Theory of Job Search.
Unemployment-Participation Tradeoff and Spatial Search with Asymmetric Changes of the Wage Distribution

Dissertation zur Erlangung des Grades eines Doktors der Wirtschaftswissenschaften

eingereicht an der Wirtschaftswissenschaftlichen Fakultät

der Universität Regensburg

vorgelegt von

Alisher Aldashev

Betreuer:

Prof. Dr. Joachim Möller

Prof. Dr. Lutz Arnold

Datum der Disputation 5. Juli 2007

Theory of Job Search. Unemployment-Participation Tradeoff and Spatial Search with Asymmetric Changes of the Wage Distribution

Alisher Aldashev

ZEW, Mannheim

Mannheim, November 14, 2007

Contents

Preface

This dissertation was written during my time as a doctorate student at the University of Regensburg and work on the project ”Flexibilität der Lohnstruktur, Ungleichheit und Beschäftigung - Eine vergleichende Mikrodatenuntersuchung für die USA und Deutschland” (under supervision of Prof. Dr. Möller), which is part of a greater project of the DFG ”Flexibilisierungspotenziale bei heterogenen Arbeitsmärkten”.

I wish to thank Prof. Dr. Joachim Möller and Prof. Dr. Lutz Arnold for their supervision and valuable comments and suggestions, which in my view helped to improve my thesis and added much to its scientific value. I benefited greatly from the discussions at the econometrics seminar and lunch seminar held at the University of Regensburg, especially with Dr. Jörg Lingens, Prof. Dr. emer. Walter Oberhofer, Prof. Dr. Rolf Tschernig (all University of Regensburg), Dr. Johannes Ludsteck (IAB), Dr. Thomas Schreck (BulwienGesa AG, München-Berlin). I also wish to thank Prof. Bernd Fitzenberger, PhD (University of Freiburg) for his thorough comments on one of my papers, which contributed much to this dissertation. I also enjoyed valuable discussions with Prof. Dr. Wolfgang Franz, Melanie Arntz, Alfred Garloff (all ZEW), Dr. Ralf Wilke (University of Leicester), Prof. Michael Burda, PhD (Humboldt University, Berlin). I also wish to thank my new colleagues at ZEW for their help: Dr. Stephan Lothar Thomsen (for his LaTeX expertise), Markus Clauss, Christian Göbel (for his tip on using Gnumeric).

Of course, this work would have never been possible without constant care and support of my family: my parents and grandparents, my sister Aisulu. And last, but certainly not least, I am indebted much to my dearest wife – Olga, not only for her help with the LaTeX formula editor, but most importantly for her love and understanding during all these years.

Chapter 1

Introduction

> Information economics has already had a profound effect on how we think about economic policy, and are likely to have an even greater influence in the future. The world is, of course, more complicated than our simple - or even our more complicated models - would suggest.
>
> J. Stiglitz, Nobel Prize lecture, December 8, 2001

Job search theory is a relatively young actor on the stage of economics and is an integral part of a broader field of economics of information. The ideas about functioning of the markets where information is costly to obtain were first enunciated by the Nobel Prize winner G. Stigler in his pioneering work ”The Economics of Information” published in the Journal of Political Economy in 1961. To explicate the very essence of what the economics of information and job search in particular all about I refer to Stigler himself:

> I propose on this occasion to address the same kinds of questions to an entirely different market: the market for new ideas in economic science. Most economists enter this market in new ideas, let me emphasize, in order to obtain ideas and methods for the applications they are making of economics to the thousand problems with which they are occupied: these economists are not the suppliers of new ideas but only demanders. Their problem is comparable to that of the automobile buyer: to find a reliable vehicle. Indeed, they usually end up by buying a used, and therefore tested, idea. Those economists who seek to engage in research on the new ideas of the science - to refute or confirm or develop or displace them - are in a sense both buyers and sellers of new ideas. They seek to develop new ideas and

> persuade the science to accept them, but they also are following clues and promises and explorations in the current or preceding ideas of the science. It is very costly to enter this market: it takes a good deal of time and thought to explore a new idea far enough to discover its promise or its lack of promise. The history of economics, and I assume of every science, is strewn with costly errors: of ideas, so to speak, that wouldn't run far or carry many passengers. How have economists dealt with this problem? That is my subject. (G. Stigler, Nobel Memorial Lecture, December 8, 1982)

The first application of the economics of information, in particular the notion that information is costly to obtain and returns in the future are uncertain, to the labor market appeared in Stigler's 1962 work. Since McCall (1970) seminal article the job search theory has become a standard tool for analyzing the decision making process of an unemployed individual who is looking for work.

An important contribution of the job search theory is an interpretation of unemployment on a microeconomic level. As Mortensen puts it:

> The theory of job search has developed as a complement to the older theoretical framework. Many writers found that the classic labor supply model with its emphasis on unilateral and fully informed choice could not explain important features of the typical individual's experience in the labor market. The experience of unemployment is an important example. Within the income-leisure choice framework, unemployment simply has no interpretation as a consequence of the assumptions that jobs are instantaneously available at market clearing wage rates known to the worker. (Mortensen (1986, p.850))

The original contribution of the search theory was the theoretic approach to the analysis of unemployment spell durations. In a partial equilibrium model (wages set by firms are considered exogenous), which stems from developments in the theory of sequential statistical decision theory, a worker is looking for a job in a decentralized labor market. Information on vacancies and the pay is imperfect and must be acquired before a worker becomes employed. Viewing this process as costly and sequential enables us to analyze variation in the unemployment spells that workers experience and in the wages received once employed. The length of time a worker spends looking for a job and the subsequent wage received once employed are both random variables with distributions which depend on the worker's individual characteristics as well as those of the environment through conditions the worker determines for acceptable employment. Because the framework has implications for the distribution of observables, likelihood functions for estimating the econometric model can be derived from the theory.

Applications of the search theory are rather broad. The partial equilibrium models concentrating on a decision-making process of an unemployed individual enable to make implications for the duration of unemployment, effect of unemployment insurance, mini-mum wages and etc. Another strand of literature looks at the problem from a different angle. Burdett and Mortensen (1998) analyze wage setting decision of firms within a search model framework. This seminal work illustrates why identical workers may receive different wages in equilibrium.

Full equilibrium models incorporate both decision-making of an unemployed individual looking for job and firm looking to fill a vacancy. Pissarides (1979) is one such example. In his work he derives a full market equilibrium which satisfies worker equilibrium condition and firm equilibrium condition. Albrecht and Axell (1984) derive an equilibrium unemployment rate which also satisfies worker and firm equilibrium conditions. Beside that, the authors were first to endogenously determine the wage offer distribution, which resulted from workers' different valuation of leisure.

Search models have been extended in different ways. An important contribution was made by Burdett (1978) who introduced the notion of the on-the-job search. The pioneering paper of Burdett (1978) and other works following afterwards helped explaining the job-to-job transitions and wage growth with the same employer. Another extension was introduced by Jovanovic (1979). Originally, it was assumed that the package of characteristics attributed to the job is known. In Jovanovic (1979), the worker does not know for sure the earning streams associated with the job or some other relevant characteristics. The worker must spend some time on that job to acquire all the relevant information. In this framework, the quit happens after the worker has learned all relevant information about the job and considers it "not a good match", i.e. the job did not met his expectations. Further extension to the basic model was the uncertainty about the offered wages. In a standard search model, the moments of the wage distribution are assumed to be known. Burdett and Wishwanath (1984) relax this assumption. In their model, workers have an expectation on the mean wage and dispersion. When they apply for a job they obtain information on the wage paid on this job. Every time they acquire information about the next job they update their expectation of the mean wage and dispersion in a Bayesian way. These extensions will not be formally presented in my work here and an interested reader is advised to follow the references.

Most of the search model frameworks imply time-invariant reservation wages. Economic reality suggests, however, that they are not. As early as in 1967, Kasper provided empirical evidence of declining reservation wages over the search span. Attempts have been undertaken to explain this phenomenon theoretically. Gronau (1971) claimed that constant reservation wage hypothesis does not hold if the infinite life horizon assumption is relaxed. However, as suggested by Mortensen (1986) this is rather an aging effect which cannot explain relatively large rates of

decline in reservation wages for relatively young workers reported in several studies. Hence, with the exception of elderly workers close to the retirement age, infinite time horizon is not a stumbling point. Mortensen (1986) provides an elegant explanation of declining reservation wages by imposing a credit market constraint. The general nonstationary job search model can be found in van den Berg (1990) where nonstationarity of reservation wages may arise due to time-dependence of any exogenous variable.

Applications of search theory give predictions about individuals' reservation wages, unemployment durations and reemployment opportunities. However, withdrawals from the labor market have received so far unfairly little attention, despite being an important indicator of labor market performance (Frijters and van der Klaauw (2006) is a notable exception). The dropouts were already mentioned in the search model of McCall (1970). However, the McCall (1970) model is static. This can well suit the decision making process whether to enter the labor market or not but fails to explain the exits from the labor market of the already participating workers. One of the objectives which I pursue in my work is to fill this gap by incorporating withdrawals from the labor market into a nonstationary job search model, and as I will show, withdrawals from labor force are a logical outcome of the nonstationary job search. I present a nonstationary job search model with the possibility of withdrawal from the labor market in Chapter 4.1.3. The outcome of the model is that lower reservation wages besides shorter unemployment duration also lead to higher exit rate from unemployment into nonparticipation. This tradeoff can be very important for unemployment insurance policy. As the simulations in Chapter 4.1.3 show higher unemployment compensation could ultimately lead to higher employment.

Another important aspect which motivated this dissertation is an asymmetric change of the wage offer distribution. In a standard search model (see for example Mortensen (1986)) reservation wages increase with the mean of the wage offer distribution and with the mean-preserving spread of the wage distribution. However, changing the spread of the distribution by holding the mean constant implies a symmetric "stretching" or "compressing" of the distribution in the tails. But what happens if the spread parameters for the left tail of the wage distribution and for the right tail may vary separately, which is usually the case when one faces the data? The complication arising here is that changing the spreads in the left and right tail unproportionately will affect the mean. To alleviate the problem I propose to use the notion of the median and the median-preserving spread. The model presented in Chapter 5 shows how predictions of the job-search theory change if asymmetric change of the wage distribution are allowed. The empirical results based on German regional data support the results of my theoretical model. This new insight into the job-search theory is very important for the regional empirical analysis where regional mean wage or dispersion are often used as regressors.

The structure of the dissertation is as follows: in Chapter 2, I present an overview of exist-

ing job-search literature with detailed description of selected works, which in my view are most innovative. In my dissertation, I concentrate on job search of the unemployed individuals, therefore Chapter 2 does not contain models allowing search on the job (for search on the job see a seminal paper Burdett (1978), respective chapters in Mortensen (1986) and Manning (2003) and citations therein. Chapter 4.1.3 presents a nonstationary search model with declining reservation wages. In this chapter I show the existence of a tradeoff between unemployment and participation, which is supported by simulations. Chapter 4 gives an overview of the econometric methods used to estimate duration models. Moreover, in this chapter the problem of an unemployment-participation tradeoff is addressed and methods of correcting the estimated failure rates are proposed. Chapter 5 gives a locational job-search model with a possibility of commuting with asymmetric changes of the wage offer distribution. Chapter 6 gives an overview of the estimation methods for count data and presents the results of the estimation of the German commuting data. Chapter 7 concludes, points out to potential drawbacks and open questions.

Chapter 2

Review of Literature

The original contribution of the search theory was the theoretic approach to the analysis of unemployment spells durations and dispersion of incomes. Theory of job search uses the tools of sequential statistical decision theory for the typical worker's problem of finding a job in a decentralized labor market. Information on vacancies and wages associated with them are considered as imperfect. This information has to be acquired before a worker can become employed. Viewing the process of acquiring information as costly is an important contribution of the economics of information and search theory. Because the search is costly an unemployed worker has to seek an optimal strategy which maximizes the present value of his future returns. Hence, with search costs present and time discounting no rational worker has an incentive to wait indefinitely for an opportunity to be employed. Besides, since the market is imperfect job offers are not immediately available. This explains the variation in unemployment durations in a search theoretical framework.

