceeds those of less skilled by more then 40%. Labour costs are roughly the same across the skills and almost double the net wage.

4.2 *Estimation Results: Overall 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 <the estimates are not reported here> to compare it with the results provided by our extension. It turns out that the structural parameters estimated with both original and extended constant-returns specifications do not significantly differ from each other, which implies that from the empirical perspective sole introduction of skill differences does not improve the estimates search frictions. Predicted theoretical offer and earnings densities (Figures A.3-4 respectively) for the extended theoretical model with constant returns have a jump at the reservation wage of the high-skilled and a spike at the maximum wage of the low-skilled, which generates a “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 ability to fit the data.

The results change when we switch to the increasing returns technology specification (the second column in Table A.1). First, the estimate of κ_1 now almost precisely matches the observed 10.53% unemployment rate of low-skilled predicting 10,73%. The model underestimates κ_2 but at the same time provides much more realistic results for κ_e and δ. Addressing Figures A.3-4 we see that the model with increasing returns implies offer and earnings densities with strictly decreasing right tail even in absence of productivity dispersion. However, the predicted earnings density is too flat implying that the formulation with identical firms also has quite a limited capacity of fitting the data. Furthermore, the initial unrestricted estimates of the model with increasing returns to scale did not meet the no mass point condition of Proposition 4, so in Table A.1 we present the estimates which are obtained by maximizing the likelihood function subject to (25). Failure to meet the requirements of Proposition 4 might suggest that even despite an improvement in comparison to the specification with constant returns technology the assumption that firms are identical is too strong.

Next we use estimate the model with two skill groups and two distinct productivity levels. As before, we also fit the original Burdett-Mortensen model with $J = 2$. Again, the results of the original Burdett-Mortensen model and our extension with constant returns almost do not differ from each other. Even though two jumps at the left tail and two spikes at the right one improve the fit of the aggregate earnings density predicted using the constant returns specification (see Figures A.5-6), locally increasing right tail of individual-specific densities still keeps this fit being far from satisfactory.

Relaxing the assumption of constant returns again changes the picture. Though, like in the case with identical firms, the unrestricted MLE still violate the no mass point condition. Therefore we perform the estimation of the model given (25). We also note that the magnitude of constraint violation at the unrestricted maximum has largely decreased with adding another productivity type. This points towards vanishing of the specification restriction when sufficiently high degree of employer heterogeneity is introduced.

The estimates of the model with increasing returns and two-point productivity dispersion are presented in the last column of Table A.2.[7] Comparing them with the estimates from the specification with identical firms and increasing returns technology two improve-

[7] Here, like in Table A.1, we report confidence intervals based on (24). This, however, does not estimate the true asymptotic covariance matrix consistently (see Section 3.4).

ments can be noticed. First, we manage to obtain a better fit for the degree of returns to scale in the whole economy. According to our estimates the homogeneity degrees are 1.11 for the "low-productive" technology and 5.44 for the "high-productive" one. Given the estimated fraction of each technology $\left[\gamma_j - \gamma_{j-1}\right]$ these estimates imply the economy-wide returns to scale at the level of 1.38. This goes in line a number of 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 to scale hypothesis and range from about 1.1 to about 1.4 (see Färe at al., 1985, Kim, 1992, and Zellner and Ryu, 1998, among many others). Second, introduction of the productivity dispersion also leads to a better fitting offer and earnings densities (see Figures A.5-6).

At the same time the extended specification still fails to match the observed unemployment rate of 3.5% among the high-skilled workers predicting the value of 5.7%.

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 increase in output coming from educating the marginal individual equals the individual's and the government's investment costs.

Following Grout (1984), who discusses the hold-up problem as a potential source of underinvestment, Acemoglu (1996) and Masters (1998) provide 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. This is also true in the present paper, since firms earn positive profits. However, there can also be overinvestment in the model, since a lower unemployment rate among high skilled workers can increase the return to human capital investment to such an degree that workers overinvest into skills.[8] The low unemployment rate for high skilled workers compared to high skilled workers can be sustained since the higher match value from meeting a high skilled worker encourages firms to create more vacancies for high skilled workers. Thus, whether there are social returns to education in an economy depends not only on the skill specific wage offer distributions but also on the skill specific unemployment rates.

