

IZA DP No. 520

**Unobserved Bilateral Search on the Labor  
Market: A Theory-Based Correction for a  
Common Flaw in Empirical Matching Studies**

Uwe Sunde

June 2002

# Unobserved Bilateral Search on the Labor Market: A Theory-Based Correction for a Common Flaw in Empirical Matching Studies

**Uwe Sunde**

*IZA Bonn and University of Bonn*

Discussion Paper No. 520

June 2002

IZA

P.O. Box 7240

D-53072 Bonn

Germany

Tel.: +49-228-3894-0

Fax: +49-228-3894-210

Email: [iza@iza.org](mailto:iza@iza.org)

This Discussion Paper is issued within the framework of IZA's research area *Mobility and Flexibility of Labor*. Any opinions expressed here are those of the author(s) and not those of the institute. Research disseminated by IZA may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent, nonprofit limited liability company (Gesellschaft mit beschränkter Haftung) supported by the Deutsche Post AG. The center is associated with the University of Bonn and offers a stimulating research environment through its research networks, research support, and visitors and doctoral programs. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public. The current research program deals with (1) mobility and flexibility of labor, (2) internationalization of labor markets, (3) the welfare state and labor markets, (4) labor markets in transition countries, (5) the future of labor, (6) evaluation of labor market policies and projects and (7) general labor economics.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available on the IZA website ([www.iza.org](http://www.iza.org)) or directly from the author.

## ABSTRACT

### **Unobserved Bilateral Search on the Labor Market: A Theory-Based Correction for a Common Flaw in Empirical Matching Studies\***

This paper develops a model of equilibrium unemployment with (unobservable) endogenous on-the-job search and (partly unobservable) endogenous search behavior by firms. The model allows to analyze crowding-out of unemployed job seekers by endogenous on-the-job search of employees, and the interaction of job search behavior and vacancy posting policies of firms. Moreover, it is shown that the neglect of endogenous but not observable behavior in the empirical literature on labor matching leads to systematically biased estimates of the matching elasticities, posing a caveat on the results of previous studies testing for constant returns of the matching function. The theoretical model presented allows to predict the direction of the bias. We propose a correction for the estimates of empirical matching functions that leads to unbiased estimates of the matching elasticities. The empirical implications of the theoretical model are tested and confirmed using German administrative data, and unbiased estimates of matching elasticities are presented.

JEL Classification: J41, J60, J63, J64, C23

Keywords: empirical matching functions, job competition, on-the-job search, hiring competition, search channels, omitted variable bias

Uwe Sunde  
IZA  
PO Box 7240  
53072 Bonn  
Germany  
Tel.: +49 228 3894 533  
Fax: +49 228 3894 210  
Email: [sunde@iza.org](mailto:sunde@iza.org)

---

\* I am very grateful to Christian Belzil, Stephan Kohns, Wendelin Schnedler, Margaret Stevens, Jan van Ours and particularly to René Fahr for helpful comments. All errors are my own. Financial support from DFG is gratefully acknowledged.

# 1 Introduction

Matching functions are at the heart of many macroeconomic models of the labor market. While the theoretical literature is well developed, the empirical literature on matching functions is still reasonably small (see Petrongolo and Pissarides (2001) for a recent survey). Most of the existing empirical literature concentrates on the question whether the assumption of constant returns in the matching function can be validated empirically. This assumption ensures that the equilibrium unemployment rate is unique along the steady state growth path. Since the existence of multiple equilibria would provide scope for policy interventions, reliable estimates of matching elasticities are therefore of considerable importance.

Recently, some contributions expressed concerns about the correctness of the conventional approach to estimating the structure of the matching technology. One branch of the literature, like Broersma and Ours (1999), argues that empirical models might be mis-specified since new hires are regressed on inappropriate stocks of job seekers and vacancies. Moreover, they emphasize the importance of accounting for job searchers not contained in the stock of unemployed which is conventionally used as proxy for the stock of job seekers.

Another caveat raised in the literature concerns the fact that endogenous behavior which is partly unobserved might render the coefficients estimated hard to interpret. In particular, Burgess (1993) and Anderson and Burgess (2000) argue that employed individuals condition their decision of whether to search on-the-job for another employment opportunity on aggregate labor market conditions. This decision is not directly observable in the data. However, such endogenous behavior affects the coefficients estimated using only observables since it affects the total number of new hires. Instead of including only information on the matching technology and therefore on the structure of labor markets, the coefficients contain the matching technology as well as the effects of endogenous changes in the respective pool of job seekers. This insight is extended in a paper by Fahr and Sunde (2001a) who account also for similar strategic behavior on the side of vacancy-posting firms which might be unobservable as well. However, the indicators used by Anderson and Burgess (2000) and Fahr and Sunde (2001a) for testing for the presence of endogenous and unobservable changes in the composition and/or size of the relevant pools of searching market participants are invalid once both sides are characterized by endogenous and unobservable behavior.

A third line of criticism addresses the assumption of random matching and undirected search underlying the conventional matching approach. As an alternative, authors like Coles and Smith (1998) and Coles and Petrongolo

(2002) assert the so-called stock-flow approach to matching according to which flows are determined by inflows into the relevant pools of searchers and vacancies instead of the pools themselves. However, also this strand of the literature is affected by issues of unobserved endogenous search behavior and the consequential neglect of relevant components of inflows into the stocks of job seekers and vacancies.

To sum up, there is no empirical approach to date that can test for simultaneous unobservable and endogenous shifts in both relevant pools of job seekers and vacancies. Nor is there a coherent study that investigates and quantifies the biases obtained by previous empirical studies of matching and job creation which neglect this issue.

This paper makes four contributions to the literature. First, we propose a simple theoretical matching model of the labor market that includes endogenous and (potentially) unobservable on-the-job search as well as (partly) unobservable search by firms. Unemployed job searchers, who are observed in the data, differ in their endogenous behavior from employed job searchers, who usually cannot be identified in data. Firms have the possibility to choose the channel through which they advertise their vacancies out of a multitude of possible search channels. Some of the channels used by firms are not observable by the statistician. The novelty of the model is that endogenous behavior on both sides affects the structure of stocks of job searchers and vacancies, and changes the composition of the relevant pools. As a result, the distributions of observable and unobservable components on either side are determined endogenously, with implications for job competition and starting wages of the different groups of job applicants.

Secondly, we show that conventional empirical matching functions regressing the flow of new hires on the observable stocks of unemployed searchers and vacancies registered at employment offices yield substantially biased estimates of the elasticities of the matching function. The reason is that important and influential explanatory variables are omitted in the estimation. In particular, the endogenous changes in the composition of observable and unobservable components of the search pools are neglected. We also show that alternative models like the stock-flow approach suffer from the same problems as the conventional approach.

Thirdly, we devise a method to obtain unbiased estimates of the true elasticities governing the matching process. This method simultaneously allows to test some crucial implications of the theoretical model. Using German administrative data, we investigate the empirical relevance of both the theoretical model and the correction for the bias due to omitted variables. Moreover, we extend the literature on the wage effects of job creation by providing testable predictions for the premium of employed over unemployed

applicants in average starting wages. The predictions and implications of the theoretical model are tested empirically and are shown to be in line with the data.

Finally, this paper presents the first coherent and thorough framework for interpreting the results of previous empirical studies of the matching process using different concepts of flow data for hirings and stocks of job seekers and vacancies, and thereby extends the work of Broersma and Ours (1999).

The remainder of the paper is organized as follows. Section 2 presents a simple matching model of the labor market with endogenous on-the-job search and alternative search channels for firms. In particular, some empirically testable predictions of the model are worked out. Section 3 shows that conventional empirical matching functions deliver biased estimates. The direction of the bias can be predicted using the theoretical framework. In Section 4 we present a way to isolate the bias and thus to compute unbiased estimates of the true matching elasticities. Using German administrative data we compute biases and unbiased estimates. Moreover, we test the empirical implications of the model presented in Section 2. Section 5 concludes.

## 2 A Simple Model of Endogenous Search

This section lays out a simple matching framework with endogenous on-the-job search and endogenous vacancy posting.

### 2.1 The Labor Market

Consider a labor market characterized by frictions like imperfect information, trading frictions etc., which are represented using a standard matching function. The matching function describes the instantaneous flow of new employer-employee relationships  $m$  as a function of people actively searching for a job,  $J$ , and of the total number of vacancies  $V$  firms try to fill. The more individuals search for a new job, and the more vacancies firms try to fill, the more new jobs are created. The matching function exhibits constant returns to scale and is written as:

$$m = m(J, V), \tag{1}$$

with  $\frac{\partial m}{\partial J} > 0$ ,  $\frac{\partial m}{\partial V} > 0$  and  $m(0, V) = m(J, 0) = 0$ .

The pool of job searchers consists of unemployed and employed individuals searching on-the-job. Individuals are *ex ante* homogenous and differ only with respect to their labor market status. However, *ex post* individuals are characterized by match heterogeneity, see Burdett and Coles (1999). In

what follows, we normalize the labor force to 1, and denote the number of unemployed (or equivalently the unemployment rate) as  $u$ , and the number of employed job seekers as  $e$ . This number can more conveniently be expressed as a share  $\phi$  of job seekers amongst the employed:  $e = (1 - u)\phi$ , where  $\phi \in [0, 1]$ .<sup>1</sup> The fraction  $\phi$  is determined endogenously by individually rational behavior, and depends positively on the arrival rate of job offers.<sup>2</sup> Intuitively, employed individuals observe the labor market situation and when many job offers become available, e.g. during a boom of the economy, they find it profitable to search for a better job. The pool of job seekers is composed of two distinct groups:  $J = u + (1 - u)\phi$ . Note that in empirical investigations, the number of persons searching on-the-job is unobservable. Moreover, every individual can only apply once at a time for a new job.<sup>3</sup>

Also firms adjust their search behavior endogenously. During all phases of the business cycle there is significant job destruction and creation, see e.g. Davis and Haltiwanger (1992). On the one hand, firms try to replace workers who dropped out for various reasons, thereby creating a pool of vacancies, most of them usually registered at employment agencies. On the other hand, firms have plans how to develop over time, to expand, to shrink or to alter their activities. While they more or less have to cope with the situation on the labor market in the former case, they can take labor market conditions into account more explicitly in the timing and planning of more fundamental re-organizations. Denote vacancies registered at employment offices by  $r$  and vacancies which are advertised through alternative channels as  $n$ . The latter type of vacancies captures vacancies related to such ventures that take the labor market conditions explicitly into account. Again, in empirical work one usually observes  $r$  while  $n$  is unobservable.<sup>4</sup> Intuitively, the latter type of vacancies represents e.g. newspaper ads, private contacts, alumni networks, headhunters etc.<sup>5</sup> The total number of vacancies is therefore given by the

---

<sup>1</sup>In the data used for the empirical application below, on average 30 percent of all matches are accessed by employed applicants. This illustrates the importance of taking employed job search explicitly into account.

<sup>2</sup>Endogeneity of on-the-job search behavior is discussed in more detail below.

<sup>3</sup>For a microfoundation of a matching function allowing for multiple applications, see Albrecht, Gautier, and Vroman (2002).