Since the pioneering work of Stigler (1962), search models have been widely used in labor market theory. Pissarides (1979) develops a search model in the presence of an employment agency. The model is grounded on rather restrictive assumptions, there is no wage variability and the separation rate is exogenous. The most important feature in his model is that two alternative search methods are possible - random search and search via an employment agency. The author shows that encouraging random search would increase the overall matching rate. In Hall (1979) there are no intermediaries, i.e. job-searchers and employers approach each other directly. Hall (1979) relaxes the assumption of a constant separation and job-finding rate. He proposes the existence of an efficient separation rate and an efficient job-finding rate as solutions to the maximization problem. Market equilibrium, given an efficient separation and job-finding rate, gives the natural unemployment rate. The classical search model was formulated by McCall (1970) where unemployed workers received job offers drawn from a non-degenerate wage offer distribution. The reservation wage in this model is a solution to an intertemporal maximization problem.

Another strand of search literature deals with search across spatial units. In Burda and Profit (1996), for example, agents do not search for vacancies only in their area of residence but can search in other districts as well. In this model distance is an important factor affecting the search effort. Individuals are set to optimize their search intensity taking the behavior of others as given. There are several works on commuting which are based on spatial search theoretical models (see e.g. van Ommeren and van der Straaten (2005)). Unlike in classical search model, which assumes wage dispersion, van Ommeren and van der Straaten (2005) allow for constant wages but dispersion of distances to work. Unemployed individuals in their model solve for maximal acceptable travel distance.

Many search theorists sought for models that could explain the dispersion in wages. Albrecht and Axell (1984) assumed two types of workers with each type having a different value attached to leisure. Because of differences in the value of leisure, reservation wages differ. The authors derive two reservation wages for the two types of workers. Firms are heterogeneous in Albrecht and Axell (1984) and differ in their productivity. The density function of productivities is assumed to be non-decreasing. Albrecht and Axell (1984) shows that differences in the value of leisure generate different wages offered by firms in equilibrium. Hence, the wage offer distribution is endogenously generated in the model. Eckstein and Wolpin (1990) extend the model of Albrecht and Axell (1984) to allow for more than two types of workers and endogenous wage offer probability which depends on the number of active firms in the market. In this model, although workers are homogenous in productivity, wage dispersion emerges due to differences in workers' tastes for leisure and efficiency of firms measured in terms of output per worker.

One of the most famous search models is probably that of Burdett and Mortensen (1998). The authors show that in the presence of the on-the-job search the dispersion of wages could arise even if all workers and firms are identical. Burdett and Mortensen (1998) and Albrecht and Axell (1984) models were synthesized by van den Berg and Ridder (1998) to account for workers and firms heterogeneity as well as for the on-the-job search. They also impose a legal minimum wage in their model. Similar approach is undertaken by Bontemps, Robin, and van den Berg (2000), who derive admissible endogenous wage distributions for a large class of productivity distributions (see also Bowlus and Grogan (2001)).

The literature review presented in this section uses notations of the original papers adapting it when necessary for the sake of conformity.

2.1 Models with Constant Wages (Urn Models)

2.1.1 Full Equilibrium Model with Employment Agency

This section describes the model of Pissarides (1979). Its important feature is that an employment agency is present in the model, although it is not always an intermediary between firms and searchers. Job searchers may choose to register at an employment agency and thus receive offers from the mediator. ”Random search” is also available to them, which means directly visiting firms and making inquiries about available vacancies. It is assumed that the unemployed always register (at the employment agency) to be entitled for the unemployment compensation. Nevertheless, some part of these jobless workers may choose the ”random search” to increase their employment chances. It is necessary, however, to introduce the cost of random search into the model, because otherwise all unemployed workers would be involved in a random search. The principle idea is that agents choose the optimal mixture of search methods to maximize the value of search.

In the same fashion, the firms in the model can also optimize their behavior with two alternatives possible: advertising the vacancies through the employment agency or privately. As Pissarides puts it: ”...firms may change the intensity of their search by switching to methods with higher job-matching probability (which in general be more expensive). It is this variable intensity of search that plays the crucial role in the determination of aggregate behavior in this model.” (Pissarides (1979), p.819)

For simplicity the author assumes that all firms offer the same wage to rule out the search on the job and rejection of job offers. Some fraction of the unemployed, $S \leqslant U$ also searches randomly (besides being registered at the agency) and x of them succeed. For the sake of simplicity it is assumed that each firm opens only one position. Total number of firms is V. R firms register their vacancy with the employment agency, $A = V - R$ firms advertise privately, thus A vacancies are available to random searchers.

Jobs are destroyed for exogenous reasons at a rate δ. Total separations are given by $\delta(L - U)$, where L is the total labor force. Each period the employment agency arranges y matches.

The agents do not know which firm would be visited by other workers. Therefore, the optimal strategy would be to choose a firm at random. The probability that a vacancy will not be searched by any of those S workers is given by $(1 - 1/A)^S$. And the probability that a firm in the set A will find a worker is given by $a = 1 - (1 - 1/A)^S$. The total number of job matchings each period is $x+y$, where x is the number of job matches through random search and y through employment agency.

The unemployment equilibrium condition implies equating the job inflows and outflows:

$$x(S, A) + y(U, R) = \delta(L - U), \tag{2.1}$$

where the left-hand side gives the total number of matchings and the right-hand side – the total number of separations.

Worker Equilibrium

The probability that a registered individual will receive an offer from the agency per period is given by $q = y/U$, and the probability that a random searcher will receive a job offers is given by $p = x/S$. In this model workers are assumed to be risk-neutral.

The models shares standards assumptions in the literature: registered unemployed receive the benefits b, random search costs c. Moreover, the agency charges workers for the job placement h.

Denote W to be the lifetime returns of the employed worker and Ω^U - the returns of the unemployed worker who is not searching randomly, which yields:

$$\begin{aligned} \Omega^U &= b + \frac{q}{1+r}W + \frac{1-q}{1+r}\Omega^U - q \cdot h \\ W &= w + \frac{\delta}{1+r}\Omega^U + \frac{1-\delta}{1+r}W. \end{aligned} \tag{2.2}$$

If they unemployed are engaged in a random search their value function can be written as:

$$\Omega^W = -c + \frac{p}{1+r}(W - \Omega^U). \tag{2.3}$$

Substituting for W and Ω^U one obtains:

$$\Omega^W = \frac{p}{r + \delta + q}(w - b + q \cdot h) - c. \tag{2.4}$$

Firm Equilibrium

Using previous notations, R firms register their vacancies with the employment agency, and A firms advertise privately. Let π denote the profit (since a firm may employ only one worker it is also profit per worker), and ρ be the cost of capital needed for one job position per period, which is a sunk cost that borne irrespective of whether the position is filled or vacant. If firms choose to advertise privately, they pay advertisement cost, α per vacancy. The profit from a filled position is thus $\pi - w - \rho$ per period, whereas the vacant position has a cost of ρ per

period. Denote Π as returns from a filled position and Ω^A as returns from a vacancy, which is not registered at the employment agency. Then:

$$\begin{aligned} \Omega^A &= -\alpha - \rho + \frac{a}{1+R}\Pi + \frac{1-a}{1+r}\Omega^A \\ \Pi &= \pi - w - \rho + \frac{\delta}{1+r}\Omega^A + \frac{1-\delta}{1+r}\Pi. \end{aligned} \tag{2.5}$$

Denote the returns from a registered vacancy by Ω_*^R, which gives:

$$\begin{aligned} \Omega_*^R &= -\rho + \frac{g}{1+r}\Pi + \frac{1-g}{1+r}\Omega_*^R - g \cdot \upsilon \\ \Pi &= \pi - w - \rho + \frac{\delta}{1+r}\Omega^R + \frac{1-\delta}{1+r}\Pi \\ \Omega_*^R &= \Omega^R + C, \end{aligned} \tag{2.6}$$

where Ω^R is the value of a vacancy before registration, C is a registration cost of a vacancy, and υ is the fee that the agency charges the firm for a successful match. ”Firms will choose the job-matching method that yields the highest returns. In equilibrium, if both methods are used by different firms, both methods must be equally attractive, so $\Omega^A = \Omega^R$.” (Pissarides (1979, p. 823))

In equilibrium:

$$\begin{aligned} \Omega^A &= 0 \\ \Omega^R &= 0. \end{aligned} \tag{2.7}$$

Market Equilibrium

Market equilibrium is given by the simultaneous satisfaction of the four equilibrium conditions:

$$\begin{cases} \Omega^W = 0 \\ \Omega^A = 0 \\ \Omega^R = 0 \\ -x(S,A) - y(U,R) - \delta U + \delta L = 0. \end{cases} \tag{2.8}$$

The first condition implies that net marginal returns from random search are zero. The next two mean that the returns from opening up more vacancies are zero. The last one implies that the unemployment pool is constant over time, i.e. inflows into unemployment pool equal the outflow to jobs.

2.1.2 Equilibrium Unemployment Rate and Employment Duration

This section describes the model of Hall (1979). A distinctive aspect of Hall (1979) model is that he stresses the notion of the unemployment and employment duration as an important factor which determines the equilibrium unemployment rate. Hall introduces the concept of an ”efficient duration” of employment. For an employer the efficient duration depends on the cost of recruiting and training; and for the worker it depends on the cost of finding new jobs. In tight markets where jobs are easy to find, workers prefer shorter jobs but this imposes higher recruiting cost on employers so they favor longer jobs.

Equilibrium Unemployment Rate

Suppose that jobs and workers are perfectly homogeneous. A vacancy in this model is instantly filled, but the unemployed must wait until a job offer arrives, which is a stochastic process. The unemployed accept the first job offer encountered. If they receive more than one, they accept one at random because jobs are homogeneous with respect to wage.

Employers make V offers to S job-seekers. The probability that a particular worker receives a particular job is $1/S$. The probability that an agent will receive no offers at all is:

$$1 - f = (1 - 1/S)^V = \left[(1 - 1/S)^{-S}\right]^{-V/S}, \tag{2.9}$$

where f is the job-finding rate. If $S \to \infty$, $(1 - 1/S)^{-S} \to e$, which yields the solution for the job-finding rate:

$$f = 1 - e^{-V/S}. \tag{2.10}$$

Define ρ as the number of vacancies needed to be opened to generate one job on average, which is $\rho = V/(f \cdot S)$. From Equation 2.10 it follows that $V/S = -\ln(1 - f)$, hence:

$$\rho(f) = -\ln(1 - f)/f. \tag{2.11}$$

The equilibrium is achieved when the flows into an out of unemployment are equated. Given a separation rate, s, this can be given as:

$$S = (1 - f)S + s \cdot E, \tag{2.12}$$

or

$$S/E = s/f, \tag{2.13}$$

where E is the number of employed workers. The unemployment rate can be given then as:

$$u = \frac{(1-f)S}{E + (1-f)S} = \frac{s}{s + f/(1-f)}. \tag{2.14}$$

It is obvious from Equation 2.14 that the unemployment rate increases with the separation rate and decreases with the job-finding rate, which is intuitively clear.

Optimal Employment Duration

From the firm side there should exist an optimal contract length. Firstly, because firms would dislike very short employment spells due to fixed costs associated with recruiting and training. Secondly, very long contracts make firms inflexible in adjusting their employment level in response to economic situation as firms would have to pay workers the compensation for premature contract termination and layoff. Hence, firms would pay lower wages for very short contracts and lower wages for very long contracts and higher wages for some intermediate length contracts. This could be given by an isocost line plotted against the contract length, which is concave and has a maximum.

For workers the tradeoff also exists. Very short jobs may be too costly because looking for a new job involves certain expenses. On the other hand very long jobs reduces workers' flexibility, hence workers would require a compensation (in terms of higher wages) for very short contracts and very long contracts to be indifferent among the job offers with various contract lengths. This would give an indifference curve for a worker, which is likely to be convex.[1] The intersection of an isocost and indifference curve gives an efficient wage and efficient contract length. Treating the expected duration of employment as a reciprocal of a separation rate, this intersection gives the solution for an efficient separation rate.

Efficient Job Finding Rate

Let $\lambda(s)$ be the probability that a job will be filled as a function of the separation rate. The expected cost associated with the job is $w \cdot \lambda(s)$. The unconditional probability that a job is unfilled is s. However, for a separation to occur, the vacancy must be filled. This gives the conditional probability that a job is unfilled $s \cdot \lambda(s)$. To keep the job filled a firm needs to hire at a rate $s \cdot \lambda(s)$, which requires a flow of offers of $\rho(f)s \cdot \lambda(s)$.