To be able to investigate the question whether there is over- or underinvestment, we first ask how many individuals a social planner would instruct to become high skilled. Let us assume that firms' profits are distributed arbitrarily among employed and unemployed

[8] This is due to the assumption of segmented labor markets for all skill groups, if we assumed a constant arrival rates across all workers, underinvestment into education would be inevitably.

workers. Since we assumed 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 the individuals incur to become high skilled workers.

Suppose the cost $\Delta c = c_2 - c_1$ of becoming high skilled is distributed according to some continuous distribution function $H(\Delta c)$ among individuals on the support $\Delta c \in [0, \infty)$. Assuming two skill levels the social planner's problem can then be written as

$$(\Delta c)^S = \arg\max_s \left[E\left[Y_j(\mathbf{l}(\mathbf{w}))\right] - m \int_0^s (\Delta c)\, dH(\Delta c) \right]$$

$$\text{s.t. } q_1 + q_2 = m \text{ and } q_2 = mH(s)$$

It follows that the socially efficient skill structure can be characterized by

$$\int_0^1 \left. \frac{\partial Y_j(\mathbf{l}(\mathbf{w}))}{\partial q_2} \right|_{q_1+q_2=m} = (\Delta c)^S,$$

which implies that social welfare is maximized if the cost the marginal individual incurs equals the output-increase generated by all firms.[9] It an be shown that for the j-type firm where $j \in \{1, 2\}$ the marginal change in output due to the marginal increase in the amount of skilled labour is

$$\left. \frac{\partial Y_j(\mathbf{l}(\mathbf{w}))}{\partial q_2} \right|_{q_1+q_2=m}$$
$$= Y_j(\mathbf{l}(\mathbf{w})) \left[\frac{\alpha_{2j}}{q_2} + \frac{2\kappa_e \alpha_{2j}}{1+\kappa_e[1-F]} \left(\frac{\partial}{\partial q_2} F \right) - \frac{\alpha_{1j}}{m-q_2} - \frac{2\kappa_e \alpha_{1j}}{1+\kappa_e[1-F]} \left(\frac{\partial}{\partial q_2} F \right) \right],$$

which implies an expected change in the total output by

$$\int_0^1 \left. \frac{\partial Y_j(\mathbf{l}(\mathbf{w}))}{\partial q_2} \right|_{q_1+q_2=m} = \int_0^{\gamma_1} \left. \frac{\partial Y_1(\mathbf{l}(\mathbf{w}))}{\partial q_2} \right|_{q_1+q_2=m} dF + \int_{\gamma_1}^1 \left. \frac{\partial Y_2(\mathbf{l}(\mathbf{w}))}{\partial q_2} \right|_{q_1+q_2=m} dF. \quad (26)$$

In order to evaluate if there are social returns to education we proceed in comparing the marginal increase in output with the cost the marginal individual incurred to become high skilled. It has to be true that the marginal worker is exactly indifferent between the two skill groups, i.e. $U_2 = U_1$. Thus, the private cost of educating oneself from the "low"

[9] Aggregate output is obtained by integrating from the firm offering the reservation wage schedule, i.e. $F_{i1}(w_i^r) = 0$, to the firm offering the maximum wage to all skill groups, i.e. $F_{iJ}(\overline{w}_i) = 1$.

to the "high" level – applying (1a) – can be written as

$$
\begin{aligned}
(\Delta c)^I &= rU_2 - rU_1 \\
&= \kappa_2 \int_{w_2^r}^{\bar{w}_2} \frac{1 - F_2(w)}{1 + \kappa_e (1 - F_2(w))} dw - \kappa_1 \int_{w_1^r}^{\bar{w}_1} \frac{1 - F_1(w)}{1 + \kappa_e (1 - F_1(w))} dw.
\end{aligned} \tag{27}
$$

Note, that (27) refers to the optimal decision of the searching 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. Applying to w_i^r the sample minimum estimator of Kiefer and Neumann (1993) and noticing that the reservation wage is given by (2) simplifies calculation of $(\Delta c)^I$ in practice.