<sup>4</sup>In the data used for the empirical application, new hires from registered vacancies account on average for 56.92 percent of all new hires, indicating that non-registered vacancies play a significant role.

<sup>5</sup>For an example, see Calvo-Armengol and Zenou (2001). Lindeboom, Ours, and Renes (1994) investigate empirically the relevance and effectiveness of alternative recruitment channels and find that alternative channels are very effective. Moreover, their results indicate that the use of informal channels is more cyclical than the use of registered vacancies, as is assumed in the model.

sum of both types:  $V = r + n$ . There is no *a priori* restriction on the number of vacancies, e.g. with respect to the size of the labor force, and firms can advertise all jobs either through registration  $r$  or alternative channels  $n$ . As a consequence of free entry, the numbers of vacancies of either type is determined by firms' optimizing behavior. The crucial difference between the two sorts of advertising vacant positions lies in the respective costs. While registration requires a constant fixed cost  $c_r$ , the costs for non-registered vacancies vary with labor market conditions and the number non-registered vacancies posted by other firms. This will be discussed in more detail below. For notational simplicity, we define  $\pi$  as the proportion of non-registered to registered vacancies,  $n = \pi r$ . This proportion  $\pi$  is determined endogenously. We can therefore write  $V = (1 + \pi)r$ .

The matching function is assumed to characterize the entire labor market. By assumption, there is no discrimination between searchers, so all applicants whether employed or unemployed can take all sorts of jobs, and can apply for registered and non-registered vacancies. Firms have no preferences for searchers of a particular employment status.<sup>6</sup> The matching function characterizing the entire labor market can therefore be written as:

$$m = m(J, V) = m(u + e, r + n) = m(u + (1 - u)\phi, (1 + \pi)r) \quad (2)$$

In order to investigate the behavior on the labor market, it is convenient to define the labor market tightness as observed by firms as the ratio of all vacancies posted over all searchers looking for new employment:

$$\theta = \frac{r + n}{u + e} = \frac{(1 + \pi)r}{u + (1 - u)\phi}. \quad (3)$$

From this, it is easy to see that  $\frac{\partial \theta}{\partial \phi} < 0$ , and that  $\frac{\partial \theta}{\partial \pi} > 0$ .

Using this notation, and the fact that there is no ranking or discrimination, the flow probability for a given job seeker to form a match instantaneously is

$$p(\theta) = \frac{m(u + e, r + n)}{u + e} = m(1, \theta), \quad (4)$$

using the constant returns to scale assumption. Note that  $\frac{\partial p}{\partial \theta} > 0$ .

---

<sup>6</sup>This is a weaker assumption than the ranking of job applicants referred to by Blanchard and Diamond (1994) and Gautier (2002), or adverse selection issues as investigated by Kugler and Saint-Paul (2001). There is also evidence that registered vacancies are preferably filled by unemployed seekers while firms use mostly alternative recruitment channels for attracting employed applicants, see Lindeboom, Ours, and Renes (1994). Making an assumption of ranking and heterogeneity of job creation reflecting this evidence would even reinforce the main results of the paper.

Analogously, noting that all individuals can do all sorts of jobs, the instantaneous flow probability of a given vacancy to be filled is given by:

$$q(\theta) = \frac{m(u + e, r + n)}{r + n} = \frac{p(\theta)}{\theta}, \quad (5)$$

with  $\frac{\partial q}{\partial \theta} < 0$ .

## 2.2 On-the-job Search

Individuals only search on-the-job for alternative jobs which are “better” than their current job, in present discounted value terms, e.g. because they involve better match quality, a higher position in the hierarchy, a leap in the career etc. Intuitively, the more jobs are available, the higher is the probability of finding such an “upgrade” job since, by assumption, the driving shocks affect all hierarchical levels in firms identically.<sup>7</sup> Since the expected gains of such a job increase relative to the search costs as the probability of finding a new job increases, more employed start searching on the job if more jobs are available.

Assume for simplicity that all individuals are *ex ante* identical, in particular they have the same reservation utility, and the same costs for searching on-the-job,  $\kappa$ . At the encounter of a new match, firm and worker realize the quality of the match which determines the productivity of the match. This match quality is irreversible.<sup>8</sup> Let the productivity  $y$  be composed of a baseline productivity  $\underline{y}$  plus a match specific mark-up  $\epsilon$ :  $y = \underline{y} + \epsilon$ , where  $\epsilon \in [0, \bar{\epsilon}]$  is identically and independently distributed across individuals according to a continuous cumulative distribution function  $F(\epsilon)$ .<sup>9</sup> Wages are set according to a fixed exogenous sharing rule of the match surplus between firms and workers, as obtained through Nash bargaining, so there is a non-degenerate wage distribution in the economy. Wages depend on productivity, and thus on the realized quality of the respective match. For simplicity, shocks affect the baseline productivity  $\underline{y}$ , and therefore leave the ordering of match qualities and hence wages unchanged across the population. However, such a change in  $\underline{y}$  alters *ceteris paribus* the total productivity of a match.

<sup>7</sup>This is similar to the argument made by Pfann (2001) for downsizing periods: Not entire layers of the hierarchical structure of a downsizing firm are removed, but all layers are slimmed more or less proportionally.

<sup>8</sup>See Pissarides (1994) for an alternative approach with on-the-job learning. Including this possibility would leave the main results unchanged, and thus we neglect it in what follows.

<sup>9</sup>In equilibrium, this distribution of match qualities is stationary and characterizes existent or available jobs in the economy at any point in time.

The value of being employed from the individual viewpoint,  $\mathcal{E}$ , increases monotonically in the wage of a job, and thus indirectly in match quality  $\epsilon$ . On the other hand, individuals searching on-the-job accept alternative offers and quit their current job if and only if the alternative job offers a higher present discounted value of earnings than the current one. Thus, the higher current match quality, the smaller the probability that such an 'upgrade' job can be encountered, or, in other words, the longer the search period needed to find a better job.<sup>10</sup> An employed worker holding a job of quality  $\epsilon^*$  is therefore indifferent between searching on-the-job and not searching if the costs of searching equal the expected benefits:

$$\kappa = (1 - \rho)p(\theta)(1 - F(\epsilon^*)) \left[ \mathcal{E}_{t+1}(w(\int_{\epsilon^*}^{\bar{\epsilon}} \epsilon f(\epsilon) d\epsilon)) - \mathcal{E}_{t+1}(w(\epsilon^*)) \right], \quad (6)$$

where  $\rho$  is the discount factor, and  $w(\cdot)$  denotes the wage earned in a job, depending on the respective match quality. Note, however, that the benefits are strictly decreasing in match quality, as the smaller probability of finding an even better match outweighs the higher expected match quality (and thus wage).<sup>11</sup>

Hence, individuals with job matches below this endogenous threshold quality  $\epsilon^*$  search on-the-job for a job paying a higher wage than their current job.<sup>12</sup> The higher the wage a worker receives in his current job, that is the higher the quality of the current match, the smaller his incentives to

---

<sup>10</sup>Given the distribution of match qualities is  $F$ , the probability of a worker with a job  $\epsilon'$  finding a job better than the current one is  $1 - F(\epsilon')$ , which is decreasing in the current match quality.

<sup>11</sup>Taking the partial derivative of the benefits of on-the-job search, the right hand side of Equation (6), with respect to  $\epsilon^*$  gives:

$$(1 - \rho)p(\theta) \left[ -f(\epsilon^*) \left[ \mathcal{E}_{t+1}(w(\int_{\epsilon^*}^{\bar{\epsilon}} \epsilon f(\epsilon) d\epsilon)) - \mathcal{E}_{t+1}(w(\epsilon^*)) \right] - \right. \\ \left. (1 - \rho)p(\theta) \left[ (1 - F(\epsilon^*)) \left[ \frac{\partial \mathcal{E}_{t+1}(\cdot)}{\partial w} \frac{\partial w(\cdot)}{\partial \epsilon} (\epsilon^* f(\epsilon^*) + 1) \right] \right] \right] < 0,$$

since  $\mathcal{E}$  is strictly increasing in the wage  $w$ , and  $w$  is strictly increasing in match quality  $\epsilon$ .

<sup>12</sup>Since endogenous on-the-job search is the focus of this model, we assume  $\kappa > 0$  has a value such that an interior solution exists. If  $\kappa = 0$ , everybody except for those individuals with a perfect match  $\bar{\epsilon}$  would search on the job. On the other hand, if the search costs were prohibitively high, nobody would search on the job. In these cases, the model would still be valid, but the most interesting results would not be obtained. A similar microfoundation for endogenous on-the-job search can be found in the paper by Boeri (1999). The presented modeling of employed search differs from Pissarides (1994), who assumes that firms decide *ex ante* whether to create high or low productivity jobs, and where the quality of a match increases over time. Burgess (1993) does not explicitly model on-the-job search, but assumes that there are always some employed that search, while in the model presented here, search is an outcome of individual optimizing behavior.

search on the job. In fact, if there is a nondegenerate wage distribution with bounded support, the individual earning the highest wage possible has never an incentive to search on-the-job as long as there are positive costs associated with search. This ensures uniqueness of the equilibrium, see also Burdett and Coles (1999).

As a consequence, the fraction  $\phi$  of employed individuals that search for alternative employment is endogenously determined by the threshold match quality  $\epsilon^*$  satisfying condition (6), with  $\phi = F(\epsilon^*)$ . Consider again the individual that is just indifferent between searching and not searching with a match quality  $\epsilon^*$ . A slight increase in the probability of finding an alternative 'better' job  $p(\theta)$  would induce this individual to search actively, and also other individuals with slightly lower match qualities who are inactive given the initial finding probability. Thus, the share of employed searching,  $\phi$ , is implicitly determined by endogenous on-the-job search behavior and is a positive function of labor market tightness  $\theta$ :

$$\frac{\partial \phi}{\partial \theta} > 0. \quad (7)$$

Formally, the value of being employed  $\mathcal{E}$  for a given individual is then characterized by the following Bellman equation:

$$\mathcal{E}_t(w(\epsilon')) = w_t(\epsilon') + (1 - \rho - \delta)\mathcal{E}_{t+1}(\epsilon') + \delta\mathcal{U} + \quad (8)$$

$$\max \left[ 0, -\kappa + (1 - \rho)p(\theta)(1 - F(\epsilon')) \left( \mathcal{E}_{t+1}(w(\int_{\epsilon'}^{\bar{\epsilon}} \epsilon f(\epsilon) d\epsilon)) - \mathcal{E}_{t+1}(w(\epsilon')) \right) \right],$$

where  $\delta$  is the exogenous rate at which matches dissolve and separate<sup>13</sup>,  $\mathcal{U}$  denotes the value of being unemployed,  $\epsilon'$  is the match quality of the current job that determines productivity on the current job, and  $\kappa$  are the individual costs of on-the-job search.