Suppose that each offer costs the firm $\mu \cdot w$, so μ is the fraction of the wage. Then the firm is interested in minimizing its cost function given by:

$$C(w, s, f) = w \cdot \lambda(s) \cdot (r \cdot s \cdot \rho(f) + 1). \tag{2.15}$$

Moreover, suppose that workers are interested in maximizing the *effective* income, where effective income is defined as:

$$y = (1-u)w = \frac{f}{(1-f)s+f}w. \tag{2.16}$$

So the effective income is the expected wage conditional upon being employed; u can be also defined as the fraction of time agents expect to be unemployed. The equation in 2.16 imposes a constraint onto 2.15, hence the efficient job-finding rate can be given as:

$$f = \arg\min\left\{\left(1+s\frac{1-f}{f}\right)y\cdot\lambda(s)\cdot(\mu\cdot s\cdot\rho(f)+1)\right\}, \tag{2.17}$$

since both y and $\lambda(s)$ are independent of f, Equation 2.17 simplifies to:

$$f = \arg\min\left\{\left(1+s\frac{1-f}{f}\right)\cdot(\mu\cdot s\cdot\rho(f)+1)\right\}. \tag{2.18}$$

In the similar fashion, the efficient separation rate solves for:

$$s = \arg\min(f) = \arg\min\left\{\left(1+s\frac{1-f}{f}\right)\lambda(s)\cdot(\mu\cdot s\cdot\rho(f)+1)\right\}. \tag{2.19}$$

The system of two equations 2.18 and 2.19 with two endogenous parameters s and f guarantee the unique solution to the natural unemployment rate given in 2.14.

The model of Hall (1979) solves for the natural unemployment rate as the outcome of efficient employment arrangements. Hence, in this sense, the natural unemployment rate is socially optimal unemployment.

2.1.3 Spatial Search with Constant Wages

This section describes the model of Burda and Profit (1996). In their model, Burda and Profit (1996) consider an individual who can determine his search activity in two dimensions: where to search and how many jobs to apply for. Each spatial unit has a job agency which mediates contacts between searchers and potential employers. Workers apply randomly to firms. Once applied, an individual is invited to an interview after which it is decided whether he is accepted or not. There is a cost, c, associated with each interview. If he applies for a job in the region j different from the region of origin, say, i, then interview cost is increased to $c + a \cdot D_{ij}$. Where, D_{ij} is the distance between the regions i and j in kilometers and a is thus a per-kilometer cost of travel. Individuals assume to optimize their search intensity in order to maximize their expected net income. Let f denote the job finding probability, r - the interest rate, and m - the search

intensity, which can be thought of as number of jobs to apply for. Then, the objective function for an individual searching in the region j is given by:

$$\max_{m_j} \left[1 - (1 - f_j)^{m_j}\right] w/r - m_j(c + a \cdot D_{ij}). \tag{2.20}$$

The first term in Equation 2.20 is the expected benefit from making m_j interviews in the region j and the second term is the total cost of those interviews. Assuming for simplicity that the expected income in unemployment equals zero, the authors derive the solution:

$$m_j = \begin{cases} f_j^{-1} \cdot \ln\Big(f_j(w/r)/(c + a \cdot D_{ij})\Big) & \text{for } f_j(w/r)/(c + a \cdot D_{ij}) \geqslant 1 \\ 0 & \text{otherwise.} \end{cases} \tag{2.21}$$

From Equation 2.21 it is seen that optimal search intensity is increasing in the wage and decreasing in discount rate and the cost of applying for the job. The effect of job finding probability on search intensity is ambiguous. One the one hand, it increases the expected benefits, but on the other hand, at given relative returns, less search is necessary to achieve the same expected benefits.

The model of Burda and Profit (1996) uses the same search principle as in Pissarides (1979). The wage is constant, thus jobs do not differ from one another, and therefore application to firms is a random process. However, in their model agents optimize their search intensity in each spatial unit given the distance between the residence location and potential employment location and difference in job-finding probability across spatial units.

2.2 Exogenous Wage Dispersion

2.2.1 Single Wage Offer Model, Discrete Time

This section describes the model of Franz (2006). The necessary condition for an unemployed individual to accept a job offer is that the wage offer exceeds his reservation wage: $w > w^R$. Suppose that an individual receives a wage offer with probability q, which depends on the general situation in the labor market and personal characteristics, such as age, sex, qualification etc (which are represented by a vector z). Moreover, it is assumed that for each individual the chances of getting a job are decreasing with the wage offer, due to increasing competition for high-paid jobs. The probability of a successful match is given as:

$$p(z, w^R) = \int_{w^R} q(z, w) f(w) dw. \tag{2.22}$$

The expected wage is given as:

$$E(w|w \geqslant w^R) = \frac{\int\limits_{w^R} wq(z,w)f(w)dw}{\int\limits_{w^R} q(z,w)f(w)dw}. \tag{2.23}$$

The expected wage is a conditional expectation, given that the wage offer exceeds the reservation wage. The present value of the reservation wage is:

$$\sum_{t=0}^{\infty} \frac{w^R}{(1+r)^t} = \frac{(1+r)w^R}{r}. \tag{2.24}$$

Denote the unemployment benefits, which an agent receives staying unemployed by b and the constant search cost per period as c, the present value of expected wages (if a person gets a job) and costs of search equal:

$$\frac{(b-c)(1+r)}{r+p(z,w^R)} + p(z,w^R) \cdot E(w|w \geqslant w^R)\frac{1+r}{r(r+p(z,w^R))}. \tag{2.25}$$

In the optimum the present value of acceptable wage offers must be equal the present value of the returns to search. Knowing that an agent accepts only those wage offers exceeding his reservation wage, in the optimum the discounted returns to search equal the discounted reservation wage, i.e. Equation 2.24 equals Equation 2.25. Solving it for w^R gives:

$$w^R = \frac{r(b-c) + p(z,w^R)E(w|w \geqslant w^R)}{r+p(z,w^R)}. \tag{2.26}$$

This relationship shows that the reservation wage decreases with search costs, increases with unemployment benefits and wages. Also a lower probability of receiving and accepting a wage offer, due to personal characteristics or labor market conditions, reduces the reservation wage. It is easy to derive from Equation 2.26. In fact:

$$\begin{aligned} \frac{\partial w^R}{\partial p(z,w^R)} &= \frac{E(w|w \geqslant w^R)(r+p(z,w)) - r(b-c) - p(z,w^R)E(w|w \geqslant w^R)}{(r+p(z,w^r))^2} = \\ &\frac{r(E\left[w|w \geqslant w^R\right] + c - b)}{(r+p(z,w^R))^2} > 0, \end{aligned} \tag{2.27}$$

as the value of employment must exceed the value of unemployment; otherwise agents would not accept job offers at all and remain constantly unemployed.

2.2.2 Multiple Wage Offer Model, Continuous Time

This section is based on the model presented in Mortensen (1986), which is an extension of McCall (1970). A distinct feature of this model is that besides wage variability it considers that in case of multiple offers agents pick the offer with the highest wage. The model assumes that workers live forever and there are no separations and quits. No recall is possible, i.e. once job offer is declined, an agent cannot go back to the offer and accept it.

Consider an agent who looks for a job. Suppose the search costs him c per time period and $\beta(h)$ is the discounting factor for time τ. The distribution of wages is given as $F(w)$. In a given period an agent could receive n job offers, with n being a random variable which follows the Poisson distribution: $q(m,\tau) = \dfrac{e^{-\lambda\tau}(\lambda\tau)^m}{m!}$, where λ is the offer arrival rate, which is assumed to be exogenous. Discounting is assumed to be continuous and the discount rate is given as: $\beta(\tau) = e^{-r\tau}$. The workers are assumed to be wealth maximizers, which is equivalent to utility maximizing in risk-neutral case. If an agent receives multiple offers he accepts the one with the highest wage. Defining $\tilde{w}_m = \max\{w_1, w_2, ..., w_m\}$, the distribution of wages accepted by an agent is an extreme value distribution $G(\tilde{w}_m)$. Assuming that $F(w)$ and $q(m,\tau)$ are time invariant, the value of search becomes time independent and the Bellman equation could written as:

$$\begin{aligned}\Omega &= (b-c)\tau + \beta(\tau)\left[\textstyle\sum_{m=1}^{\infty} q(m,\tau)\int\limits_0^{\infty} \max\left[\Omega, W(w)\right] g(\tilde{w}_m)dw + q(0,\tau)\Omega\right] = \\ &= (b-c)\tau + \beta(\tau)\left[\textstyle\sum_{m=1}^{\infty} q(m,\tau)\int\limits_0^{\infty} \max\left[0, W(w)-\Omega\right] g(\tilde{w}_m)dw + \Omega\right].\end{aligned} \tag{2.28}$$

where Ω denotes the value of search, b - value of leisure per period, $W(w)$ - value of employment. The value of employment can be written as:

$$W(w) = w\tau + \beta(\tau)W(w) = \frac{w\tau}{1-\beta(\tau)}. \tag{2.29}$$

It could be seen that $W(w)$ is monotonically increasing in wage. The value of search as a function of a wage satisfies Blackwell's sufficient conditions for a contraction: (1) Ω is monotonic in w, (2) Ω is discountable. This implies that only one fixed point exists where $\Omega(W(w)) = W(w)$. Defining the reservation wage as the wage at which an agent is indifferent between accepting a job and continuing search, this implies: $W(w^R) = \Omega$, which is unique, where w^R stands for the reservation wage.

The probability of receiving more than one offer within an infinitesimal interval of time is virtually zero ($\tau \to 0$). Hence, in continuous time, the Bellman equation in 2.28 considerably simplifies:

$$r\Omega = (b - c) + \lambda \int_0^\infty \max[0, W(w) - \Omega] dF(w). \tag{2.30}$$

Knowing that: $w = rW(w)$ and $W(w^R) = \Omega$, we obtain: $w^R = r\Omega$, which yields:

$$w^R = b - c + \frac{\lambda}{r} \int_{w^R}^\infty (w - w^R) dF(w). \tag{2.31}$$

Applying integration by parts to Equation 2.31 yields:

$$w^R = \frac{r}{r+\lambda}(b - c) + \frac{\lambda}{r+\lambda} E(w) + \int_0^{w^R} F(w) dw. \tag{2.32}$$

One could see from Equation 2.32 that the reservation wage depends on the wage distribution $F(w)$. Usually the distribution is characterized by the first two moments: the mean as a location parameter and the variance as a scale parameter. In the search literature it is conventional to use the mean-preserving spread as a scale parameter which characterizes the shape of the distribution due to Rotschild and Stiglitz (1970). Let s_1 and s_2 be mean-preserving spreads of the wage offer distribution $F(w)$. By definition of the mean-preserving spread, $\int_0^\infty w dF(w; s_1) = \int_0^\infty w dF(w; s_2)$. It can be shown that $\int_0^x F(w; s_1) dw > \int_0^x F(w; s_2)$ for $s_1 > s_2$, which implies that the reservation wage is:

- increasing in the value of time, i.e. if leisure becomes more valuable, agents increase their reservation wages;
- decreasing in the cost of search, i.e. when search becomes more costly, agents are willing to accept lower-paid jobs;
- increasing in the mean of the wage offer distribution; the magnitude of the effect is less than unity in absolute value, i.e. when the mean of the wage offer distribution goes up the reservation wage will also increase but not as much;

- increasing in the mean-preserving spread, i.e. when dispersion of wages go up, agents would not accept lower-paid jobs anymore; this is "... the consequence of the fact that the worker has the option of waiting for an offer in the upper tail of the wage distribution" (Mortensen (1986, p. 865)).

2.3 Endogenous Wage Dispersion

2.3.1 Wage Dispersion Due to Worker Heterogeneity

This section is based on the paper of Albrecht and Axell (1984). An important feature of their work is that unlike many previous models, in their setting the wage distribution is endogenously determined. Moreover, they introduce heterogeneity of workers who differ in their value of leisure.

Individuals

Individuals are assumed to maximize their lifetime utility. Let the utility function be given in the form:

$$U = C + \nu L, \tag{2.33}$$

where U stands for utility, C for consumption, L denotes leisure, and parameter ν is attributed to a "consumption value" of leisure. L in the model is a binary choice variable taking the value of zero when an agent is working and one when he is searching. Individual's wage rate is denoted by w; and a non-wage income consists of "dividends", θ, which do not depend on the individual's employment status; [2] unemployment compensation is denoted by b. Individual's utility choice becomes:

$$U = \begin{cases} w + \theta & \text{if working at a wage } w \\ \theta + b + \nu & \text{if searching .} \end{cases} \tag{2.34}$$

The economy in the model in any period consists of k individuals and n firms. Firms are assumed to live forever, but individuals exits the economy at the end of any period at a rate τ. The key assumption of the model is that there are two types of individuals in the economy who differ in the value they attach to leisure.