Using our structural estimates we evaluate both (26) and (27) to see whether present skill structure is efficient. First we find that indeed a marginal change of the skill structure towards a larger share of skilled workers induces an increase in output. Furthermore, the expected marginal increase in output is big enough to offset private costs of this increase. Namely, we find that the expected marginal change in output makes DM 3252.78 per month whereas the individual costs of shifting from "low" to "high" education level amount to DM 281.48 per month. This implies that at least within the simplest possible framework with two distinct skill levels we find support for the hypothesis that there is underinvestment in the economy or put differently that the social returns to education exceed the private returns since individuals capture only part of the return of their human capital investment.[10]

Consequently, our results imply that policies promoting higher education would be welfare-improving. In addition to that, differentiating the output twice we compute the expected second order effect of the marginal shift of the skill structure towards high skills. It turns out that at the estimated parameter constellation the second order effect is negative, amounting to DM -11.76, which points at the decreasing returns of possible education-promoting policies.

Finally we have to notice that the ultimate marginal value of the difference between the expected output and private costs of education can also depend on the state subsidies that are paid out to the educating institutions. According to OECD statistics the expenditures for educating one low skilled worker to a high skilled worker by 1998 were DM

[10] Note, that we are only able to capture the direct effect of the change in the skill on output. Due to our assumption of a constant offer arrival rate we are not able to evaluate the impact a change of the skill structure has on the matching probabilities of workers and firms and hence on the measure of employed and productive workers. Since this indirect effect is of second order it should not alter our main finding.

164 per month.[11] These expenditures, however, need to be taken into account only if the government subsidy depends on the number of students. Though, in our particular case, even it were so, the output enhancing effect of the marginal increase of the fraction of high-skilled workers would still be sufficient to counter the total private and government costs of DM 445.48 per month.

The present paper offers a new approach to measure social returns to education within an equilibrium framework which takes the skill specific unemployment risk explicitly into account. Although we are able to provide consistent estimates of whether there is over- or underinvestment in an economy, our framework does not allow us to determine the source of the inefficiency. The underinvestment found could either be caused by the hold-up problem workers face when making their investment decision or by a human capital externality due to a education spillover. If compared to the literature we can thus only compare whether other studies found underinvestment or not.

Up to now the major part of empirical work identifying a human capital externality has been performed using a Mincer regression approach. Acemoglu and Angrist (2000) estimate a reduced-form "Mincerian" equation that is consistent with the model of Acemoglu (1996) in which positive externality is a theoretical result. Applying an instrumental variable technique they find almost no support for social increasing returns. Moretti (2004) provides additional elaboration on the method of Acemoglu and Angrist (2000) and develops own version of the appropriate econometric model. Considering in detail the issues of omitted variable bias, endogeneity and exogenous shocks Moretti (2004) backs the existence of positive externality using the same data of Acemoglu and Angrist (2000). Finally, within the same framework Dalmazzo and de Blasio (2003) consider locally disaggregated markets rather then one aggregate labour market and find significant positive effect of human capital externalizes.

The "Mincerian" approach has, however, two main drawbacks that can possibly lead to the ambiguity found in the empirical studies mentioned above. First of all, its' per-

[11] Following the International Standard Classification of Education (ISCED) of the OECD in 1998 there where about 2.5 million (full-time equivalent) individuals enrolled in post secondary or tertiary education. The total expenditure for these education levels amounted to around 25 million DM per year. Given an average enrolment duration of 4.9 years to complete a tertiary education we estimate that the government spends around 49,000 DM for the education of an individual enrolled in tertiary education. Assuming the annual interest rate of 4% the flow value of the government expenditure on a monthly basis is given by DM 164 per month. Source: Education Database at http://www1.oecd.org/scripts/cde/members/EDU_UOEAuthenticate.asp

formance depends of the availability of appropriate instrumental variables and on the identifying assumptions concerning the time-varying nature of endogenous dependence between pecuniary human capital externality and unobserved individual abilities (see Moretti, 2004). Secondly, Ciccone and Peri (2002) theoretically demonstrate that even with the constant returns to scale technology the "Mincerian" estimates are biased whenever workers of different skill groups are imperfect substitutes. On grounds of this critique Ciccone and Peri (2002) develop a "constant-composition" approach which assumes that human capital accumulation does not change the skill structure of the workforce so that social returns to education can appear only due to the increasing average level of human capital in the society. Implementing this approach empirically they do not find support to this hypothesis.