### 2.3 Search Behavior of Firms

Firms can either search via posting vacancies registered at the employment office,  $r$ , or through alternative channels, that is posting non-registered vacancies  $n$ . Note that if firms decide to create a vacancy, they have to choose whether to register it or not. Once it is registered, they cannot use other

---

<sup>13</sup>This separation probability is taken to be exogenous in what follows. Endogenizing it along the lines of Mortensen and Pissarides (1994) would be possible. However in the current context we are primarily interested in hiring activity and crowding out effects on this side rather than firing or equilibrium unemployment.

recruitment channels than the employment office to search for applicants.<sup>14</sup> Denoting the value of a filled job for a firm as  $\mathcal{J}$ , the value of a vacancy of type  $i = r, n$  is described by:

$$\rho \mathcal{V}_t^i = -c_i + q_i(\theta) (\mathcal{J}_{t+1} - \mathcal{V}_{t+1}^i) . \quad (10)$$

Note that both types of vacancies create the same jobs. However, the vacancies differ with respect to the costs they cause to the firm. Registered vacancies exhibit a fix cost  $c_r$ . In contrast, non-registered vacancies exhibit costs that on the one hand increase directly with the amount of non-registered vacancies posted in the market,  $\pi$ , and that on the other hand decrease with the intensity of on-the-job search  $\phi$ :  $c_n(\pi, \phi)$  with  $\frac{\partial c_n}{\partial \pi} > 0$  and  $\frac{\partial c_n}{\partial \phi} < 0$ . Intuitively, newspaper adds, headhunters etc. become the more expensive the more firms demand their services, so the costs increase in  $\pi$ . On the other hand, the more firms can expect that some employed workers start thinking about a switch, the lower the costs faced for vacancies posted on alternative channels. Fewer newspaper adds are required as more people read them, fewer headhunters need to be contacted, etc.

The productivity of a filled job is determined by the quality of a match, so firms cannot *ex ante* decide upon the surplus of a job. The expected value of a filled job for a firm is therefore given by the following Bellman equation:

$$\begin{aligned} \mathcal{J}_t(\epsilon) = & \int_0^{\epsilon} (y(\epsilon) - w(\epsilon)) f(\epsilon) d\epsilon + \\ & [1 - \rho - \delta - \phi p(\theta) P(\phi)] \mathcal{J}_{t+1}(\epsilon) + (1 - \rho) [\delta + \phi p(\theta) P(\phi)] \mathcal{V}_{t+1} , \end{aligned} \quad (11)$$

where  $y = \underline{y} + \epsilon$  is the gross product that the job creates to the firm, and  $w$  is the result of the wage bargaining, hence  $y - w$  is the net surplus of the job accruing to the firm, which is strictly increasing in  $\epsilon$ .  $\phi = F(\epsilon^*)$  is the fraction of employed job seekers, and  $P(\phi) = \left[ 1 - F \left( \int_0^{\epsilon^*} \epsilon f(\epsilon) d\epsilon \right) \right]$  is the expected probability that an average employed job seeker finds an acceptable alternative job offer. Intuitively,  $\phi p(\theta) P(\phi)$  is the joint probability that an individual searches on the job, and obtains an alternative offer, which is also acceptable. For notational convenience, we write  $G(\phi) = \phi P(\phi)$ .<sup>15</sup>

<sup>14</sup>Relaxing this assumption would open the possibility that firms find several applicants for one vacancy at the same time, which would unnecessarily complicate the analysis. See also Van Ours (1995) for an alternative view that employers use different recruitment channels for the same vacancy.

<sup>15</sup>Note that  $G(\phi)$  can be re-written in terms the threshold match quality  $\epsilon^*$  as

$$\mathcal{G}(\epsilon^*) = F(\epsilon^*) \left[ 1 - F \left( \int_0^{\epsilon^*} \epsilon f(\epsilon) d\epsilon \right) \right] = F(\epsilon^*) [1 - F(E(\epsilon|\epsilon < \epsilon^*))],$$

## 2.4 Steady State Equilibrium

In steady state, the value of a vacancy of either type for a firm must equal zero, since otherwise it would be profitable to open more vacancies or reduce the number of vacancies, respectively. That is, as a consequence of free entry,  $\mathcal{V}^i = 0$  for  $i = r, n$ . This allows to combine Equations (10) and (11) to get:

$$c_r = q(\theta) \frac{\int_0^{\bar{\epsilon}} (y(\epsilon) - w(\epsilon)) f(\epsilon) d\epsilon}{\rho + \delta + p(\theta)G(\phi)}. \quad (12)$$

Moreover, the free entry condition implies that the costs for both types of vacancies must be equal in steady state, as the benefits  $q(\theta)(\mathcal{J} - \mathcal{V}_i)$  are equal as well. Otherwise, due to the assumption that every applicant can do every job, firms could gain by changing the structure of their vacancies. This implies:

$$c_r = c_n(\theta, \pi). \quad (13)$$

In steady state, by definition, neither the level of unemployment nor the level of total employment changes over time. With the exogenous rate of destruction of employment relationships,  $\delta$ , the former condition is satisfied when inflows into unemployment exactly outweigh outflows from unemployment:

$$\delta(1 - u) = u\theta q(\theta), \quad (14)$$

The latter condition requires that the total outflow from employment into unemployment (through exogenous separation) or directly into new employment (through on-the-job-search) exactly equals the new inflows into employment through all types of vacancies:

$$(1 - u) [\delta + p(\theta)G(\phi)] = (1 + \pi)r q(\theta). \quad (15)$$

The system of the four Equations (12), (13), (14), and (15) determines the behavior of the four unknown variables of the model,  $u, e, r$  and  $n$  (or alternatively  $u, \phi, r$  and  $\pi$ ) in steady state.

We are now prepared to state the main results of the model. The primary aim of the remainder of this section is to develop implications that can be tested empirically. Considering the data that will be available to verify the empirical validity of these implications, it becomes apparent that correlations of the key variables of the model during the transition from one steady state

---

where  $E(\cdot)$  denotes the expectations operator. The behavior of  $G$  with respect to  $\phi$  (or that of  $\mathcal{G}$  with respect to  $\epsilon^*$ , respectively) depends on the steady state distribution of match qualities  $F$ . From  $G \in [0, 1]$ ,  $G(0) = 0$  and  $1 > G(1) = 1 - F(E(\epsilon)) > 0$ ,  $G$  must be increasing at least within some range of the support for low  $\phi$ .

to another are of particular interest. The reason is that the data used for the empirical analysis consist of observations for different occupation-year cells. These observations can be interpreted as realizations of (different) steady states. Therefore, correlations of variables can be obtained using variation between the occupation-year cells, i. e. different steady states. For simplicity, assume in the following that the shock driving the transition from one steady state to another affects the baseline productivity  $\underline{y}$ .<sup>16</sup>

The (first block of) main results of the model is stated in the following proposition:

**Proposition 1.** *As consequence of an unexpected positive (negative) change in the baseline productivity of a match,  $\underline{y}$ ,*

- (i) the number of registered vacancies  $r$  increases (decreases);*
- (ii) the intensity of on-the-job search  $\phi$  increases (decreases);*
- (iii) the number of unregistered vacancies  $n$  increases (decreases);*
- (iv) and the number of unemployed job seekers  $u$  decreases (increases) in equilibrium.*

*Proof.* Ad (i): If  $\underline{y}$  increases, the labor market becomes tighter, as firms post more vacancies in total, since according to Equation (12) the expected benefits of a vacancy outweigh the costs. This is counteracted by a higher labor market tightness  $\theta$ . Note, that from Equation (11) a higher  $\underline{y}$  raises the value of a filled job invariably. Given Equation (10) and the assumption that any applicant can fill a vacancy, the value of a vacancy of either type increases, so more vacancies are registered in order to keep the free entry condition satisfied.

Ad (ii): Following an increase in  $\underline{y}$ , labor market tightness increases, as was shown in (i). However, as a result of Equation (7), this triggers more on-the-job search.

Ad (iii): The third claim follows from condition (13) and the fact that  $c_n$  increases in  $\pi$  and decreases in  $\phi$ : as on-the-job search  $\phi$  increases,  $\pi$  has to increase to guarantee equality of the costs for vacancies of either type.

Ad (iv): Consider the effects of an increase in  $\underline{y}$  on Equation (12): labor market tightness  $\theta$  increases. However, then for condition (14) to hold for a higher value of  $\theta$ , the steady state level of unemployed job seekers  $u$  has to fall. □

---

<sup>16</sup>Strictly speaking, these shocks are unexpected in the sense that at any time  $t$ :  $E_t(\underline{y}_{t+1}) = \underline{y}_t$ , that is the expected future baseline productivity of a match is the same as the current baseline productivity. Note also, that match qualities are irreversible, so once realized, qualities (or their distribution) do not change over time.

Note that these qualitative results do not hinge on the precise form of the steady state distribution of match qualities  $F$ . With respect to the empirical implementation, it is illustrative to summarize these results in form of correlations between the main variables of the model:

**Corollary 1.** *The main variables of the model have the following correlations across different steady states:*

- (a)  $\pi$  and  $r$  are positively correlated;
- (b)  $\phi$  and  $r$  are positively correlated;
- (c)  $\phi$  and  $u$  are negatively correlated;
- (d)  $\pi$  and  $u$  are negatively correlated, and
- (e)  $u$  and  $r$  are negatively correlated.

To see the economic intuition behind the positive correlation between registered and alternative (non-registered) vacancies, think about the consequences of an exogenous positive shock to productivity  $\underline{y}$ . Clearly, in order for Equation (12) to hold, labor market tightness  $\theta$  has to increase, so firms will post more vacancies in steady state. *A priori* they are indifferent between posting  $r$  or  $n$  vacancies. Note, that higher labor market tightness will prompt more on-the-job search due to Equation (7), which in turn leads to more  $n$  vacancies (higher  $\pi$ ) as a result of condition (13). However, it is not possible that the total number of non-registered vacancies is raised while the number of registered vacancies is lowered. To see this, think about a firm considering which vacancy to post immediately after the shock. For  $n$  vacancies to be preferable, it would be necessary that  $\phi$  is higher. However,  $\phi$  cannot increase without more vacancies being posted. Thus, at least some firms will register their additional vacancies after the shock, which prompts more on-the-job search and makes non-registered vacancies more attractive. If the additional vacancies would be posted through alternative channels instead of registration at the employment office, the adjustment to the new steady state would have to rely on some sort of self-fulfilling prophecy, as firms would post more  $n$  vacancies hoping to create sufficient on-the-job search. But even then, they would not decrease the number of registered vacancies at the outset and create new  $n$  instead of the  $r$  vacancies, since the costs of  $n$  vacancies increase in their number (i. e. in  $\pi$ ).

Considering the correlation between the intensity of employed job search and the level of registered vacancies, it becomes clear from this result that on-the-job search is actually initiated by more registered vacancies. Moreover,

there are no restrictions for employed job seekers as to which vacancies they can fill. On the other hand, intensified search decreases labor market tightness and therefore leads firms to post more vacancies, including registered ones, in order to maintain the equilibrium labor market tightness, hence the positive correlation between  $\phi$  and  $r$ .