Consider a two-point wage distribution in the model with w_0 and w_1 denoting the low and the high wage respectively, and let γ be the share of firms offering the low wage. In equilibrium, w_0 must be the reservation wage of those individuals who derive low utility from leisure, and w_1 must be the reservation wage of those who derive high utility from leisure.

Individuals who impute low value to leisure will accept the first wage offer encountered, whereas those who impute high value to leisure will search until they are offered w_1. Let β denote the fraction of individuals with low value of leisure. The amount of search in the economy - measured by the unemployment rate - is therefore an increasing function of γ.

Consider a search behavior of an individual with the value of leisure ν_0 who has drawn a wage w_0. If he rejects w_0, then he "earns" leisure worth ν_0, a non-wage income $\theta + b$, and with probability $1 - \tau$ draws a wage in the next period. The wage sampled from a subsequent drawing equals w_1 with the probability $1 - \gamma$; and if w_1 is actually drawn, then it is accepted, resulting in an expected future life-time utility of $(w_1 + \theta)/\tau$. Otherwise, he continues the search.

The value of rejecting w_0 is therefore:

$$\begin{aligned} V &= \nu_0 + b + \theta + (1-\tau)\left[\frac{(1-\gamma)(w_1+\theta)}{\tau} + \gamma V\right] = \\ &= \frac{\nu_0 + b}{1-\gamma(1-\tau)} + \frac{(1-\gamma)(1-\tau)}{1-\gamma(1-\tau)}\frac{w_1}{\tau} + \frac{\theta}{\tau}. \end{aligned} \tag{2.35}$$

Naturally, w_0 is the reservation wage for the individual with the value of leisure ν_0. Hence:

$$V = (w_0 + \theta)/\tau. \tag{2.36}$$

Solving Equations 2.35 and 2.36 for w_0 yields:

$$w_0 = \frac{(1-\gamma)(1-\tau)}{1-\gamma(1-\tau)} w_1 + \frac{\tau(\nu_0 + b)}{1-\gamma(1-\tau)}. \tag{2.37}$$

If w_1 is the reservation wage for ν_1 individuals, then in equilibrium the value of rejecting w_1 must equal the value of accepting w_1:

$$(\nu_1 + b + \theta)/\tau = (w_1 + \theta)/\tau \Rightarrow w_1 = \nu_1 + b. \tag{2.38}$$

Equations 2.37 and 2.38 give the solution for the wage distribution. In fact, one could see that wages are endogenously determined by the set of exogenous parameters.

Firms

Firms produce according to the linear production function $y = p \cdot \ell$, where ℓ denotes the amount of labor a firm hires and p is the productivity of labor, which is distributed across firms according to a distribution function $A(p)$, with the corresponding density function $a(p)$. p is normalized to lie within a unit interval. Let $\ell(w)$ be per period labor supply to a firm offering a wage w, then the profit of this firm solves for:

$$\Pi(w, p) = (p - w) \cdot \ell(w). \tag{2.39}$$

For a firm to be profitable $p > w$ must hold. So firms whose productivity is less than the lowest reservation wage, w_0, are out of the market, therefore only a fraction $1 - A(w_0)$ is active. Among active firms a fraction γ offers w_0 and a fraction $1 - \gamma$ offers w_1. A profit-maximizing firm needs to choose between offering a wage w_0 or w_1.

The authors further define a wage-indifference point, p^*. A firm with productivity p^* is indifferent between offering w_0 or w_1. Consequently, firms with $w_0 < p \leqslant p^*$ will offer w_0 and firms with $p^* < p < 1$ will offer w_1. The equilibrium condition is thus:

$$\gamma = \frac{A(p^*) - A(w_0)}{1 - A(w_0)}. \tag{2.40}$$

The nominator reflects the share of firms offering w_0 among *all* firms. Dividing it by $1 - A(w_0)$ gives the share of firms offering w_0 among *active* firms. The solution for "indifference productivity" is:

$$p^* = \frac{w_1 \ell(w_1) - w_0 \ell(w_0)}{\ell(w_1) - \ell(w_0)}. \tag{2.41}$$

Next, Albrecht and Axell (1984) derive $\ell(w_0)$ and $\ell(w_1)$. If a firm offers w_0, only individuals with the low value of leisure will accept that. Each period $\tau \cdot k \cdot \beta$ individuals with low value of leisure enter the economy. Letting $\mu \equiv \frac{k}{n}\Big[1 - A(w_0)\Big]$, there are $\tau \cdot \mu \cdot \beta$ individuals with

the value of leisure ν_0 per active firm entering the economy each period. All of those searchers contacting a firm will accept an offer. Hence, $\ell(w_0)$ can be computed as the sum of the $\tau \cdot \mu \cdot \beta$ agents who accept the low wage in the current period, the $(1-\tau)\tau \cdot \mu \cdot \beta$ surviving individuals who accepted w_0 in the previous period, and so on. Therefore:

$$\ell(w_0) = \tau\mu\beta \left[1 + (1-\tau) + (1-\tau)^2 + ...\right] = \mu\beta. \tag{2.42}$$

For a firm offering w_1, all individuals contacting this firm accept the offer and the number of contacts per firm per period is the sum of:

$\tau\mu\beta$	ν_0 individuals entering the economy
$\tau\mu(1-\beta)$	ν_1 individuals entering the economy
$\tau\mu(1-\beta)\gamma(1-\tau)$	ν_1 individuals who have searched once
$\tau\mu(1-\beta)\gamma^2(1-\tau)^2$	ν_1 individuals who have searched twice,

and so on...

The resulting sum is then:

$$\tau\mu\beta + \tau\mu(1-\beta)[1 + \gamma(1-\tau) + \gamma^2(1-\tau)^2 + ...] = \tau\mu\beta + \frac{\tau\mu(1-\beta)}{1-\gamma(1-\tau)}. \tag{2.43}$$

As a result $\ell(w_1) = \mu\beta + \dfrac{\mu(1-\beta)}{1-\gamma(1-\tau)}$.

The derivation of the equilibrium unemployment rate is rather straightforward. In any period there are $\tau \cdot k(1-\beta)\gamma$ agents who search for the first time, $\tau \cdot k(1-\beta)\gamma^2(1-\tau)$ agents who search for the second time, and so on. Thereby, the equilibrium unemployment rate can be given as:

$$u = \tau(1-\beta)\gamma[1 + \gamma(1-\tau) + \gamma^2(1-\tau)^2 + ...] = \frac{\tau(1-\beta)\gamma}{1-\gamma(1-\tau)}, \tag{2.44}$$

with

$$\frac{du}{d\gamma} = \frac{\tau(1-\beta)}{[1-\gamma(1-\tau)]^2} > 0. \tag{2.45}$$

So the equilibrium unemployment rate is an increasing function of the share of firms offering a low wage as required. The cutoff productivity can then be derived as:

$$p^* = \frac{w_1\ell(w_1) - w_0\ell(w_0)}{\ell(w_1) - \ell(w_0)} = w_1 + \frac{(w_1 - w_0)\ell(w_0)}{\ell(w_1) - \ell(w_0)} = w_1 + \frac{(w_1 - w_0)\beta}{(1-\beta)/[1-\gamma(1-\tau)]}. \tag{2.46}$$

Using Equations 2.37 and 2.38 one could establish that $w_1 - w_0 = \dfrac{\tau(\nu_1 - \nu_0)}{1-\gamma(1-\tau)}$, hence:

$$p^* = v_1 + b + \frac{\tau(\nu_1 - \nu_0)\beta}{1-\beta} \tag{2.47}$$

The comparative statics show that the equilibrium unemployment rate increases with the unemployment benefit. However, strikingly, if one can discriminate the unemployment insurance between individuals with low value of leisure and high value of leisure, increase in unemployment compensation for agents who impute low value to leisure decreases the equilibrium unemployment rate!

Efficiency Issues

Albrecht and Axell (1984) established that an increase in unemployment compensation leads to higher unemployment for a broad class of productivity distribution functions. However, they show further that the socially optimal level of unemployment is not necessarily zero.

Define the social objective function as per capita utility:

$$U^* = C^* + \nu_1 u. \tag{2.48}$$

The first term in Equation 2.48 represents per capita consumption and the second term is the value of leisure per capita. Per capita production (which is equal to per capita consumption) is the sum of production from low-wage firms and high-wage firms. Production from low-wage firms is the product of (i) the number of firms offering w_0, which is equal to $n[A(p^*)]$, (ii) labor supply to low-wage firms, $\ell(w_0)$ and (iii) the average productivity of firms offering w_0, which is $\int_{w_0}^{p^*} p dA(p)/[A(p^*) - A(w_0)]$. Hence the total production from low-wage

firms is $n\ell(w_0)\int_{w_0}^{p^*} pdA(p)$. In the same fashion total production from high-wage firms is $n\ell(w_1)\int_{p^*}^{1} pdA(p)$.

Equilibrium per capital consumption can be given therefore as:

$$C^* = \frac{n}{k}\left[\ell(w_0)\int_{w_0}^{p^*} pdA(p) + \ell(w_1)\int_{p^*}^{1} pdA(p)\right] = \\ = \frac{1}{\mu[1-A(w_0)]}\left\{\ell(w_0)\int_{w_0}^{1} pdA(p) + [\ell(w_1)-\ell(w_0)]\int_{p^*}^{1} pdA(p)\right\}. \tag{2.49}$$

Or alternatively,

$$C^* = \beta\int_{w_0}^{p^*}\frac{pdA(p)}{1-A(w_0)} + \frac{(1-\beta)(1-\gamma)}{1-\gamma(1-\tau)}\int_{p^*}^{1}\frac{pdA(p)}{1-A(p^*)}. \tag{2.50}$$

Defining the unemployment rate among the ν_1 individuals as $u_1 = \tau\gamma/[1-\gamma(1-\tau)]$ yields:

$$C^* = \beta\int_{w_0}^{1}\frac{pdA(p)}{1-A(w_0)} + (1-\beta)(1-u_1)\int_{p^*}^{1}\frac{pdA(p)}{1-A(p^*)}. \tag{2.51}$$

Then the per capita utility is given by:

$$U^* = \beta\int_{w_0}^{1}\frac{pdA(p)}{1-A(w_0)} + (1-\beta)\int_{p^*}^{1}\frac{pdA(p)}{1-A(p^*)} - u\int_{p^*}^{1}\frac{(p-\nu_1)dA(p)}{1-A(p^*)}. \tag{2.52}$$

As stated before unemployment is likely to rise with unemployment compensation. It is clear from Equation 2.52 that increase in the equilibrium unemployment rate leads to a utility loss. However, one must keep in mind that the wage distribution also changes with b. Increase in unemployment compensation raises the reservation wages. Searchers sort out less productive firms which cannot match their wage aspirations and workers are consequently employed by more productive firms. Hence, the change in the wage distribution drives inefficient firms out of the market. This results in a utility gain. The total effect is unclear, but what is important from this analysis is that unemployment compensation is not necessarily socially undesirable.

2.3.2 Equilibrium Wage Dispersion with Identical Workers

The Burdett-Mortensen model is one of the major contributions to the theory of labor economics of the last decade. The model was able to explain the long-standing question why observationally equivalent workers are still paid different wages in equilibrium. As D. Margolis puts it:

> Like researchers in many other fields of science, some labor economists have been looking for a ”universal theory of everything”, or at least insofar as concerns labor market outcomes like employment and unemployment, wage distributions, firm size, seniority returns, and so on. The enthusiasm with which the literature has adopted the Burdett-Mortensen (1998) and Pissarides (2000) frameworks suggests that some macro-based labor economists, especially in Europe, believe they have found their holy grail. (D. Margolis, Annotation to the book of Mortensen (2003)).