Abstracting from the application to social returns, our results also appear to be in line with those of Falk and Koebel (1999) who show that output is a positive and increasing function of skills and that output effect dominates in explaining the shift away from unskilled labour in Germany.

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 constant and increasing returns to scale.

The main theoretical contribution of this paper is that allowing production function to have any degree of homogeneity returns to scale we are able to generate a decreasing wage offer density. Subsequent introduction of employer heterogeneity leads to further improvement of the shape of 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.

Theoretical solution of our extension suggests a structural econometric model that allows estimating not only search frictions inherent to the labour market but also the parameters of the production function. Richness of the theoretical model enables us to estimate all parameters of interest using wage and duration data only, which requires no additional information on the output.

We apply our model to learn whether there is over- or underinvestment into human capital in Germany. Our results suggest that cost of the marginal shift of the skill composition of the workforce towards the larger share of skilled workers is lower than the expected increase in output, induced by this shift. This suggests that social returns to education exceed private return and that a policy designed to promote higher education would be welfare improving.

REFERENCES

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

Acemoglu, D., and J., Angrist "How Large are Human Capital Externalities? Evidence from Compulsory Schooling Laws", NBER Macroannual (2000), 9-59.

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, "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.

CICCONE, A., AND G., PERI, "Identifying Human Capital Externalities: Theory with Application to US Cities", IZA Discussion Paper No. 488 (2002).

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

DALMAZZO, A., AND G., DE BLASIO, "Social Returns to Education: Evidence from Italian Local Labor Market Areas", IMF Working Paper WP/03/165 (2003).

FALK, M., AND B., KOEBEL, "Curvature Conditions and Substitution Pattern among Capital, Energy, Materials and Heterogeneous Labour", ZEW Discussion Paper No. 99-06 (1999).

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.

KREMER, M., "The O-ring Theory of Economic Development", *Quarterly Journal of Economics* 108 (August 1993), 551-575.

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.

MORETTI, E., "Estimating the Social Return to Higher Education: Evidence from Longitudinal and Repeated Cross-Sectional Data", *Journal of Econometrics* 121 (2004), 175-212.

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, B.J. Christensen, P. Jensen, N.M. Kiefer, and D.T. Mortensen, eds., *Panel Data and Structural Labour 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).

OECD, Education Database at http://www1.oecd.org/scripts/cde/members/EDU_UOE Authenticate.asp.

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.

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 Earnings Densities"

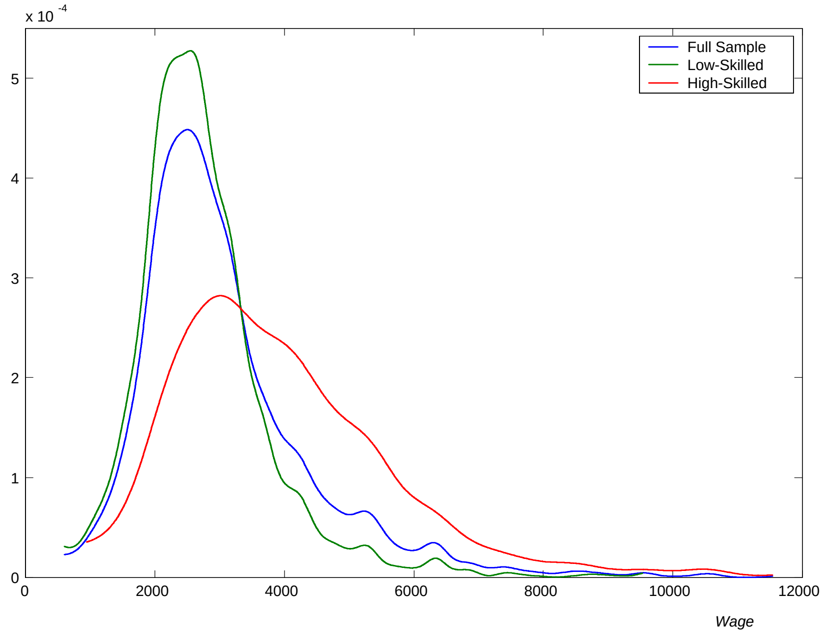


Figure A.2: "Kernel Estimates of Labour Cost Densities"