Next, consider the effects of a shock to productivity by examining Equation (12), and neglect for the moment that there is on-the-job search (that is the term  $\phi p(\theta)$  in the denominator). By applying the implicit function theorem, it is easy to show that, in steady state, an increase in productivity leads *ceteris paribus* to a higher level of market tightness  $\theta$ . Noting that  $\theta q(\theta) = p(\theta)$  by Equation (5), this leads to more inflows into employment, so that for the steady state condition (14) to be satisfied, the level of unemployment has to decrease. Now make the same experiment taking on-the-job search into account. Obviously, since the denominator is larger and increases in  $\theta$ , labor market tightness has to increase by less than in the previous case without on-the-job search, and unemployment decreases by less. However, it is clear that the fraction of employed job searchers increases while the pool of unemployed job searchers decreases after a positive shock to productivity. Thus, we have a negative correlation between  $\phi$  and  $u$ . The result is interesting since it provides a prediction on crowding-out effects: While more on-the-job search undeniably increases competition for vacancies, the net effect is a 'crowding-in' rather than crowding out, since in steady state equilibrium more employed job search is associated with lower unemployment levels.

The results presented so far illustrate the workings of the model, and will be confronted with data below. However, endogenous on-the-job search behavior has two further effects: Job competition as a result of endogenous search leads to a proportional crowding-out of unemployed job seekers in the hiring process.<sup>17</sup> As opposed to the 'crowding-in' result of Proposition 1, according to which more on-the-job search is associated with fewer unemployed job seekers, and, due to the constant search intensity, lower levels of unemployment, the *fraction* of job accessions won by unemployed *decreases* as on-the-job search increases. Another result is even more interesting: Quite intuitively, the average starting wage of previously employed newly hired workers is higher than the average starting wage of unemployed applicants. However, when the number of employed job seekers goes up, and thus their total number of new matches as well as the share of new hires made up for by employed applicants increase, this differential in average starting wages *increases*. That is, contrary to what one would expect, a higher number of

---

<sup>17</sup>This result has also been shown by Burgess (1993) and Anderson and Burgess (2000).

employed applicants drives up their premium in starting wages. This increase in the wage premium is entirely supply-side driven and not induced by any sort of ranking or increase in the relative demand for employed applicants. These two results are summarized in the following proposition:

**Proposition 2.** *An increase (decrease) in on-the-job search intensity,  $\phi$ , following an unexpected positive (negative) productivity shock to  $\underline{y}$ , decreases (increases) the fraction of job accessions by unemployed applicants, and increases (decreases) the differential in average starting wages between formerly employed and formerly unemployed recruits.*

*Proof.* To proof the first claim, write job accessions by employed applicants as  $(1 - u)p(\theta)G(\phi)$ . Likewise, job accessions by unemployed applicants are  $up(\theta)$ , and total hirings are the sum of these two expressions. However, careful observation reveals that an increase in  $\theta$  following a rise in  $\underline{y}$ , and the resulting surge in on-the-job search  $\phi$  increases the job accessions of employed by more than those of unemployed, since both terms increase equally in  $\theta$ , but the former additionally increases in  $\phi$ . The opposite is true for a negative productivity shock.

The second claim follows from the search behavior of employed workers. As was seen in Equation (6), the threshold match quality below which individuals search for alternative employment increases if labor market conditions become more favorable, that is if  $\theta$  increases. However, this means that those who start searching as a consequence of a shock to baseline productivity  $\underline{y}$  forego a comparatively better job than those who already searched before the shock. Of course, they are only prepared to accept a job offering a better match than their current one, so *on average* the quality of matches won by employed job seekers has to increase as a consequence. Since higher match quality directly translates into higher wages due to the fix sharing rule between workers and firms, average starting wages of previously employed recruits increase as more 'picky' applicants are in the pool of employed job seekers. On the other hand, there is no change in the average starting wage of previously unemployed hires following a shock to  $\underline{y}$ . Hence, the differential widens after a positive shock, and reduces after a negative shock.  $\square$

The last claim is interesting on its own. It reflects not only a testable implication of the basic motivation for on-the-job search in the model, but it also highlights a novel insight the model can provide to the discussion about the effects of search frictions and job creation on wages. A growing body of research is devoted on equilibrium wage dispersion in matching models with endogenous on-the-job search, see Postel-Vinay and Robin (2002) for a recent example. Another branch of the literature is concerned about the

effect of (net) job creation on the level of wages, particularly entry wages (see Belzil (2000)). However, there are no results to date investigating the cyclicity (with respect to job creation) of the *premium* in starting wages earned by formerly employed over formerly unemployed workers. The model presented here provides a framework that allows to make testable predictions about how this premium should be correlated with other important variables, which complements the existing literature in this respect. These predictions will be confronted with the data below.

### 3 The Empirical Relevance of Unobserved Endogenous Search Behavior

Below in section 3.2 it will be shown how the empirical content of the predictions of the model presented in the previous section can be tested. But before that, we briefly address the effects of neglecting on-the-job search and the use of alternative recruitment channels in empirical estimations of the matching function.

#### 3.1 A Common Problem With Empirical Matching Studies

As was mentioned earlier, the problem of empirical studies of the matching function is that endogenous search behavior, which lies at the heart of the theoretical model presented in the previous section, is, at least partly, unobservable by the researcher. This section presents the consequences of neglecting unobservable endogenous determinants for empirical estimates of the matching function in terms of biased and inconsistent results.

To be specific, we assume in the following that both the number of workers searching on-the-job search,  $e$ , and the number of non-registered vacancies,  $n$ , are not observable. On the other hand, the pools of unemployed job seekers,  $u$ , registered vacancies  $r$  and the number of successful matches,  $m$ , are observable. We introduce the following notation for the shares of the respective groups of searchers and vacancies in the overall pools:

$$\tau := \frac{u}{u + e}, \text{ and hence } (1 - \tau) = \frac{e}{u + e}, \quad (16)$$

and

$$\sigma := \frac{r}{r + n}, \text{ and hence } (1 - \sigma) = \frac{n}{r + n}. \quad (17)$$

Note also the relations  $\tau = \frac{u}{u+(1-u)\phi}$ ,  $\sigma = \frac{1}{1+\pi}$  and  $(1 - \sigma) = \frac{\pi}{1+\pi}$ . Moreover, denote the inverse of the shares of the observable stocks of job searchers and vacancies by:

$$\tilde{\phi} := \frac{1}{\tau} \text{ and } \tilde{\pi} := \frac{1}{\sigma} \quad (18)$$

Note that since  $\phi \in [0, 1]$ , we have  $\tilde{\phi} \in [1, \frac{1}{u}]$  which implies that for  $\tilde{\phi} = 1 \Leftrightarrow \tau = 1$  there is no on-the-job search ( $\phi = 0$ ), and that for  $\tilde{\phi} = \frac{1}{u} \Leftrightarrow \tau = u$  everybody in the labor force actively searches for a new job ( $\phi = 1$ ). Of course,  $\tilde{\phi}$  and  $\tilde{\pi}$  are also both determined endogenously. For example, from Equation (7) it follows that if  $\frac{\partial \phi}{\partial \theta} > 0 \Leftrightarrow \frac{\partial \tilde{\phi}}{\partial \theta} > 0$ .

Using this notation, labor market tightness as given by Equation (3) can alternatively be expressed as  $\theta = \frac{r+n}{u+e} = \frac{(1+\pi)r}{u+(1-u)\phi} = \frac{\tilde{\pi}r}{\tilde{\phi}u}$ . The matching function introduced in Equation 2 can then be stated as:

$$m = m(u + e, r + n) = m(u + (1 - u)\phi, (1 + \pi)r) = m(\tilde{\phi}u, \tilde{\pi}r). \quad (19)$$

Obviously, when one estimates a matching function of the sort of Equation (19) using only  $u$  and  $r$  as explanatory variables, variation in the unobserved parts of the pools of job searchers and vacancies leads to variation in observed matches. To see the argument more formally, consider a fully parameterized version of the matching function satisfying all assumptions made in the previous section. In particular, consider a Cobb-Douglas type matching function as is done in most of the empirical literature:<sup>18</sup>

$$m_t = AJ_t^\alpha V_t^\beta, \quad (20)$$

where  $m_t$  is the number of new hires between period  $t$  and  $t + 1$ ,  $A$  is total matching efficiency, and  $J$  and  $V$  are the numbers of job seekers and vacancies at period  $t$ , respectively. One is interested in estimates of the matching elasticities  $\alpha$  and  $\beta$ . Thus, the conventionally estimated empirical matching function has the form:

$$\ln m_t = \ln A + \alpha \ln J_t + \beta \ln V_t + \varepsilon_t, \quad (21)$$

with  $\varepsilon$  being a normally distributed error term:  $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ . However, instead of observing  $J$  and  $V$ , virtually all empirical studies have only data about unemployed job seekers  $u$  and registered vacancies (or vacancies contained in a help-wanted index)  $r$ .

Remember that  $J = u + e = \tilde{\phi}u$  and  $V = r + n = \tilde{\pi}r$ , and consider the assumptions that all applicants can take on any job and that firms have no

<sup>18</sup>See Warren (1996), Yashiv (2000), and Fahr and Sunde (2001b) for more general specifications.

preferences for applicants with a certain labor market status. This is reflected in identical elasticities of the matching function with respect to unemployed and employed job seekers ( $\alpha$ ) and registered and non-registered vacancies ( $\beta$ ). Hence, neglecting time indices, Equation (21) can be decomposed into observable and unobservable components:

$$\ln m = \ln A + \alpha \ln \tilde{\phi} + \alpha \ln u + \beta \ln \tilde{\pi} + \beta \ln r + \varepsilon. \quad (22)$$

Both,  $\tilde{\phi}$  and  $\tilde{\pi}$  are unobservable. Consequently, estimating the matching function conventionally, that is with only  $u$  and  $r$  as explanatory variables, amounts to omit relevant variables in the regression leading to a bias in the estimated elasticities. As is shown in more detail in Appendix A using OLS one obtains the following estimates for the elasticities:<sup>19</sup>

$$\hat{\alpha} = \alpha + \frac{\alpha \text{cov}(\ln u, \ln \tilde{\phi})}{\text{var}(\ln u)} + \frac{\beta \text{cov}(\ln u, \ln \tilde{\pi})}{\text{var}(\ln u)}, \quad (23)$$

and

$$\hat{\beta} = \beta + \frac{\alpha \text{cov}(\ln r, \ln \tilde{\phi})}{\text{var}(\ln r)} + \frac{\beta \text{cov}(\ln r, \ln \tilde{\pi})}{\text{var}(\ln r)}. \quad (24)$$

Obviously, the estimates obtained by the standard procedure do not deliver the coefficients of interest. Moreover, it is not clear *ex ante* in which direction the bias goes. However, the theoretical model presented before can give some guidance as to whether one should expect the coefficient estimates to suffer from upward or downward bias. This is done next, before setting out a strategy that allows to obtain unbiased estimates of the variables of interest.

One additional remark is in order. The result that empirical studies neglecting endogenous unobservable on-the-job search and vacancy posting deliver biased estimates is not limited to studies estimating conventional U-V-type matching functions. Also studies exploiting micro data on unemployment duration (see Petrongolo (2001) for a recent example) and studies emphasizing the stock flow approach (like Coles and Smith (1998) and Coles and Petrongolo (2002)) are affected by the bias due to omitting unobservable variables. These studies do not explicitly account and control for changes in the composition of the pools of job searchers competing for a new job. Neither do these alternative approaches reflect changes in the composition of vacancies actually available (but partly not observable) which affect labor market tightness and therefore individual hazard rates.<sup>20</sup> While so far, these

<sup>19</sup>Appendix A.1 contains a more general version taking indirect effects into account.