This section is based on a seminal paper of Burdett and Mortensen (1998) (a nice overview of the Burdett and Mortensen (1998) model can be found in Manning (2003) from where the notation is borrowed). Consider an economy with M_F firms and M_W workers. Let M_F and M_W be fixed and denote $M \equiv M_F/M_W$. It is assumed that each firm opens but one vacancy. The inflow of job offers to unemployed workers is a stationary Poisson process with the arrival rate λ. Employed workers may search on the job. Workers are homogeneous. Let the arrival rate of job offers to employed workers be also λ. Jobs are also destroyed for exogenous reasons at rate δ. Equilibrium unemployment rate is then $u = \dfrac{\delta}{\delta + \lambda}$.

The fraction of workers receiving wage w or less is given by:

$$G(w; F) = \frac{\delta F(w)}{\delta + \lambda(1 - F(w))}, \tag{2.53}$$

where $F(w)$ is the distribution of wage offers, which is to be endogenously determined. The separation rate of workers in a firm paying a wage w is $\delta+\lambda(1-F(w))$. The first term represents the flow of employed workers into non-employment and the second term is the flow of workers to other firms offering a higher wage than w.

The inflow of workers to a firm paying a wage w is:

$$\frac{\lambda}{M}[u + (1 - u)G(w; F)] = \frac{\delta\lambda}{M[\delta + \lambda(1 - F(w))]}, \tag{2.54}$$

where $\lambda u/M$ is the rate of recruiting from non-employment and $\dfrac{\lambda}{M}(1 - u)G(w; F)$ is the flow of workers from other firms paying less than w.

The steady-state labor supply of workers to a firm is then:

$$L(w;F) = \frac{\delta\lambda}{M[\delta + \lambda(1 - F(w))]^2}. \tag{2.55}$$

Employed workers are identical in productivity. Hence, any worker employed at a firm generates a product worth p. The profits of a firm solve:

$$\pi(w;F) = \frac{\delta\lambda(p - w)}{M[\delta + \lambda(1 - F(w))]^2}. \tag{2.56}$$

Each firm sets a wage to maximize profits. In steady-state profits of firms should be equal irrespective of the wage set to preserve a non-degenerate wage distribution.[3] Suppose that workers value their leisure at b, which would be the reservation wage for non-employed workers. Then the steady-state level of profits is given by:

$$\pi(w;F) = \frac{\delta\lambda(p - b)}{M[\delta + \lambda]^2}. \tag{2.57}$$

This gives solution to the equilibrium wage offer distribution:

$$F(w) = \frac{\delta + \lambda}{\lambda}\left[1 - \sqrt{\frac{p - b}{p - w}}\right], \tag{2.58}$$

and the distribution of observed wages:

$$G(w) = \frac{\delta}{\lambda}\left[\sqrt{\frac{p - b}{p - w}} - 1\right]. \tag{2.59}$$

The expected observed wage in the economy is then:

$$E(w) = \frac{\delta}{\delta + \lambda}b + \frac{\lambda}{\delta + \lambda}p. \tag{2.60}$$

The lowest observed wage in the economy is hence b and the highest is $p - (p - b)\left(\frac{\delta}{\delta + \lambda}\right)^2$.

The phenomenon of wage dispersion has been attracting much attention of many labor economists. Human capital theory attributed these differences in pay to variation in human capital across workers (see the seminal work by Mincer (1974)). However, "... observable worker characteristics that are supposed to account for productivity differences typically explain no more than 30 percent of the variation in compensation across workers..." (Mortensen (2003, p. 1)). Although controlling for firm-specific effects can explain about 70% of wage variation the question still remains: Why firms follow different wage policies and why similar workers within one firm are still paid differently? The seminal paper by Burdett and Mortensen (1998) presents a model where identical workers are still paid differently in equilibrium. The model has been modified in several ways to make predictions about the shape of the wage distribution more consistent with observed data (see e.g. Bontemps, Robin, and van den Berg (2000) and van den Berg and Ridder (1998)); but even in its simplest form the model still sheds much light onto the phenomenon of wage dispersion.

2.4 Conclusion and Empirical Relevance

Models presented in this chapter give an overview of developments in the job search theory. Search theory has been successfully applied for various empirical questions. For example, Burda and Profit (1996) apply the theory of locational search to estimate the matching functions using the Czech data (the rate at which unemployed workers are matched with the available vacancies). Kiefer and Neumann (1979) empirically test the hypothesis of the search theory that the reservation wage is constant over time.

Most interesting are probably the so-called structural models. For example, in the unemployment duration analysis, each parameter of the so-called hazard function (see Section 4.1) $\lambda(1 - F(w^R))$ is estimated. Using the functional relationship between unemployment duration and the hazard function, the reservation wage and other exogenous variables, and wage offer function and observed wages it is possible to estimate separately λ, w^R, and $F(w)$, which are called structural parameters (therefore the notion "structural model"). For identification of the model reservation wages should be observed[4] (or partially observed) and structural form restrictions are applied. Therefore not many datasets allow identification of the structural parameters of the model. A recent example of the structural model can be found in Frijters and van der Klaauw (2006). Authors introduce the nonparticipation option into van den Berg (1990) model (a special case of this model can be found in Section 3.1). The authors use the GSOEP data for the period 1989-1995. The main difference of this empirical work is that the authors explicitly allow for exits into nonparticipation. Frijters and van der Klaauw (2006) define the withdrawals as being coded in the GSOEP data by "maternity leave", "housewife or house-

husband" or "other". The reservation wages of the unemployed individuals are reported in the data (if an individual was unemployed at the date of the interview). Unemployment benefits are partially observed and are comprised of unemployment insurance and unemployment assistance. If not, the authors impute them by regressing the log of benefit on the previous wage and other individual characteristics. Using the GSOEP data the authors aim at identifying the wage offer distribution, the job offer arrival rate, the discount rate, and the instantaneous utility of nonparticipation. The wage offer distribution above the reservation wage can be identified through the accepted post-unemployment wages after the unemployment spell. The tail below the reservation wage cannot be identified (see Flinn and Heckman (1982)). Knowing the distribution of offered wages and the reemployment hazard one could identify the arrival rate (up to certain normalization). The instantaneous utility of nonparticipation can be identified from the length of the unemployment spell until withdrawal from the labor force and the reservation wage before withdrawal. The discount rate can be identified if the reservation wage and its first derivative is observed. The authors find that the wage distribution facing unemployed workers shifts downwards with the unemployment spell. The authors argue that the shift of the wage distribution can be attributed to the loss of skills. Moreover, they find that the fastest loss of skills occurs during the first year of unemployment. Hence, the authors suggest that the measures aimed at halting the loss of skills should be taken during the first year of unemployment.

An important phenomenon, which is yet missing in the nonstationary job search models, in my view, is a tradeoff existing between duration of unemployment and withdrawals from the labor market. In the next chapter I will demonstrate the existence of a tradeoff between unemployment and participation. Hence, a change in a variable which results in reduction in unemployment duration would result in more exits into nonparticipation and therefore reduce the participation rate. This tradeoff poses certain problems for comparison of labor market performance and estimation of survival rates. Empirical methods for correcting the estimated job-finding probability will be discussed in Section 4.2. The relevance of this tradeoff to policy issues is also discussed.

Chapter 3

Nonstationarity in the Theory of Job Search and Withdrawals from the Labor Market

3.1 Theoretical Framework

The model built here is in principle a dynamic variant of McCall (1970) (see Section 2.2.2). Unemployed workers are identical and live forever. They possess the knowledge about the parameters of the wage offer distribution, but they have no information when job offers arrive and what wages are associated with them. Once accepted by a firm, workers must immediately reply (accept the job or decline), so no waiting is allowed. Once the job is rejected it cannot be recalled. Searching involves a direct cost c per period. Hence, by not participating agents can always enjoy the "pure" leisure (without incurring the cost c). The arrival rate is assumed to be declining over time, which is the special case of van den Berg (1990). The probability that an agent receives job offers is given by a Poisson probability distribution:

$$q(m, \tau, \lambda(t)) = \frac{e^{-\lambda(t)\tau}(\lambda(t)\tau)^m}{m!}. \tag{3.1}$$

Searchers discount at a rate $\beta(\tau)$. When an agent receives m job offers, he picks the best one. Define $\tilde{w}_m = \max\{w_1, w_2, ...w_m\}$. The distribution of accepted offers is an extreme value distribution $G(\tilde{w}_m)$ with a density function $g(\tilde{w}_m)$.

The Bellman equation for the optimal value of search can be given as:

$$\Omega(t) - (b-c)\tau =$$
$$\beta(\tau)\Big[\sum_{m=1}^{\infty} q(m,\tau,\lambda(t)) \int_0^{\infty} \max\left[\Omega(t), W(w)\right] dG(\tilde{w}_m) + q(0,\tau,\lambda(t))\Omega(t+\tau)\Big] =$$
$$\beta(\tau)\Big[\sum_{m=1}^{\infty} q(m,\tau,\lambda(t)) \int_0^{\infty} \max\left[0, W(w) - \Omega(t)\right] dG(\tilde{w}_m) +$$
$$q(0,\tau,\lambda(t))\Omega(t+\tau) + \sum_{m=1}^{\infty} q(m,\tau,\lambda(t))\Omega(t)\Big]. \tag{3.2}$$

In continuous time the function of reservation wage can be given as:

$$w^R(t) = b - c + \frac{\lambda(t)}{r} \int_{w^R(t)}^{\infty} (w - w^R(t)) dF(w) + \frac{dw^R(t)}{r \cdot dt}, \tag{3.3}$$

where r stands for the discount rate. It is assumed that the arrival rate declines over time and equals zero at time $\bar{T}$, formally, $\lambda(t \geqslant \bar{T}) = 0$ (in principle $\bar{T}$ could be equal to infinity). If nonparticipation option is unavailable, the lowest possible reservation wage is $b - c$ (the unemployed has the option of rejecting all offers and keep searching). Consider now the case when the unemployed may withdraw from the labor market, i.e. quit searching. In this case, the unemployed will still "earn" the value of leisure, b, (which is the utility of nonparticipation) but without incurring the search cost c. This implies that there is a critical time, denote it as t^* , such that $w^R(t^*) = b$ and $w^R(t > t^*) < w^R(t^*) < w^R(t < t^*)$. Since $w^R(t > t^*) < b$ it is not optimal anymore for a worker to search after t^* and at that time the unemployed worker drops out of the labor market. The *potential* search time is then the maximum amount of time an agent is willing to allocate to his search and is equal to t^* .

Figure 3.1 shows the reservation wage as a function declining over time. Time at which the reservation wage equals the value of leisure, b, is the potential search time. At t^* the worker drops out of the labor market (does not actively search) and receives b. The reservation wage equation can be rewritten as a differential equation:

$$w^R(t) = \max\Big[b, b - c + \frac{\lambda(t)}{r} \int_{w^R(t)}^{\infty} (w - w^R(t)) dF(w) + \frac{dw^R(t)}{r \cdot dt}\Big]. \tag{3.4}$$

The model shows that at t^*, when an agent drops out of the labor force, his reservation wage equals b. Solving the differential equation in 3.3 for $w^R(t)$ and substituting b for $w^R(t^*)$ gives the solution for t^*. The potential search time implies the maximum time an unemployed worker is willing to allocate to his search. This means that after t^* the worker withdraws from the labor

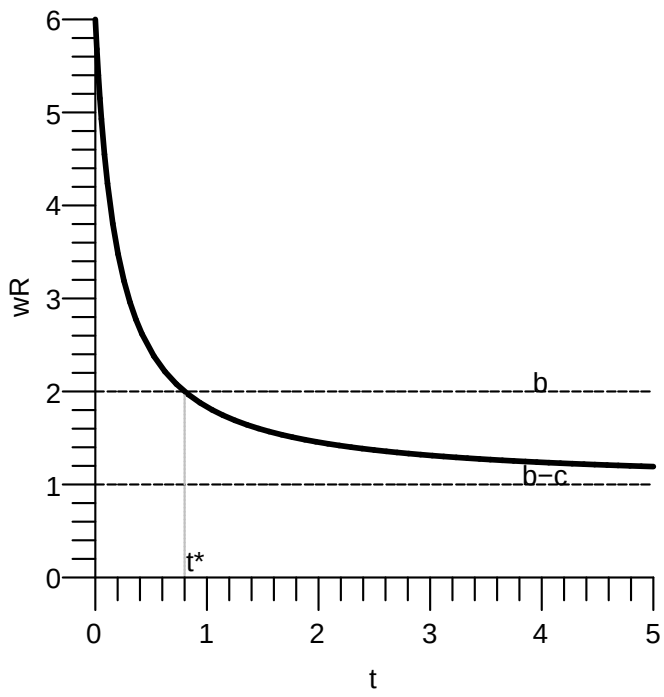


Figure 3.1: Declining Reservation Wage

market under condition that he has not accepted a job within this period. An important result is that the dropout time is not a random variable, it is the *choice*[5] variable for searchers. The potential search time, t^*, is known at any point in time and it does not depend on whether an unemployed worker finds a job or not, but is *observed* only if a worker fails to find employment until time t^*. Looking at Figure 3.1 one can see that exogenous factors which push the reservation wage up also increase the potential search time.