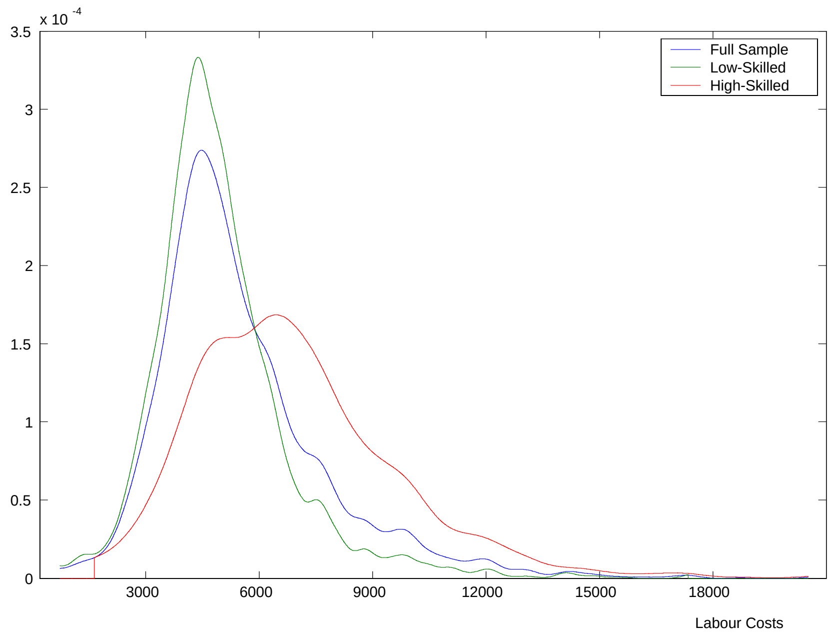


Table A.1: "Estimation Results: Homogeneous Firms"

	Specification			
	Constant Returns*		Increasing Returns	
κ_{u1}	6.3787	[5.9812, 6.7763]	8.3235	[7.7169, .9301]
κ_{u2}	14.1253	[12.5470, 15.7036]	18.2779	[15.7899, 20.7659]
κ_e	0.1527	[0.1352, 0.1703]	2.2996	[1.8870, 2.7121]
δ	0.0066	[0.063, 0.0068]	0.0041	[0.0039, 0.0044]
ξ			2.0000	[1.7859, 2.2141]
α_1	0.6824		1.3955	
$\ln(\mathcal{L})$	-66903.716		-65314.79	