<sup>20</sup>As for the stock-flow approach, new inflows into the pools of vacancies and active job seekers are measured inaccurately by concentrating only on inflows into unemployment

shortcomings might have been treated as measurement problems related to the data, the model presented above gives a clear indication as to which distortions are to be expected from disregarding systematic endogenous behavior in empirical work.

### 3.2 Testable Implications of the Model

From inspection of Equations (23) and (24) it becomes clear that there are four correlations that have to be known before one can make statements about the true parameters of the model. Before it is shown how these correlations can be obtained empirically, we briefly consult the theoretical model about indications of their direction. One has to keep in mind, however, that the data used for the empirical evaluation of the model are obtained for certain reference dates only, whereas the model is in continuous time. In what follows, the observations in the data are therefore interpreted as realizations of steady state equilibria.

Consider first the estimate of  $\hat{\alpha}$  in Equation (23). From the assumptions about the matching function,  $\alpha$  and  $\beta$  should be positive, and the variance of  $\ln u$  is positive by definition. Therefore, one is primarily interested in the covariances. The theoretical model would predict that  $\ln u$  and  $\ln \phi$  should be negatively correlated. This result can be obtained more indirectly by noting that  $\tilde{\phi} = \frac{1}{\tau}$ . Therefore, one would expect that  $\frac{d\tilde{\phi}}{du} \sim -\frac{d\tau}{du}$ . But from the definition of  $\tau$  in Equation (16),  $\frac{d\tau}{du} = \frac{\partial\tau}{\partial u} + \frac{\partial\tau}{\partial\phi} \frac{\partial\phi}{\partial u}$ . Taking derivatives, one can see that  $\frac{\partial\tau}{\partial u}$  is positive, while  $\frac{\partial\tau}{\partial\phi}$  is negative. However, from the Corollary 1 we know that  $\phi$  and  $u$  are negatively correlated, so the entire second term is positive. Thus, from the theoretical model one indeed expects  $cov(\ln u, \ln \tilde{\phi})$  to be negative.

Analogously, according to the prediction of the model,  $cov(\ln u, \ln \tilde{\pi})$  should be negative as well. Following a similar argument as before,  $\tilde{\pi} = \frac{1}{\sigma}$ , and therefore  $\frac{d\tilde{\pi}}{du} \sim -\frac{d\sigma}{du}$ . Again,  $\frac{d\sigma}{du} = \frac{\partial\sigma}{\partial\pi} \frac{\partial\pi}{\partial u} + \frac{\partial\sigma}{\partial r} \frac{\partial r}{\partial u}$ . Taking derivatives reveals that  $\frac{\partial\sigma}{\partial\pi} < 0$  and that  $\frac{\partial\sigma}{\partial r} = 0$ . Thus, from Corollary 1, one finds that the first term is negative, while the second term equals zero.

Consequently, the theoretical model predicts that the conventional estimation approach leads to downwardly biased estimates of the matching

---

and registered vacancies, as is done in the studies mentioned in the text. However, while showing that the results obtained this way are biased is straightforward, correcting for the bias in a stock-flow framework requires more specific information and assumptions on the quantitative effects of unobserved endogenous behavior. Since the data used below do not contain detailed flow information, we proceed by presenting empirical results and a correction strategy for the conventional matching approach.

elasticity with respect to the pool of searchers:

$$\hat{\alpha} < \alpha . \tag{25}$$

The downward bias is a result of neglecting unobserved variation both in the stock of job seekers and vacancies. Both, more on-the-job search and more search of firms through non-registered vacancies are associated with lower unemployment levels, whose coefficient therefore underestimates the true effect of the pool of job seekers on employment flows.

Next, turn to the estimate of  $\hat{\beta}$  from Equation (24). The correlation between registered vacancies and the inverse of the share of unemployed job seekers,  $cov(\ln r, \ln \tilde{\phi})$ , should be positive according to the model. As before, this can be seen by noting that  $\frac{d\tilde{\phi}}{dr} \sim -\frac{d\tau}{dr}$  and decomposing  $\frac{d\tau}{dr} = \frac{\partial\tau}{\partial\phi} \frac{\partial\phi}{\partial r} + \frac{\partial\tau}{\partial u} \frac{\partial u}{\partial r}$ . The first term is negative, since  $\frac{\partial\tau}{\partial\phi} < 0$  and from Corollary 1. The second term is negative as  $\frac{\partial\tau}{\partial u} > 0$  and by the negative correlation of  $u$  and  $r$  from Corollary 1. Concluding from the theoretical model one would therefore expect that the correlation between  $\tilde{\phi}$  and  $r$  is positive.

From the discussion about vacancy posting by firms, one would expect  $cov(\ln r, \ln \tilde{\pi})$  to be positive as well. More precisely, since  $\frac{d\tilde{\pi}}{dr} \sim -\frac{d\sigma}{dr}$ , we examine  $\frac{d\sigma}{dr} = \frac{\partial\sigma}{\partial r} + \frac{\partial\sigma}{\partial\pi} \frac{\partial\pi}{\partial r}$ . As  $\frac{\partial\sigma}{\partial r} = 0$  and  $\frac{\partial\sigma}{\partial\pi} < 0$ , and from the positive correlation between  $\pi$  and  $r$  in Corollary 1, the model indeed predicts  $r$  and  $\tilde{\pi}$  to be correlated positively.

Therefore, the estimate of the elasticity of matches with respect to vacancies obtained in conventional empirical matching functions is expected to be upward biased:

$$\hat{\beta} > \beta . \tag{26}$$

This bias reflects the fact that the unobserved variation in non-registered vacancies is stronger than the variation in registered vacancies, while both move into the same direction. Given these testable implications, we now turn to their empirical evaluation.

## 4 Empirical Application

The purpose of this section is twofold. On the one hand, it develops a way for correcting the estimates of the elasticities of the empirical matching function, which allows to impute the "correct" unbiased coefficients. Secondly, this allows to test the predictions of the theoretical model for their empirical content. The fundamental assumption underlying the following analysis is that there is only one matching function on the labor market with the same stable parameters for all types of searchers and vacancies.

## 4.1 True Matching Elasticities

The empirical strategy proposed here exploits the fact that the data allow to identify several concepts of new matches, that is, several alternative measures for the flows used as the dependent variable. In particular, in the data, which are described in detail in Fahr and Sunde (2001b), we can distinguish all successful matches within a given period of time,  $m$ , from successful matches of formerly unemployed persons,  $m_u$ , and from successful matches resulting from registered vacancies,  $m_r$ . The basic idea is that matches from unemployment  $m_u$  or from registered vacancies  $m_r$  as dependent variables lead to different biases in the parameter estimates than the bias obtained by regressing  $m$ . In particular, consider a matching function for  $m_u$ :

$$\ln m_u = \ln A_u + \alpha_u \ln u + \beta_u \ln r + \beta_u \ln \tilde{\pi} + \varepsilon_1. \quad (27)$$

By definition, only unemployed job seekers can explain flows from unemployment into employment, employed job seekers do not have to be considered. The respective estimates of the elasticities obtained by OLS are:

$$\hat{\alpha}_u = \alpha_u + \frac{\alpha_u \text{cov}(\ln u, \ln \tilde{\pi})}{\text{var}(\ln u)}, \quad (28)$$

and

$$\hat{\beta}_u = \beta_u + \frac{\beta_u \text{cov}(\ln r, \ln \tilde{\pi})}{\text{var}(\ln r)}. \quad (29)$$

Analogously, a matching function for  $m_r$  can be written as:

$$\ln m_r = \ln A_r + \alpha_r \ln u + \alpha_r \ln \tilde{\phi} + \beta_r \ln r + \varepsilon_2. \quad (30)$$

Note that only registered vacancies, not non-registered vacancies, are to be included as explanatory variable for the stock of vacancies. This leads to the following OLS estimates of the matching elasticities:

$$\hat{\alpha}_r = \alpha_r + \frac{\alpha_r \text{cov}(\ln u, \ln \tilde{\phi})}{\text{var}(\ln u)}, \quad (31)$$

and

$$\hat{\beta}_r = \beta_r + \frac{\beta_r \text{cov}(\ln r, \ln \tilde{\phi})}{\text{var}(\ln r)}. \quad (32)$$

By assumption, the true matching elasticities characterizing these alternative specifications of the matching function are the same as in the matching function for all new matches presented in Equation (22). Therefore,

$\alpha = \alpha_u = \alpha_r$  and  $\beta = \beta_u = \beta_r$ .<sup>21</sup> Estimates of all coefficients  $\hat{\alpha}$ ,  $\hat{\alpha}_u$ ,  $\hat{\alpha}_r$ ,  $\hat{\beta}$ ,  $\hat{\beta}_u$ , and  $\hat{\beta}_r$  can be obtained from the data, but all of these estimates are biased estimates of the true parameters  $\alpha$  and  $\beta$ . Taking a closer look at the structure of the biases reveals, however, that it is possible to impute the true parameter from the biased estimates. For example, one can obtain the bias stemming from neglecting employed job seekers in the estimation simply by subtracting the respective coefficient estimates:

$$\hat{\alpha} - \hat{\alpha}_u = \frac{\alpha \text{cov}(\ln u, \ln \tilde{\phi})}{\text{var}(\ln u)}. \quad (33)$$

Analogously, one can isolate the remaining elements of the bias:

$$\hat{\alpha} - \hat{\alpha}_r = \frac{\alpha \text{cov}(\ln u, \ln \tilde{\pi})}{\text{var}(\ln u)}, \quad (34)$$

$$\hat{\beta} - \hat{\beta}_u = \frac{\beta \text{cov}(\ln r, \ln \tilde{\phi})}{\text{var}(\ln r)}, \quad (35)$$

$$\hat{\beta} - \hat{\beta}_r = \frac{\beta \text{cov}(\ln r, \ln \tilde{\pi})}{\text{var}(\ln r)}. \quad (36)$$

Using these estimates of the biases from estimates of matching functions with all hirings ( $m$ ), hirings from unemployment ( $m_u$ ), and hirings from registered vacancies ( $m_r$ ) as dependent variables, the correct elasticities of the matching function can therefore be calculated. This is done below.

## 4.2 Empirical Relevance

The results so far not only imply that the true unemployment elasticity of matching  $\alpha$  is underestimated by  $\hat{\alpha}$ , but also that this bias is larger than the bias obtained when using hirings from unemployment as dependent variable, i. e.  $\hat{\alpha}_u > \hat{\alpha}$  when  $\text{cov}(\ln u, \ln \tilde{\pi})$  as implied by the model. This is consistent with previous findings in the literature: studies using all hires as dependent variable invariably found systematically lower unemployment elasticities than studies using outflows from unemployment, see Broersma and Ours (1999) for an overview. However, instead of relying on such indirect evidence, we now test the validity of the model implications directly.

---

<sup>21</sup>Note that this is essentially the same as assuming that the fraction of new hires made up by unemployed applicants,  $\frac{m_u}{m}$  is proportional to the fraction of job seekers who are unemployed,  $\tau$ . Likewise, the same holds for vacancies, i. e.  $\frac{m_r}{m} \sim \sigma$  etc. Broersma and Ours (1999) make the same assumption.