3.2 Unemployment Participation Tradeoff

The instantaneous probability or hazard of transition from unemployment into employment can be written as:

$$\phi(t) = \lambda(t)\Big[1 - F(w^R(t))\Big] \tag{3.5}$$

Measures which reduce reservation wages of the unemployed workers increase the hazard of exit from unemployment into employment. However, as it was already mentioned in Section 3.1, lower reservation wage imply shorter potential search. Consider a group of identical workers (call them group 1) with a search cost c_1. These workers have identical reservation wages w_1^R, and the length of potential search for them is t_1^*. The fraction of exits from unemployment

into employment is $1-\exp\left(-\int_0^{t_1^*}\phi(t)dt\right)$. Consider a second group of identical workers of the same size for whom the search cost is c_2, such that $c_1 < c_2$, so that $w_1^R > w_2^R$ and $t_1^* > t_2^*$. The second group has a higher hazard rate (of exiting from unemployment into employment) for a given point in time as $w_1^R > w_2^R$. However, this does not necessarily mean that the *absolute number* of exits from unemployment into employment of the second group is larger. Workers with reservation wages w_2^R who have not accepted any job until t_2^* withdraw from the labor market. Workers with reservation wages w_1^R do not drop out at t_2^* but continue their search until t_1^*. Whether the number of transitions into employment during $t_1^* - t_2^*$ may compensate the difference in the number of exits from unemployment into employment during t_2^* is ambiguous. Figures 3.2 and 3.3 show possible scenarios when reservation wages change.

Figure 3.2 shows the first possible scenario. The reservation wage of the first group is w_1^R (dashed curve) and the potential search time t_1^*, and the share of exits into employment is P_1. The reservation wages of the second group is w_2^R (solid curve), the potential search time is t_2^* but the share of exits into employment is P_2, which is greater than P_1. Figure 3.3 shows the second possible scenario. Here the potential search of the second group is t_2^*, such that the share of exits into employment P_2 is smaller than P_1.

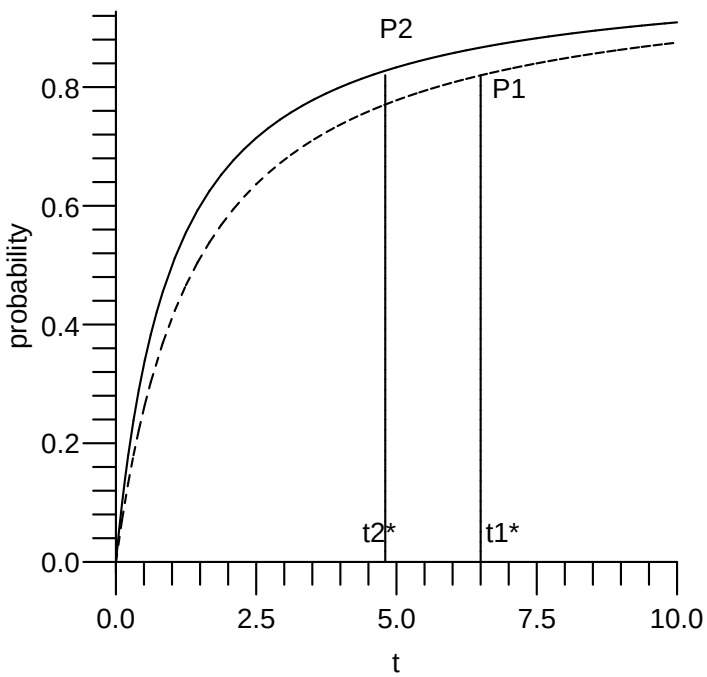


Figure 3.2: The tradeoff between the job-finding hazard and potential search time: Scenario 1

This tradeoff between the probability of finding a job and the length of potential search poses certain problems when one tries to compare performances of different labor markets. If we look at Scenario 2 and consider only unemployment durations then we might get an impression that

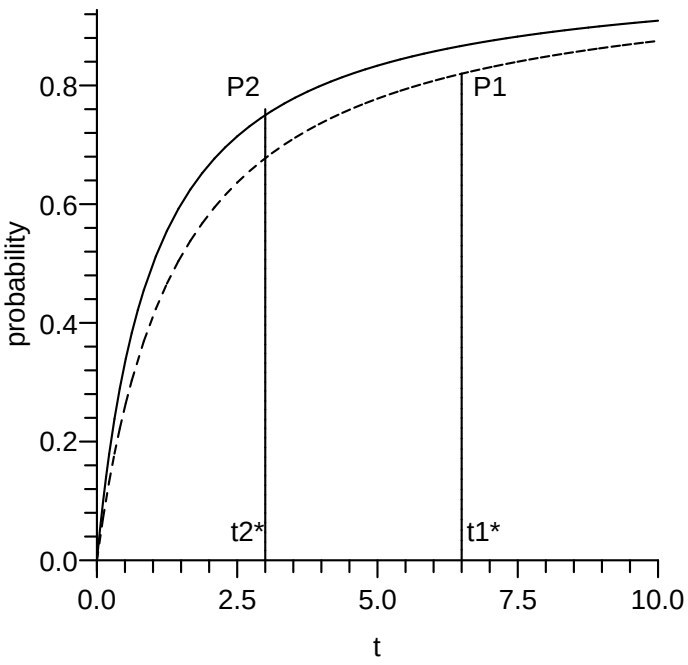


Figure 3.3: The tradeoff between the job-finding hazard and potential search time: Scenario 2

group 2 have better job-finding chances as the reservation wage of the group 2 is always above that of the group 1 for any fixed t. However, once we incorporate the change in potential search time, we could see that the overall job-finding probability of group 2 is lower.

3.3 Simulations

For the numerical simulations the function of the arrival rate was chosen to be linear in time and be of the form:

$$\lambda(t) = \begin{cases} \lambda - t \cdot k \text{ if } t \cdot k < \lambda \\ 0 \text{ otherwise} \end{cases}$$

Consider a point T, such that $T \cdot k = \lambda$ and $\lambda(T) = 0$. The parameter values were chosen to be: $\lambda = 0.018$, $k = 0.01$, $T = 18$ months. Moreover, the reservation wage function would have the following form:

$$w^R(t) = \max\Big[b, b + \alpha - c + \frac{\lambda(t)}{r} \int_{w^R(t)}^{\infty} (w - w^R(t)) dF(w) + \frac{dw^R(t)}{r \cdot dt}\Big], \tag{3.6}$$

where α is the unemployment benefit which the unemployed worker receives while searching but not in the case of nonparticipation. The wage offers were set to be uniformly distributed for

the sake of simplicity with the infimum $w_L = 400$ and supremum $w_U = 5000$. Utility of non-participation and search costs are chosen to be equal to 400 each. The amount of unemployment insurance benefits (UI) was chosen to be equal 300 in the base model, in Model 1, UI = 350 and in Model 2, UI = 200. Reservation wage paths of three models are shown on Figure 3.4. In Model 3, all parameters were chosen to be the same as in the base model except for the arrival rate which is equal to 0.02 in Model 3. Theoretical considerations in Section 3.1 would imply that $w^R(T) = b - c$ which would serve as an initial condition for solving the differential equation 3.6. For each model the potential search time and the cumulative job finding probability at the time of withdrawal were calculated, which can bee seen in the table below.

Table 3.1: Parameters and results of the simulated models

Model	Unemployment insurance (UI)	Arrival rate (λ)	Potential search time	Cumulative job-find. prob.
Base	300	0.018	7.10	0.64
1	350	0.018	10.70	0.74
2	200	0.018	1.50	0.24
3	300	0.020	9.10	0.76

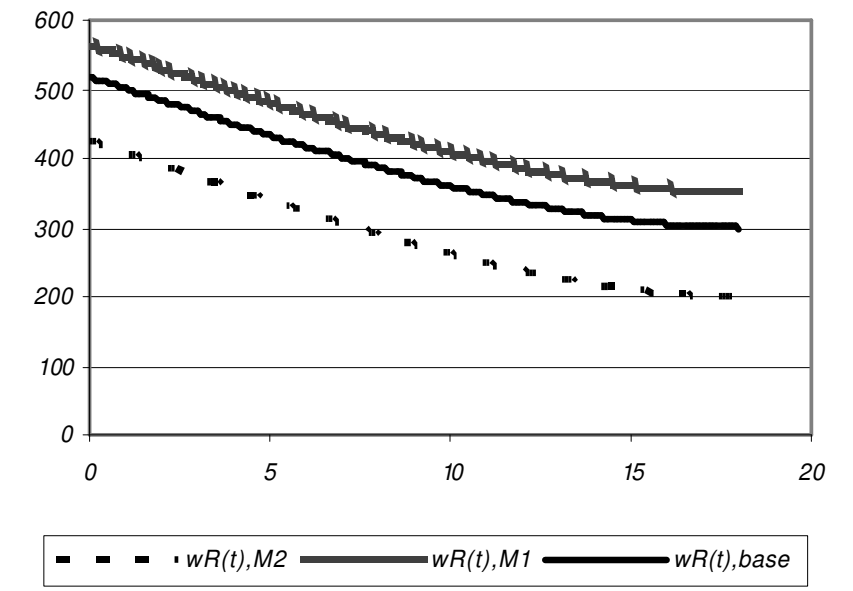


Figure 3.4: Reservation wage paths

It is apparent from Figure 3.1 that an increase in the unemployment insurance benefits pushes reservation wages up. However, Table 3.1 shows that the longer potential search induced by an increase in the unemployment insurance benefits more than compensates the negative effect

of higher reservation wages on employment prospects. These results are partially in accord with the empirical results of Frijters and van der Klaauw (2006). These authors find that a reduction of unemployment benefits by 50% increases the percentage of individuals finding a job within 12 months from 60.3 to 66.5% which is statistically insignificant. However, the estimated effect of the benefits on the probability of an exit into nonparticipation is positive (a 50% reduction of unemployment benefits reduced the share of exits into nonparticipation from 6.1 to 5.2%). This seems to be in discord with the theory. The problem could have arisen from the definition of the unemployment insurance benefits. The definition of benefits in Frijters and van der Klaauw (2006) covered the unemployment insurance benefits and unemployment assistance (which is of unlimited duration). Recalling Equation 3.6 it seems more plausible that unemployment insurance benefits affect only α and unemployment assistance affects the utility of nonparticipation.

Concerning policy implications, the simulation results show that reducing the level of unemployment insurance benefits may even worsen the situation due to a higher transition rate into nonparticipation. Further work is required to identify the set of parameters which make unemployment drop with higher unemployment compensation.

Chapter 4

Empirical Estimation of Duration Models

4.1 Estimation Methods

Due to a tradeoff between unemployment and participation, as it was shown in Chapter 4.1.3, unemployment duration or survival rate might not always be good indicators of labor market performance. The empirical model presented in this chapter shows that care should be taken when interpreting the results of a duration model estimation as the conditional job-finding probability estimated in a duration model may be significantly different from an unconditional job-finding probability when the dropout rate is non-negligible.

The theoretical model set up in this work gives insight into the search process of workers. The exogenous factors determine the reservation wage and thus the probability of finding a job. Moreover, it has been shown that the parameters of the model affect the time an unemployed spends looking for a job. Unfortunately however, there is no closed-form solution for the reservation wage, as it involves solving integrals with variable limits of integration. Therefore, we cannot determine the elasticities of reservation wage with respect to the set of exogenous parameters analytically.

An offer is accepted only when the offered wage exceeds the reservation wage. Hence, the probability that a worker is ”matched” at time t is given by:

$$\sum_{m=0}^{\infty} \frac{(\lambda(t)t)^m e^{-\lambda(t)t}}{m!} \Big(1 - F^m(w^R)\Big) \tag{4.1}$$

One could see that when no application is successful, i.e. $m = 0$, the probability in Equation 4.1 is zero. Equation 4.1 states that the probability of a match goes up with the arrival rate and goes down with the reservation wage. Hence, without knowing the respective elasticities, the sign of the effect of any parameter on the probability of a match cannot be determined. For instance, when the arrival rate in the economy rises, agents will increase their reservation wages at the same time, so they become pickier and could search longer until they have found a suitable job.[6]
.