* *95% Confidence intervals in square brackets*

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

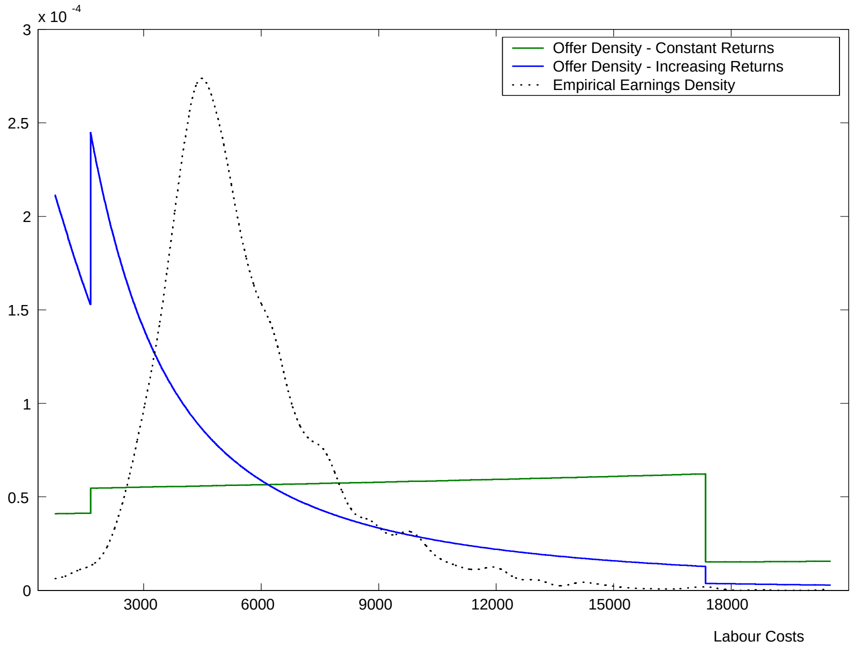


Figure A.4: "Theoretical Earnings Densities Predicted by the Model: Homogeneous Firms"

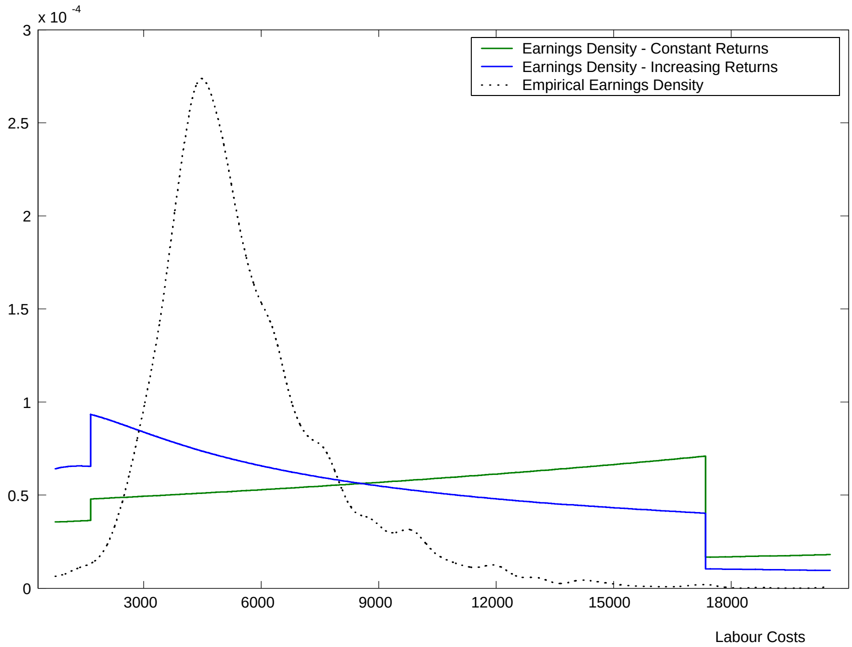


Table A.2: "Estimation Results: 2-Point Employer Heterogeneity"

	Specification			
	Constant Returns*		Increasing Returns	
κ_{u1}	7.1572	[6.6825, 7.6319]	7.5995	[7.0683, 8.1307]
κ_{u2}	15.7879	[13.8682, 17.7076]	16.5693	[14.4455, 18.6931]
κ_e	0.9858	[0.9118, 1.0597]	1.7426	[1.5640, 1.9212]
δ	0.0054	[0.0052, 0.0056]	0.0049	[0.0046, 0.0051]
ξ_1			1.1116	[1.0962, 1.1271]
ξ_2			5.4378	[3.3641, 7.5116]
α_{11}	0.6163		0.7103	
α_{12}	0.7119		3.8787	
$\overline{w}_{11}$	6332.		6332.	
$\overline{w}_{21}$	9955.		9955.	
γ_1	0.9174		0.9388	
$\ln(\mathcal{L})$	-64020.80		-63618.85	

* *95% Confidence intervals in square brackets*

Figure A.5: "Aggregate Wage Offer Densities: 2-Point Employer Heterogeneity"