Due to the fact that all estimates used for the calculations are available, it is straightforward to use them to compute the desired estimates of the deep parameters of the matching function. Moreover, the calculation of the different biases allows to test the predictions of the theoretical model presented in the previous section.

Consider Table 1, which is based on German data described in detail in Fahr and Sunde (2001a,b).<sup>22</sup> The table contains the estimates of empirical matching functions of all the specifications needed for the computation of the true elasticities.<sup>23</sup>

Using the previous results, the imputed components of the empirical biases are:

$$\begin{aligned}\hat{\alpha} - \hat{\alpha}_u &= -0.005, \\ \hat{\alpha} - \hat{\alpha}_r &= -0.191, \\ \hat{\beta} - \hat{\beta}_u &= 0.153, \text{ and} \\ \hat{\beta} - \hat{\beta}_r &= 0.133.\end{aligned}$$

Consequently, the desired coefficients of the matching function are:

$$\alpha = 0.616, \text{ and } \beta = 0.162.$$

Equipped with these numbers, the correlations that allow to test the

---

<sup>22</sup>The data for stocks of unemployed and registered vacancies contain yearly observations from 1980 through 1995 for 40 occupational groups, and are taken from official labor market statistics. While the data on vacancies, unemployment levels and the like are available on the occupational level, observations for matches ( $m$ ,  $m_u$ ,  $m_r$ , and matches from non-employment  $m_x$  which are included for robustness) are calculated from German social security records on the individual level. For a detailed description of the data and results for different levels of disaggregation, see Fahr and Sunde (2001b).

<sup>23</sup>Results for the constant term (total matching efficiency) are not discussed for brevity, see Lindeboom, Ours, and Renes (1994) for an interpretation as the inverse of the speed of matching. Note, however, that there is a potential problem of multicollinearity if  $u$  and  $r$  are highly correlated. A (negative) correlation is predicted by the model, see Proposition 1. Indeed, controlling for systematic differences across different occupations,  $u$  and  $r$  are negatively correlated in the data. However, the symptoms of multicollinearity are weak, in particular both explanatory variables have highly significant coefficients and plausible values comparable to the existing literature, so we follow the literature and neglect the issue of multicollinearity. See also Appendix A.1.

implications from the theoretical model can be computed:

$$\begin{aligned} \frac{\text{cov}(\ln u, \ln \tilde{\phi})}{\text{var}(\ln u)} &= -0.008, \\ \frac{\text{cov}(\ln u, \ln \tilde{\pi})}{\text{var}(\ln u)} &= -0.310, \\ \frac{\text{cov}(\ln r, \ln \tilde{\phi})}{\text{var}(\ln r)} &= 0.944, \text{ and} \\ \frac{\text{cov}(\ln r, \ln \tilde{\pi})}{\text{var}(\ln r)} &= 0.821. \end{aligned}$$

For statistical inference, also the variance of the coefficients has to be computed. However, this is not a straightforward matter, because the empirical error variance obtained from the estimated model is also biased as a result from omitting (unobservable) variables in the estimation. In Appendix A.2, two alternative ways of imputing an unbiased estimate of the variance of the errors are presented. The first alternative is conservative in the sense that it deliberately overestimates the error variance, thus leading to actual significance levels that are higher than those indicated by conventional critical values. The second alternative allows to compute the error variance accurately, however imposing more restrictive assumptions. Standard errors of the coefficient estimates using both methods are presented in Table 1.<sup>24</sup>

Taking flows from non-employment  $m_x$  instead of flows from unemployment as dependent variable, which arguably contains less measurement error in the flow data used for the estimation, one obtains for the biases:  $\hat{\alpha} - \hat{\alpha}_u = 0.009$ ,  $\hat{\alpha} - \hat{\alpha}_r = -0.191$ ,  $\hat{\beta} - \hat{\beta}_u = 0.040$ , and  $\hat{\beta} - \hat{\beta}_r = 0.133$ . The matching elasticities in this case are then calculated as  $\alpha = 0.602$  and  $\beta = 0.303$ .<sup>25</sup>

Obviously, all empirical correlations and effects are as predicted by the theoretical model. The only exception is the positive correlation between the stock of unemployed and on-the-job search with matches from non-employment as dependent variable instead of matches from unemployment. A possible explanation for this is that  $m_x$ , unlike  $m_u$ , depends also on other seekers than the registered unemployed explicitly contained in the estimation. The variation in unemployed seekers might then be not sufficient to explain the entire variation in  $m_x$ , leading to a somewhat downward biased estimate of the respective elasticity  $\hat{\alpha}_u$ . Thus, the calculated bias  $\hat{\alpha} - \hat{\alpha}_u$  is

<sup>24</sup>Including a linear time trend did not change the results qualitatively. Time trends were found to be significant and negative.

<sup>25</sup>Again, adding time trends did not alter the results.

**TABLE 1: EMPIRICAL AGGREGATE MATCHING FUNCTIONS BY SOURCES OF FLOWS AND IMPUTED UNBIASED ELASTICITIES**

**Dependent Variable: Hirings (per occupation and year) by sources**

	(1) total	(2) from un- employment	(3) from non- employment	(4) from registered vacancies	(A) Imputed Elasticities and standard errors using (1),(2),(4)	(B) Imputed Elasticities and standard errors using (1),(3),(4)
	$m$	$m_u$	$m_x$	$m_r$		
$\ln u^a$	0.420 (0.014)	0.425 (0.022)	0.411 (0.016)	0.611 (0.029)	0.616 (0.055)* (0.038)**	0.602 (0.029)* (0.029)**
$\ln r^b$	0.448 (0.012)	0.295 (0.018)	0.436 (0.013)	0.315 (0.024)	0.162 (0.045)* (0.031)**	0.303 (0.024)* (0.024)**
Const.	3.024 (0.197)	2.766 (0.346)	2.857 (0.233)	1.358 (0.267)		
$R^2$	0.854	0.621	0.821	0.614		
N	640	640	640	640		
e'e	0.158	0.371	0.191	0.619		

Note: Source and data description: Fahr and Sunde (2001a). Data are of yearly frequency for 1980-1995, the level of aggregation is occupation (40 groups). Robust standard errors are in parentheses.

\* Standard errors computed using Alternative I (conservative approach) presented in Appendix A2.

\*\* Standard errors computed using Alternative II (exact approach) presented in Appendix A2.

<sup>a</sup> Var (ln u) = 1.187.

<sup>b</sup> Var (ln r) = 1.744.

too large when  $m_x$  are used as dependent variable.<sup>26</sup> However, using either  $m_x$  or  $m_u$ , the estimates of the elasticity of matches with respect to job seekers are downward biased, while the true elasticity with respect to vacancies is lower than the estimate suggests, precisely as was predicted by the model.

Finally, it remains to empirically verify the predictions of Proposition 2. We check the first claim by regressing the log of the fraction of new hirings accounted for by previously unemployed applicants on some proxy for baseline productivity. Since  $\underline{y}$  is unobserved, we resort to using registered vacancies as an instrument: from Proposition 1 and Corollary 1 registered vacancies are (positively) correlated with  $\underline{y}$ , and, by assumption, there is no ranking or any other restriction to job accessions by unemployed or employed job seekers. Hence, rank and order conditions for an instrument are satisfied by registered vacancies. According to the claim, the elasticity of the fraction of unemployed job accessions with respect to registered vacancies should be negative. The regression results are contained in Table 2. It is obvious that there is evidence for relative crowding out as predicted by Proposition

<sup>26</sup>An alternative explanation is that in times of high unemployment more employed workers start searching for a new job as a consequence of more intensive use of advance notice of layoffs, c. f. Garibaldi (1998). This effect, while present in the estimation for matches from unemployment, might be even stronger for matches from non-employment.

**TABLE 2: EVIDENCE FOR RELATIVE CROWDING-OUT OF UNEMPLOYED JOB SEEKERS BY ON-THE-JOB SEARCHERS**

**Dependent Variable: Fraction of Hirings (per occupation and year) from Unemployment**  
 **$\ln(m_u/m)$**

	$\ln(m_u/m)$				
<b><math>\ln r</math></b>	-0.151 (0.014)	-0.154 (0.014)	-0.154 (0.018)	-0.156 (0.019)	-0.149 (0.013)
<b><math>\ln u</math></b>			0.005 (0.025)	0.003 (0.025)	
Const.	-0.228 (0.114)	-0.753 (0.324)	-0.258 (0.201)	-0.770 (0.363)	
Linear time trend		0.006 (0.004)		0.006 (0.004)	-2.638 (0.303)
Occupation Dummies	no	no	no	no	yes
$R^2$	0.179	0.183	0.179	0.183	0.895
N	640	640	640	640	640

Note: Source and data description: Fahr and Sunde (2001a). Data are of yearly frequency for 1980-1995, the level of aggregation is occupation (40 groups). Robust standard errors are in parentheses.

2, and that this finding is robust against different specifications. Adding a time trend does not affect the results at all, while adding the number of unemployed seekers is insignificant, suggesting that variation in the share of matches won by unemployed applicants is mainly driven by the endogenous job competition mechanism described in the model.<sup>27</sup>

Next, we test the claim that the premium in terms of average starting wages on a new job, which previously employed recruits receive as compared to unemployed recruits, increases in the number of employed job searchers. In order to do so, we regress the logged ratio of average starting wages of employed over unemployed hirings on the logarithm of the share of matches won by employed job seekers. This estimation setting appears to be the most direct test of the hypothesis: as was shown in the first part of Proposition 2, the share of job accessions of employed applicants increases as the number of on-the-job searchers increases, while the share of matches of unemployed seekers decreases. If the claim is true, the wage premium of employed hires over unemployed hires should increase as on-the-job search intensity, and thus the share of employed job matches increases. The results of different specifications of this setting are presented in Table 3. The data strongly support the claim made in the proposition. Note that evidence in favor of

<sup>27</sup>The results provide evidence that hiring firms rank employed applicants before unemployed applicants, in the sense that if a vacancy is applied for simultaneously by an unemployed and an employed applicant, the job is offered to the latter. Intuitively, if firms rank applicants, the fraction of matches of unemployed seekers entirely depends on the behavior of the employed, since they directly affect the probability of finding a new job. See also Anderson and Burgess (2000).

**TABLE 3: THE EMPLOYED JOB SEARCHER PREMIUM IN AVERAGE STARTING WAGES AND ON-THE-JOB SEARCH**

Dependent Variable: Average Starting Wage Markup of Employed Searchers				
	$\ln(w_E/w_U)$			
$\ln(m_E/m)$	0.075 (0.029)	0.068 (0.028)	0.079 (0.029)	0.074 (0.028)
$\ln u$			-0.005 (0.005)	0.008 (0.005)
Const.	0.198 (0.029)	-0.338 (0.103)	0.252 (0.063)	-0.267 (0.105)
Linear time trend		0.006 (0.001)		0.006 (0.001)
$R^2$	0.073	0.121	0.075	0.125
N	640	640	640	640

Note: Source and data description: Fahr and Sunde (2001a). Data are of yearly frequency for 1980-1995, the level of aggregation is occupation (40 groups). Wage data are deflated daily wages collected at individual levels from social security records, aggregated to cell means of occupation-year cells. Robust standard errors are in parentheses.

the proposition also indirectly supports the microfoundation for employed job search: if more employed workers search, more of them will find a job increasing the share of hirings from the pool of the employed. On the other hand, the more employed search, the higher the average quality of their current match as a consequence of the non-degenerate distribution of job match quality. Hence, they require, on average, a higher starting wage to accept a job offer, leading to a supply-induced increase in the wage premium of formerly employed recruits.