Simplifying Equation 4.1 yields:

$$\Pr(T < t) = 1 - e^{-\lambda(t)t\Big(1 - F(w^R)\Big)}, \tag{4.2}$$

where $\Pr(T < t)$ means that the time when an agent finds a job, T, happens before t. Of course, if the solution for the reservation wage existed in closed form or could be observed, estimation of Equation 4.2 would have been straightforward. However, many levels of recursion and nonlinearities make it impractical for empirical applications.

Before going into explanation of the estimation methods it would be helpful to introduce following definitions and notations: $\Phi(t) = \Pr(T < t)$. In duration analysis this is called a failure function, which specifies the probability that the spell ”fails” before time t. In our case, the ”failure” is a change of state for a worker from ”unemployed” to ”employed”. The survival (or survivor) function, $S(t) = 1 - \Phi(t)$, is the probability that the ”failure” does not occur before time t. The hazard function, $\phi(t) = \dfrac{d\Phi(t)/dt}{S(t)}$, is ”... the rate at which spells will be completed at duration t, given that they last until t” (Kiefer (1988, p. 651)).[7] A tractable approach to tackle the unemployment model empirically is to define the hazard function as:

$$\phi(t) = \lambda(t)\Big(1 - F(w^R)\Big). \tag{4.3}$$

Then one must specify the functional relationship between the hazard function and the set of exogenous parameters. A convenient starting point is the exponential specification. As Kiefer puts it: ”The exponential [distribution] is simple to work with and to interpret, and is often an adequate model for durations that do not exhibit much variation (in much the same way that the linear regression model is simple and adequate if the data do not vary enough to reveal important nonlinearities).” (Kiefer (1988, p. 552))[8] Other specifications are considered in the next sections: Section 4.1.1 describes the exponential and Weibull distributions, Section 4.1.2 gives an overview of the Cox method, Section 4.1.4 describes the Kaplan-Meier product-limit estimator.

4.1.1 Parametric Methods

Exponential Distribution

When the duration distribution is specified the model could be estimated by the method of Maximum Likelihood. For an exponential distribution the hazard function can be given as:

$$\phi(x, \beta) = e^{x'\beta} \tag{4.4}$$

The expected duration, $E(t)$ is $e^{-x'\beta}$. The log-likelihood function can be given as:

$$\ln L = \sum d_i x_i' \beta - \sum t_i e^{x_i'\beta}, \tag{4.5}$$

where $d = 0$ if an observation is censored and $d = 1$ if it is uncensored. The respective derivatives of the likelihood function are:

$$\begin{aligned} \frac{\partial \ln L}{\partial \beta} &= \sum d_i x_i' - \sum t_i e^{x_i'\beta} x_i' \\ \frac{\partial^2 \ln L}{\partial \beta \partial \beta'} &= -\sum t_i e^{x_i'\beta} x_i x_i'. \end{aligned} \tag{4.6}$$

One could see that the log-likelihood function is concave, so numerical maximization is straightforward to apply.

Weibull Distribution

The Weibull model specifies the hazard function as:

$$\phi(\lambda, t, \beta) = \lambda\alpha(\lambda t)^{\alpha-1}. \tag{4.7}$$

To facilitate estimation the following transformation would be useful: $(t\lambda)^\alpha = e^\omega$. Then the distribution function for the Weibull is: $F(\omega) = 1 - \exp(-e^\omega)$. Letting $\lambda = \exp(x'\beta)$ and $\alpha = 1/\sigma$:

$$\omega_i = \frac{\ln t_i - x_i\beta}{\sigma}. \tag{4.8}$$

It is obvious from Equation 4.7 that the Weibull distribution is simply a generalization of the exponential distribution, where the latter is a special case of the Weibull distribution when $\alpha = 1$.

The observed random variable is:

$$\ln t_i = \sigma\omega_i + x_i\beta. \tag{4.9}$$

The log-likelihood can be given by:

$$\ln L = \sum_i \Big[d_i \Big(\frac{\ln t_i - x_i\beta}{\sigma} - \ln\sigma \Big) - \exp\Big(\frac{\ln t_i - x_i\beta}{\sigma} \Big) \Big], \tag{4.10}$$

where $d_i = 0$ if an observation is censored and $d_i = 1$ otherwise.

4.1.2 Proportional Hazard Specification and Semiparametric Estimation

Proportional hazard models, namely the Cox regression (developed by Cox(1972, 1975)), are popular in econometrics of duration data as the baseline hazard needs not to be specified which makes the model rather flexible. In this model the hazard function is specified as:

$$\phi(t, x, \beta, \phi_0) = \lambda(x, \beta)\phi_0(t), \tag{4.11}$$

where ϕ_0 is the "baseline" hazard, corresponding to $\lambda(x, \beta) = 1$. Specifying the arrival rate as $\lambda(x, \beta) = \exp(x'\beta)$, yields:

$$\phi(t, x, \beta, \phi_0) = e^{x'\beta}\phi_0(t). \tag{4.12}$$

The algorithm of the Cox regression can be explained as follows. The completed durations are ordered, $t_1 < t_2 < \ldots < t_n$ [9] the conditional probability that observation 1 "fails" at duration t_1, given that any of the n observations could have been concluded at t_1 is $\frac{\phi(t_1, x_1, \beta)}{\sum_{i=1}^n \phi(t_1, x_i, \beta)}$. In the proportional hazard model this expression reduces to:

$$\frac{\lambda(x_1, \beta)}{\sum_{i=1}^n \lambda(x_i, \beta)}, \tag{4.13}$$

which 4.13 is the contribution of the shortest spell to the likelihood function. In the same fashion, the contribution of the jth duration (given that the spells are ordered as described above) is $\frac{\lambda(x_j, \beta)}{\sum_{i=j}^n \lambda(x_i, \beta)}$. The likelihood function is then:

$$\ln L(\beta) = \sum_{i=1}^n \Big(\ln\lambda(x_i, \beta) - \ln\Big[\sum_{j=1}^n \lambda(x_j, \beta) \Big] \Big). \tag{4.14}$$

4.1.3 Competing Risks

Competing risks imply that there are several possible ways in which the spell can end or several possible destination states from the state where the person or process dwells. For example, in medical research the treatment of the disease may end with recovery or death, in political science the legislative term (for example in parliament) may end due to retirement or assumption of another public office (be it a minister or a governor). In terms of the theoretical model presented in Chapter the unemployed worker may become employed or withdraw from the labor force. The conventional and most widely used method of estimating the competing risk models is to assume independence of risks. In this case, one first estimates the model treating the failures due to *risk 1* as uncensored, considering the failures due to *risk 2* as censored. Secondly, one estimates the model with failures due to *risk 2* as uncensored, treating the failures due to *risk 1* as censored. This could seem quite strong an assumption. However, for conventional estimation techniques the *conditional* independence suffices, i.e. independence is assumed after having controlled for all the covariates. Put it another way, the errors in the first and the second regression must be uncorrelated. If the assumption of independence of errors is fulfilled, the competing risk model can be handled as a single risk model with other risks treated as censored (see among others Katz (1986), Gonzalo and Saarela (2000), Qian and Correa (2003), Wolff (2003), Wolff and Trübswetter (2003), Fitzenberger and Wilke (2004), Wolff (2004)).

The dependent competing risks are still a field with white spots as identification arises as a problem, since once a person leaves his current state via risk i, then the duration until the exit via risk j is not observed. Although much work has been done to identify the model, overall identifiability is still debatable. Hausman and Han (1990) devised methods to handle the dependent competing risks. They estimated the full model with a bivariate error distribution. They claim the model is identified only when there are as many continuous regressors as there are competing risks. Their likelihood function had to be maximized with respect to exogenous parameters β_1 and β_2, respective variances σ_1 and σ_2, the coefficient of correlation ρ, and the unobserved mean durations $\bar{t}_1$ and $\bar{t}_2$.

One of the approaches to handle the competing risks is to allow the correlation of errors due to unobserved heterogeneity. Hence, once unobserved heterogeneity is accounted for, the errors are not correlated anymore (see Hougaard (1987) for an overview). These are also known as *frailty* models. For example, using French data van den Berg, van Lomwel, and van Ours (2003) estimated a model with independent errors but correlated heterogeneity terms. However, they could not reject the hypothesis of heterogeneity independence. Gordon (2002) used generalized dependent risk model of government office duration but also commented that the generalized dependent risk model did not offer improvement over a stochastically independent risk model.

4.1.4 Nonparametric Methods: Kaplan-Meier Product-Limit Estimator

Nonparametric methods does not use underlying theory or restrict the model to any specific distribution. They are more ”data-suited” than ”theory-suited”. However, graphical representation of nonparametric estimates of survivor function could be a good starting point before going into parametric methods as it may give some hints about the functional form of the hazard function. For the sample of n observations with no censoring the survivor function is:

$$\hat{S}(t) = n^{-1}. \tag{4.15}$$

When censoring is present Equation 4.15 should be modified. We need to order *completed* (uncensored) durations in the sample in such a way that $t_1 < t_2 < \ldots < t_k$ (from the shortest to the longest), with $K < n$ if at least one observation is censored (the spell is not completed at the calendar time the study terminates) or if there are ties (two or more observations have the same duration). Define h_j as be the number of completed spells of duration t_j, for $j = 1, \ldots, K$. Let m_j be the number of censored observations between t_j and t_{j+1}; and let n_j be the number of spells either completed or censored after duration t_j.

$$n_j = \sum_{i \geqslant j}^{K} (m_i + h_i). \tag{4.16}$$

The hazard $\phi(t_j)$ is the probability that a spell ends at t_j, given that it lasted until t_j. The estimator for $\phi(t_j)$ is:

$$\hat{\phi}(t_j) = h_j / n_j, \tag{4.17}$$

which is the number of ”failures” at duration t_j divided by the number of ”potential failures” at duration t_j. The estimator of the survivor function is:

$$\hat{S}(t_j) = \prod_{i=1}^{j} \frac{n_i - h_i}{n_i} = \prod_{i=1}^{j} (1 - \hat{\phi}_i), \tag{4.18}$$

which is called the Kaplan-Meier (due to Kaplan and Meier (1958)) or *product-limit* estimator. ”Essentially, this estimator is obtained by setting the estimated conditional probability of completing a spell at t_j equal to the observed relative frequency of completion at t_j” (Kiefer (1988, p. 659)). One could see from Equation 4.18 that the estimator does not involve any explanatory variables and does not account for heterogeneity of the data. Hence to apply the product-limit estimation, the data should be split into homogenous groups.

4.2 The Kaplan-Meier Estimator and Withdrawals from the Labor Market

Consider the next simulation scenario. I generated 1000 spells of 1000 unemployed individuals all of whom eventually find employment, i.e. all spells are uncensored. The job finding probability (I would call it the counterfactual here) is represented by the solid curve in Figure 4.1. Now suppose that for a group of workers who find jobs between $t = 508$ and $t = 585$ the potential search time, t^* is 508. These workers have not found jobs until $t = 508$, hence, they withdraw from the labor market at this point in time (in this artificial data there are about 250 workers who have not found job until $t = 508$ but *would have found* it at $508 < t < 585$ if they had searched longer). If we treat these observations as censored the estimated job finding probability would correspond to dashed curve in Figure 4.1 (Kaplan-Meier estimates). However, it does not correspond to the true job finding probability. The real cumulative job finding probability is about 75 percent, since some 250 unemployed workers out of 1000 withdraw from the labor market and *never* become employed (represented by the dotted curve in Figure 4.1). But if we look at the dashed line, the estimated probability equals one at highest value of t. This is because duration models estimate the conditional job finding probability.