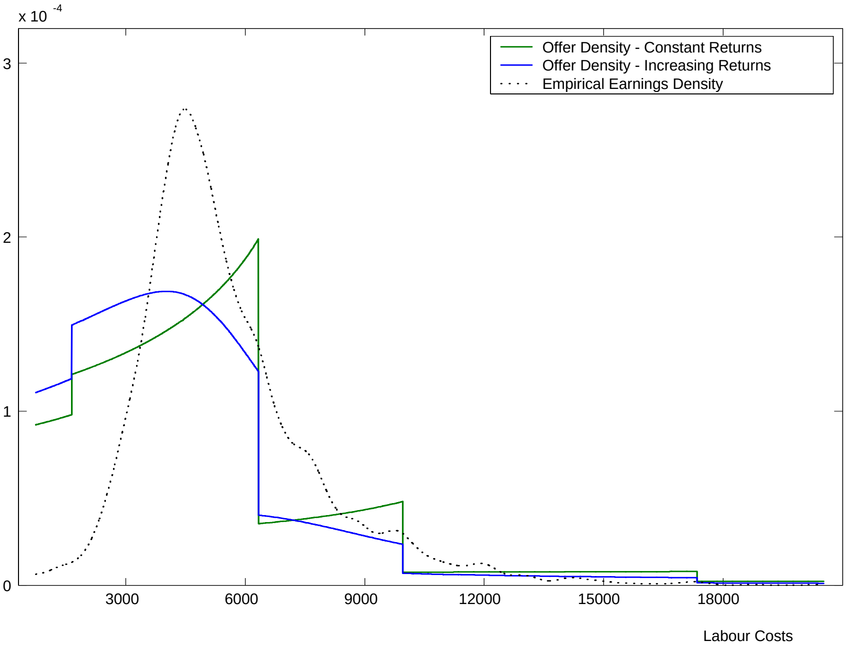
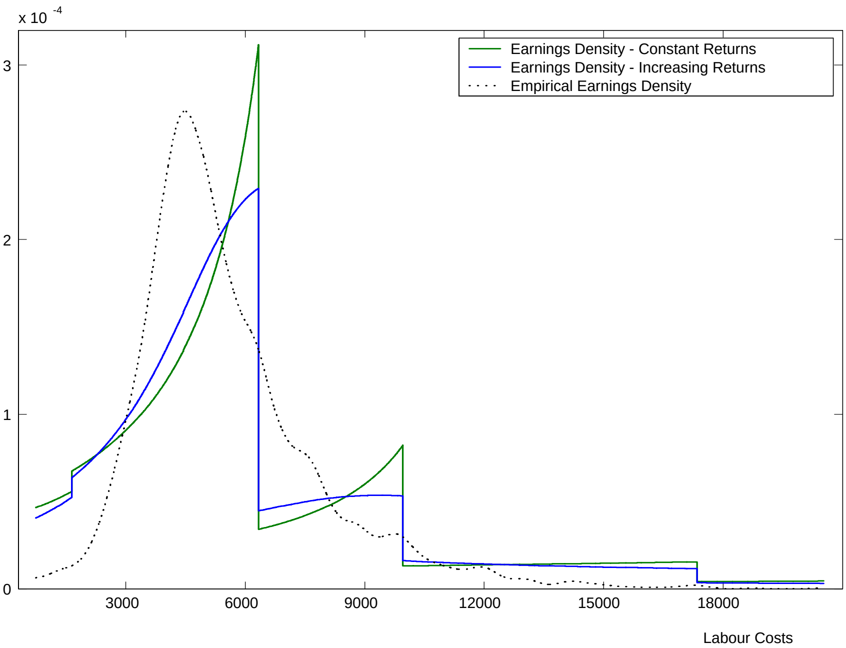


Figure A.6: "Theoretical Earnings Densities Predicted by the Model: 2-Point Employer Heterogeneity"



Proof of Proposition 5.

Define

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

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

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

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

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

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

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

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)

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

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

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

The first order condition can therefore be written as

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

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

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

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

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

Rearranging gives

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

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 we get

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

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

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

Otherwise, we get the following solution for the quadratic function, i.e.

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

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

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

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

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

Hence the cumulative wage offer distribution is given by

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

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

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

The slope of the wage offer density is given by

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

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. ■