To conclude, the main results of the empirical evaluation of the implications of the theoretical model are strongly supportive of the model. All empirical correlations, with the one exception mentioned, exhibit the sign predicted by the theoretical model, indicating that the theoretical model presented in section 2 indeed describes empirically relevant mechanisms. This conclusion is corroborated by the findings of empirical tests of Proposition 2: Substantive endogenous job competition seems to shape the structure of job accessions, and the wage premium earned by employed applicants when entering a new job increases in the number of employed job seekers.

## 5 Conclusion

This paper has shown that previous empirical studies of matching models of job creation do not deliver the correct elasticities of the matching function. In particular, empirical estimations of conventional matching functions regressing employment flows on stocks of unemployed searchers and registered

vacancies, as well as estimations of stock-flow matching functions, all suffer from the problem that important determinants of the matching process, like employed job seekers or alternative search channels, are unobservable and thus not included as regressors.

In order to investigate this issue, we have developed a simple model of equilibrium unemployment and vacancy posting that takes unobservable endogenous behavior on both sides of the labor market explicitly into account. The implications of the model with respect to the effects of endogenous on-the-job search and vacancy posting on the composition of the pools of seekers and job vacancies can predict the direction of the bias pervading previous empirical studies. Moreover, the theory makes predictions about the structure of job accessions with respect to the employment status of applicants, and about the cyclicity of starting wages and wage premia for employed applicants.

In the second part of the paper, it was shown how the bias can be removed and estimates of the true parameters of the matching function can be retrieved. In an empirical application using German administrative data, it is shown that, in line with the predictions of the theoretical model, conventional empirical studies tend to underestimate the searcher-elasticity of new hirings and to overestimate the vacancy-elasticity. Also other implications of the theoretical model are supported by the data: various correlations implied by the model are validated empirically, job competition seems to play an important role, and the supply-induced cyclicity of the premium in average starting wages of employed job accessions over unemployed job accessions provides strong evidence for the theoretical motivation for on-the-job search based on match quality.

The paper poses a strong caveat on the use of conventional empirical matching functions, as well as alternative approaches based on hazard functions or stock-flow notions, investigating the job creation process. The model allows to put the interpretation of previous results obtained by using such conventional or alternative approaches into perspective. This applies particularly for studies trying to test for constant returns in the matching process while neglecting important but unobservable endogenous behavior like on-the-job search or search of firms through several recruitment channels.

## References

- ALBRECHT, J., P. GAUTIER, AND S. VROMAN (2002): “Matching With Multiple Applications,” *mimeo*.
- ANDERSON, P. M., AND S. M. BURGESS (2000): “Empirical Matching Functions: Estimation and Interpretation Using State-Level Data,” *Review of Economics and Statistics*, 82(1), 90–102.
- BELZIL, C. (2000): “Job Creation and Job Destruction, Worker Reallocation, and Wages,” *Journal of Labor Economics*, 18(2), 183–203.
- BLANCHARD, O. J., AND P. DIAMOND (1994): “Ranking, Unemployment Duration, and Wages,” *Review of Economic Studies*, 61(3), 417–437.
- BOERI, T. (1999): “Enforcement of Employment Security Regulations, on-the-Job Search and Unemployment Duration,” *European Economic Review*, 43(1), 65–89.
- BROERSMA, L., AND J. V. OURS (1999): “Job Searchers, Job Matches, and the Elasticity of Matching,” *Labour Economics*, 6(1), 77–93.
- BURDETT, K., AND M. COLES (1999): “Long-Term Partnership Formation: Marriage and Employment,” *Economic Journal*, 109(456), F307–F333.
- BURGESS, S. M. (1993): “A Model of Competition Between Unemployed and Employed Job Searchers: An Application to the Unemployment Outflow in Britain,” *Economic Journal*, 103(420), 1190–1204.
- CALVO-ARMENGOL, A., AND Y. ZENOU (2001): “Job Matching, Social Networks and Word-of-Mouth Communication,” *CEPR Discussion Paper*, 2797.
- COLES, M., AND B. PETRONGOLO (2002): “A Test Between Unemployment Theories Using Matching Data,” *CEPR Discussion Paper*, 3241.
- COLES, M. G., AND E. SMITH (1998): “Marketplaces and Matching,” *International Economic Review*, 39(1), 239–254.
- DAVIS, S., AND J. HALTIWANGER (1992): “Gross Job Creation, Gross Job Destruction and Employment Reallocation,” *Quarterly Journal of Economics*, CVII(3), 819–863.
- FAHR, R., AND U. SUNDE (2001a): “Strategic Hiring Behavior in Empirical Matching Functions,” *IZA Discussion Paper*, 320.

- (2001b): “Disaggregate Matching Functions,” *IZA Discussion Paper*, 335.
- GARIBALDI, P. (1998): “Search Unemployment With Advance Notice,” *International Monetary Fund Working Paper*, 98/108.
- GAUTIER, P. (2002): “Search Externalities in a Model with Heterogeneous Jobs and Workers,” *Economica*, 273(1), 21–40.
- GREENE, W. (1997): *Econometric Analysis*. Prentice Hall, London, 3rd edn.
- KUGLER, A. D., AND G. SAINT-PAUL (2001): “How do Firing Costs affect Worker Flows in a World with Adverse Selection?,” *mimeo*.
- LINDEBOOM, M., J. V. OURS, AND G. RENES (1994): “Matching Employers and Workers: An Empirical Analysis on the Effectiveness of Search,” *Oxford Economic Papers*, 46(1), 45–67.
- MORTENSEN, D., AND C. PISSARIDES (1994): “Job Creation and Job Destruction in the Theory of Unemployment,” *Review of Economic Studies*, 61(1), 397–415.
- PETRONGOLO, B. (2001): “Reemployment Probabilities and Returns to Matching,” *Journal of Labor Economics*, 19(3), 716–741.
- PETRONGOLO, B., AND C. PISSARIDES (2001): “Looking Into the Black Box: A Survey of the Matching Function,” *Journal of Economic Literature*, 39(2), 390–431.
- PFANN, G. (2001): “Downsizing,” *IZA Discussion Paper*, 307.
- PISSARIDES, C. (1994): “Search Unemployment with On-the-Job Search,” *Review of Economic Studies*, 61(3), 457–475.
- POSTEL-VINAY, F., AND J.-M. ROBIN (2002): “Equilibrium Wage Dispersion with Worker and Employer Heterogeneity,” *Econometrica*, forthcoming.
- RAO, C. (1973): *Linear Statistical Inference and Its Applications*. Wiley, New York, 2nd edn.
- VAN OURS, J. (1995): “An Empirical Note on Employed and Unemployed Job Search,” *Economics Letters*, 49, 447–452.
- WARREN, R. S. (1996): “Returns to Scale in a Matching Model of the Labor Market,” *Economics Letters*, 50, 135–142.

YASHIV, E. (2000): “The Determinants of Equilibrium Unemployment,”  
*American Economic Review*, 90(5), 1297–1322.

# A Econometric Issues

## A.1 A.1 Estimation Bias

This appendix works out the consequences of regressing the matching function only on the observable stocks of unemployed job seekers  $u$  and registered vacancies  $r$  instead of also including employed job seekers  $e$  and non-registered vacancies  $n$  as explanatory variables in some detail. In order to do that, pose the problem in a somewhat more abstract form. In particular, consider the model:

$$y = X_1\beta_1 + X_2\beta_2 + X_3\beta_3 + X_4\beta_4 + \varepsilon, \quad (38)$$

where  $y$  is a vector of the dimension  $N \times 1$  reflecting the observed matches.  $N$  is the number of observations. The  $X_i$  are  $N \times 1$  vectors with  $X_1$  being (the log of) the stock of unemployed ( $u$ ),  $X_2$  being (the log of) the stock of registered vacancies,  $X_3$  being (the log of)  $\tilde{\phi}$ , and  $X_4$  (the log of)  $\tilde{\pi}$ . Since neither  $\tilde{\phi}$  nor  $\tilde{\pi}$  can be observed, rewrite the econometric model in terms of observables  $X = (X_1 \ X_2) = (u \ r)$ , which is a  $(N \times 2)$  matrix consisting of the  $N \times 1$  components  $u$  and  $r$ , and unobservables  $Z = (X_3 \ X_4) = (\tilde{\phi} \ \tilde{\pi})$ . Accordingly, denote  $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$  and  $\gamma = \begin{pmatrix} \beta_3 \\ \beta_4 \end{pmatrix}$  and write Equation (38) as:

$$y = X\beta + Z\gamma + \varepsilon, \quad (39)$$

Regressing  $y$  only on  $X$ , as is standard in estimations of empirical matching functions, delivers the following OLS estimator:

$$\hat{\beta} = (X'X)^{-1}X'y \quad (40)$$

$$= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'Z\gamma + (X'X)^{-1}X'\varepsilon \text{ and}$$

$$E(\hat{\beta}) = \beta + (X'X)^{-1}X'Z\gamma. \quad (41)$$

where  $E$  is the expectations operator. Note that:

$$X'X = \begin{pmatrix} X'_1X_1 & X'_1X_2 \\ X'_2X_1 & X'_2X_2 \end{pmatrix}$$

Using standard results on the inverse of partitioned matrices, one can show that:<sup>28</sup>

$$E(\hat{\beta}) = \beta + \begin{pmatrix} b_{31} + \frac{b_{21}s_{21}b_{31} - b_{21}s_{23}}{s_{22} - s_{21}b_{21}} & b_{41} + \frac{b_{21}s_{21}b_{41} - b_{21}s_{24}}{s_{22} - s_{21}b_{21}} \\ \frac{-s_{13}b_{21} + s_{23}}{s_{22} - s_{21}b_{21}} & \frac{-s_{14}b_{21} + s_{24}}{s_{22} - s_{21}b_{21}} \end{pmatrix} \gamma, \quad (42)$$

<sup>28</sup>The inverse of a partitioned matrix can be written as follows, cf. Rao (1973), p. 33:

$$\begin{pmatrix} A & B \\ B' & D \end{pmatrix} = \begin{pmatrix} A^{-1} + FE^{-1}F' & -FE^{-1} \\ -E^{-1}F' & E^{-1} \end{pmatrix},$$

where  $E = D - B'A^{-1}B$ , and  $F = A^{-1}B$ .

where  $E$  is the expectations operator,  $s_{12} = X_1'X_2$ ,  $b_{12} = (X_1'X_1)^{-1}X_1'X_2$ , etc. In the regression of  $y$  on  $X$  it is implicitly assumed that the observables  $X_1$  and  $X_2$  (ie.  $u$  and  $r$ ) are not perfectly correlated.<sup>29</sup> For illustration purposes, it is assumed in the text that  $s_{21} = 0$ , which simplifies the estimator considerably. However, this assumption is inconsequential, since the strategy to reveal biases and unbiased estimates of the coefficients of interest is valid also in the general case. Calculating the estimates gives:

$$\begin{aligned}\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} &= \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \frac{b_{13}+b_{21}s_{21}b_{13}-b_{21}s_{23}}{s_{22}-s_{21}b_{21}} & b_{14} + \frac{b_{21}s_{21}b_{14}-b_{21}s_{24}}{s_{22}-s_{21}b_{21}} \\ \frac{-b_{21}s_{13}+s_{23}}{s_{22}-s_{21}b_{21}} & \frac{-b_{21}s_{14}+s_{24}}{s_{22}-s_{21}b_{21}} \end{pmatrix} \begin{pmatrix} \beta_3 \\ \beta_4 \end{pmatrix} \\ &= \begin{pmatrix} \beta_u \\ \beta_r \end{pmatrix}.\end{aligned}\quad (43)$$

Analogously, one obtains for matching functions estimated only for new matches of formerly unemployed  $m_u$  (Equation 27):

$$\hat{\beta}^u = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} b_{14} + \frac{b_{21}s_{21}b_{14}-b_{21}s_{24}}{s_{22}-s_{21}b_{21}} \\ \frac{-b_{21}s_{14}+s_{24}}{s_{22}-s_{21}b_{21}} \end{pmatrix} \beta_4,\quad (44)$$

and likewise for the matching function for  $m_r$  (Equation 30):

$$\hat{\beta}^r = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \frac{b_{13}+b_{21}s_{21}b_{13}-b_{21}s_{23}}{s_{22}-s_{21}b_{21}} \\ \frac{-b_{21}s_{13}+s_{23}}{s_{22}-s_{21}b_{21}} \end{pmatrix} \beta_3.\quad (45)$$

In terms of notation in the text,  $x_u \equiv x_1$ ,  $x_r \equiv x_2$ ,  $x_e \equiv x_3$ , and  $x_n \equiv x_4$ , etc. Note also that the assumption of a unique and stable matching technology for all participants in the labor market means that  $\beta_u = \beta_{\bar{\phi}}$ , and  $\beta_r = \beta_{\bar{\pi}}$ . After a view manipulations, it is possible to reveal all relevant coefficients using Equations (43) to (45). In particular, estimates of the unbiased coefficients of interest can be obtained by:

$$\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \hat{\beta}^u - [\hat{\beta} - \hat{\beta}^r].\quad (46)$$

Using these coefficients and estimation results from Equations (43) to (45), there are four equations in the four unknown correlations  $s_{13}$ ,  $s_{14}$ ,  $s_{23}$ , and  $s_{24}$ , allowing to solve for the correlations predicted by the theoretical model.

## A.2 Error Variance

In order to pursue statistical tests about the imputed coefficients, one needs an estimate of the variance of the residuals  $\sigma^2$ . A standard candidate for this would be:<sup>30</sup>

$$\hat{\sigma}^2 = \frac{e_x'e_x}{N-2},\quad (47)$$

<sup>29</sup>This assumption was already discussed in footnote 23.

<sup>30</sup>See Greene (1997), p. 403.

where  $e_x$  is the  $(N \times 1)$  vector of residuals obtained by estimating  $y$  only on  $X$ . Note that

$$e_x = M_x y = M_x(X\beta + Z\gamma + \varepsilon) = M_x Z\gamma + M_1 \varepsilon,$$

with  $M_X = I - X'(X'X)^{-1}X'$ . One can show that (with  $E$  being the expectations operator):

$$E(e'_x e_x) = \gamma' Z' M_x Z \gamma + (N - 2)\sigma^2. \quad (48)$$

Due to the fact that  $Z$  is unobservable (remember, it contains employed job seekers  $e$  and non-registered vacancies  $n$ ), it is not possible to obtain a direct estimate of this error variance.

Nevertheless, there are ways to construct an estimate of the error variance that allow for inference. We next present two alternative approaches. The first approach gives a conservative measure of the variance that overestimates the true value. The second alternative allows to compute the value accurately under some more assumptions.

Consider  $\gamma' Z' M_x Z \gamma$  and substitute for the definition of  $M_x$ :

$$\begin{aligned} & \gamma' Z' (I - X'(X'X)^{-1}X') Z \gamma \\ &= \gamma' Z' Z \gamma - \gamma' Z' X (X'X)^{-1} X' Z \gamma \\ &= \gamma' Z' Z \gamma - \gamma' \begin{pmatrix} s_{13} & s_{14} \\ s_{23} & s_{24} \end{pmatrix}^{-1} \begin{pmatrix} b_{13} & b_{14} \\ b_{23} & b_{24} \end{pmatrix} \gamma. \end{aligned} \quad (49)$$

Hence, an estimate of the variance according to Equation (48) would be

$$(N - 2)\hat{\sigma}^2 = E(e'_x e_x) + \gamma' \begin{pmatrix} s_{13} & s_{14} \\ s_{23} & s_{24} \end{pmatrix}^{-1} \begin{pmatrix} b_{13} & b_{14} \\ b_{23} & b_{24} \end{pmatrix} \gamma - \gamma' Z' Z \gamma. \quad (50)$$

Taking a closer look at this estimate, one realizes that the first term on the RHS is available from the estimation of the conventional matching function ( $y$  on  $X$ ). Moreover, it has been shown before, that the components of  $\gamma$ , as well as all covariances and partial regression coefficients contained in the second term on the RHS can be imputed. The last term contains the variances and covariances of the unobservables combined with the respective parameters:

$$\gamma' Z' Z \gamma = (\beta_3 \beta_4) \begin{pmatrix} s_{33} & s_{34} \\ s_{43} & s_{44} \end{pmatrix} \begin{pmatrix} \beta_3 \\ \beta_4 \end{pmatrix} = (\beta_{\tilde{\phi}} \beta_{\tilde{\pi}}) \begin{pmatrix} s_{\tilde{\phi}\tilde{\phi}} & s_{\tilde{\phi}\tilde{\pi}} \\ s_{\tilde{\pi}\tilde{\phi}} & s_{\tilde{\pi}\tilde{\pi}} \end{pmatrix} \begin{pmatrix} \beta_{\tilde{\phi}} \\ \beta_{\tilde{\pi}} \end{pmatrix}$$

However, the parameters  $\beta_{\tilde{\phi}}$  and  $\beta_{\tilde{\pi}}$  are positive (they have already shown to be imputable). Moreover, from the theoretical model it can be expected that the covariances contained are positive (cf. Equation (13)), so that from this and the quadratic form, it is clear that the last term subtracts a positive number from the RHS. Hence, even though this number cannot be computed because  $Z$  is unobservable, one can calculate an estimate of the variance that overestimates the true variance by simply neglecting the last term on the RHS. The results will be robust

in the sense that such a procedure delivers significance levels for the tests that are lower than the true size of the test (i.e.  $t$ -values obtained would be downward biased).

An alternative approach involves estimation results for the matching functions using flows from unemployment and flows from registered vacancies, Equations (27) and (30). Assume that all three processes can be described by the same variance of errors. This assumption can be justified by the fact that the two specifications of matching functions with  $m_u$  and  $m_r$  as dependent variables are nested in the specification with  $m$  as regressand. Then, analogous estimates of the error variance in Equation (50) from the alternative specifications of the matching function deliver:

$$(N - 1) \hat{\sigma}^2 = E(e'_x e_x)|_{m_u} + \beta_3(s_{13} \ s_{23}) \begin{pmatrix} b_{13} \\ b_{23} \end{pmatrix} \beta_3 - \beta_3^2 s_{33}, \quad (51)$$

for flows from unemployment, and

$$(N - 1) \hat{\sigma}^2 = E(e'_x e_x)|_{m_r} + \beta_4(s_{14} \ s_{24}) \begin{pmatrix} b_{14} \\ b_{24} \end{pmatrix} \beta_4 - \beta_4^2 s_{44}, \quad (52)$$

respectively. Furthermore, assume that unobservable stocks are uncorrelated, that is  $s_{34} = \text{Ncov}(\tilde{\phi}, \tilde{\pi}) = 0$ . Under this assumption, subtracting Equation (50) from the sum of Equations (51) and (52) and rearranging gives:

$$\begin{aligned} N\hat{\sigma}^2 &= E(e'_x e_x)|_{m_u} + E(e'_x e_x)|_{m_r} - E(e'_x e_x) \\ &\quad - \beta_3\beta_4(s_{13}b_{14} + s_{23}b_{24} + s_{14}b_{23} + s_{24}b_{23}) \\ &= E(e'_x e_x)|_{m_u} + E(e'_x e_x)|_{m_r} - E(e'_x e_x) \\ &\quad - \beta_{\tilde{\phi}}\beta_{\tilde{\pi}}(s_{u\tilde{\phi}}b_{u\tilde{\pi}} + s_{r\tilde{\phi}}b_{r\tilde{\pi}} + s_{u\tilde{\pi}}b_{u\tilde{\phi}} + s_{r\tilde{\pi}}b_{r\tilde{\phi}}). \end{aligned} \quad (53)$$

All elements of this expression are either known or can be imputed, allowing to explicitly calculate an estimate of the error variance of the model.

# IZA Discussion Papers

No.	Author(s)	Title	Area	Date
503	R. Winter-Ebmer A. Wirz	Public Funding and Enrolment into Higher Education in Europe	3	05/02
504	L. Cappellari S. P. Jenkins	Modelling Low Income Transitions	4	05/02
505	T. K. Bauer	Migration, Sozialstaat und Zuwanderungspolitik	1	05/02
506	P. Díaz-Vázquez D. Snower	Can Insider Power Affect Employment?	3	05/02
507	E. Fehr A. Falk	Psychological Foundations of Incentives	5	05/02
508	C. Belzil J. Hansen	Unobserved Ability and the Return to Schooling	6	05/02
509	A. Kunze	The Timing of Careers and Human Capital Depreciation	1	06/02
510	E. S. Prasad	Wage Inequality in the United Kingdom, 1975-99	2	06/02
511	F. Büchel H. Battu	The Theory of Differential Overqualification: Does it Work?	1	06/02
512	C. Belzil J. Hansen	A Structural Analysis of the Correlated Random Coefficient Wage Regression Model	6	06/02
513	C. Belzil J. Hansen	Earnings Dispersion, Risk Aversion and Education	6	06/02
514	F. Schneider	The Size and Development of the Shadow Economies of 22 Transition and 21 OECD Countries	4	06/02
515	J. Hurley R. Vaithianathan T. F. Crossley D. Cobb-Clark	Parallel Private Health Insurance in Australia: A Cautionary Tale and Lessons for Canada	3	06/02
516	H. Bonin	Eine fiskalische Gesamtbilanz der Zuwanderung nach Deutschland	7	06/02
517	E. Tekin	Child Care Subsidies, Wages, and Employment of Single Mothers	3	06/02
518	P. Carneiro J. J. Heckman	The Evidence on Credit Constraints in Post-Secondary Schooling	5	06/02
519	S. Cohen Z. Eckstein	Labor Mobility of Immigrants: Training, Experience, Language and Opportunities	1	06/02
520	U. Sunde	Unobserved Bilateral Search on the Labor Market: A Theory-Based Correction for a Common Flaw in Empirical Matching Studies	1	06/02