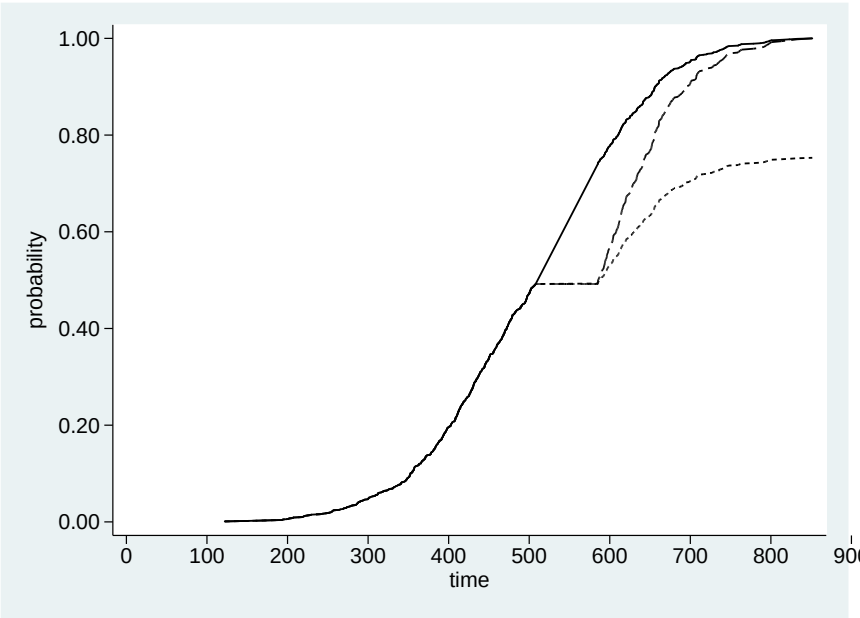


Figure 4.1: Job-finding Probability Simulation

Recall the Kaplan-Meier estimator of the survivor function given in Equation 4.18. If an individual withdraws from the labor market at t_j then at t_j this observation is treated as censored and at t_{j+1} this individual is no more ”at risk”. As a result the more individuals withdraw the lower is n_j at large t and hence lower the survival rate and higher the job finding probability.

If we do account for the dropouts, i.e. consider them ”at risk” then the job finding probability

would be represented by the dashed curve in Figure 4.1, which in this case overestimates the real share of job findings as it estimates the *conditional* job finding probability, in this case - conditional upon staying in the labor market and not withdrawing. In a situation where dropouts are not negligible it may be advisable to use the unconditional job finding probabilities as a basis of comparison of the labor market performance.

To illustrate the idea consider the share of job findings at time t to be given as:

$$p(t) = \frac{E_t}{U_t + E_t + D_t}, \tag{4.19}$$

where E_t denotes the number of workers who find jobs at time t, U_t - who remain unemployed, and D_t - who withdraw from the labor market. The survival rate at time T is then given by:

$$S(T) = \prod_{t=0}^{T}(1 - p(t)), \tag{4.20}$$

and the job finding rate: $1 - S(T)$. However, if we estimate the duration model using conventional methods the denominator in Equation 4.19 would be $U_t + E_t$ as the withdrawn workers, D_t, would not be considered 'at risk'.

When we estimate the duration model using the Kaplan-Meier method, the survival rate is given by:

$$S(T)^c = \prod_{t=0}^{T}(1 - p(t)^c)), \tag{4.21}$$

with $p(t)^c = \dfrac{E_t}{U_t + E_t}$ being a conditional probability of finding a job. The unconditional probability of employment at time t can be given as:

$$p(t) = p(t)^c(1 - p(t)^d), \tag{4.22}$$

where $p(t)^d$ is the unconditional probability of withdrawing from the labor market. The unconditional dropout probability can be given as:

$$p(t)^d = p(t)^{dc}(1 - p(t)), \tag{4.23}$$

with $p(t)^{dc} = \dfrac{D_t}{U_t + D_t}$ being a conditional probability of dropout. It is easily verifiable that solving the system of Equations 4.22 and 4.23 for $p(t)$ would yield 4.19.

To find the unconditional job finding probability we would just need to recode the end of spells for the withdrawing individuals to $t = \infty$, i.e. they never find jobs.[10] For practical reasons it suffices to replace the end of spell for the withdrawing individual with $t = \max(t)$.

For illustration I estimated conditional and unconditional job finding probabilities for unemployed German workers. For empirical estimation the 1975-2001 data sample of the IABS dataset has been used (description see in Appendix A.12 and Bender, Haas, and Klose (2000)). The analysis is restricted to unemployed females. The comparison is drawn between four groups: singles in West Germany, married in the West, singles in East Germany, and married in the East. The data contain the information when the unemployment spells ended with an employment begin. As a proxy for withdrawals from the labor market the coding ”not available to the labor market” has been used. If an unemployment spell is interrupted for less than 30 days (during which the unemployed worker is not observed in the sample) and after that is coded again as unemployed the two broken spells are considered as one continuous spell (see Fitzenberger and Wilke (2004)). If the spell ended in a way other than employment or withdrawal (for example, the entitlement period for unemployment insurance benefits ended) it is considered as censored. Naturally the unconditional job finding probabilities are lower than the conditional ones. Figures 4.2 - 4.5 show the estimated conditional and unconditional probabilities.

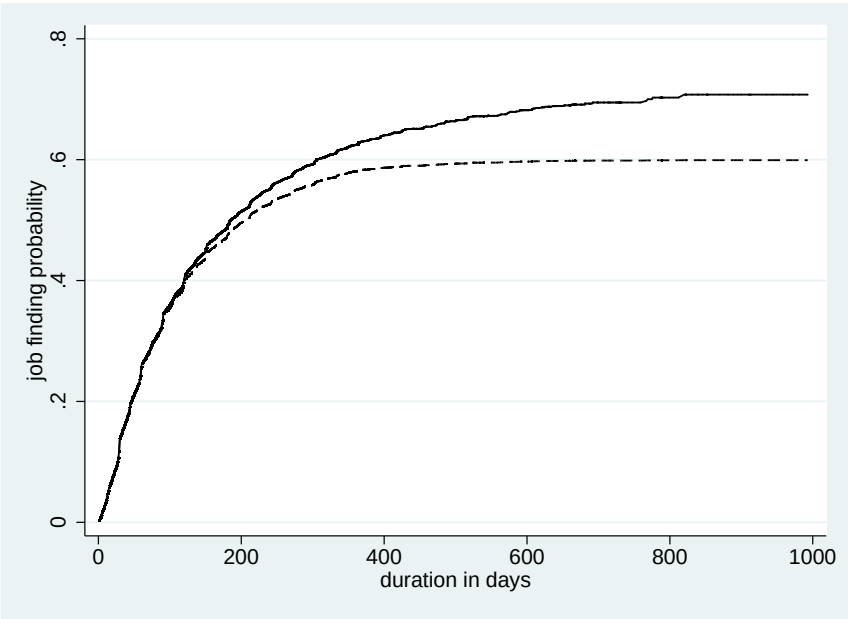


Figure 4.2: Single females, West Germany. Solid line - conditional, dashed line - unconditional job finding probability.

The conditional and unconditional probabilities do not differ much with duration of unemployment less than a year. This implies that in the first year of unemployment the withdrawal rate is negligible. However, after about one year of unemployment duration the discrepancy between the conditional and unconditional probabilities grows as more and more unemployed withdraw from the labor market. Comparing Figures 4.4 and 4.2 we could see that single females in West Germany have higher employment prospects than their Eastern counterparts. After one year the cumulative job finding probability is about 0.6 for females in the West and about 0.4 for females

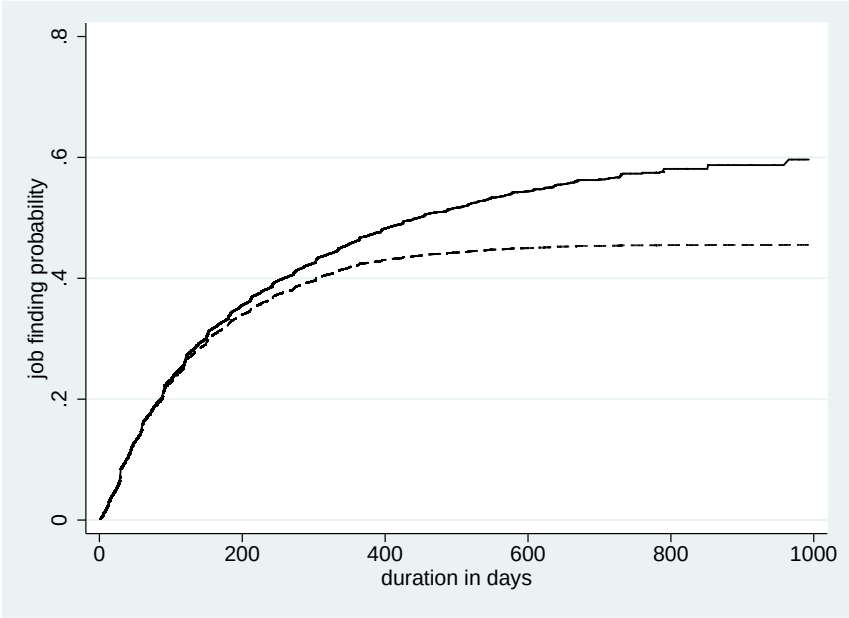


Figure 4.3: Married females, West Germany. Solid line - conditional, dashed line - unconditional job finding probability.

in the East. If we look at Figures 4.5 and 4.3 we could see that the withdrawal rate for married females in the East is greater than for their Western counterparts (this could be seen by a larger gap between the estimated conditional and unconditional probabilities). Higher withdrawal rate in the East indicates that the returns to search relative to the utility of nonparticipation are lower in the East. This could be driven by many factors which can not be answered here (lower wages, lower arrival rate due to fewer vacancies, faster skill loss, higher utility of nonparticipation). If we look at the conditional job finding probability for married females in East Germany, we get an impression that after a year of unemployment almost a 30 percentage point improvement in job finding probability is possible (from about 0.4 at $t = 365$ to about 0.7 at $t = 900$). However, in reality (if we look at the unconditional job finding probability) after about 400 days of unemployment no more improvement is possible. This implies that the increase in the cumulative job finding probability after about 400 days is artificially driven by a diminishing number of persons ”at risk”.

In the light of the theoretical model, the estimation results tell us that the potential search time, t^*, for many female workers is about 400 days. After 400 days of unemployment their employment chances become low, the returns to search do not cover the search cost anymore, and they withdraw from the labor market which results in a flat job finding probability after t^*. This could be for example driven by skill depreciation, which would be in accordance with findings of Frijters and van der Klaauw (2006) who find that the fastest loss of skills occurs in the first year of unemployment. If this is true, this could mean that after a year of unemployment skills depreciate so much that employment chances become very low and unemployed workers quit

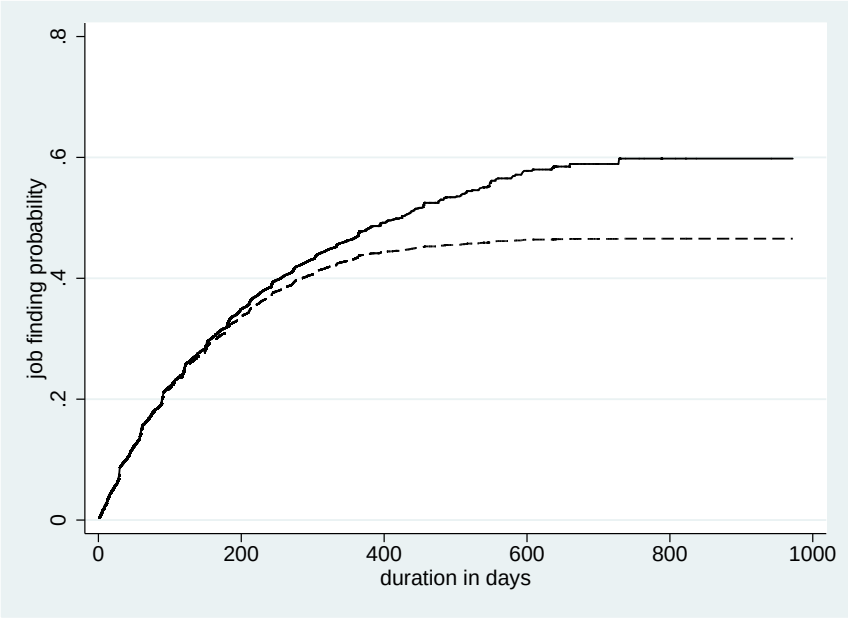


Figure 4.4: Single females, East Germany. Solid line - conditional, dashed line - unconditional job finding probability.

the labor market. This implies that the gain from active labor market policies that halt a skill loss (or improve skills) is greater with an early intervention than with a late one.

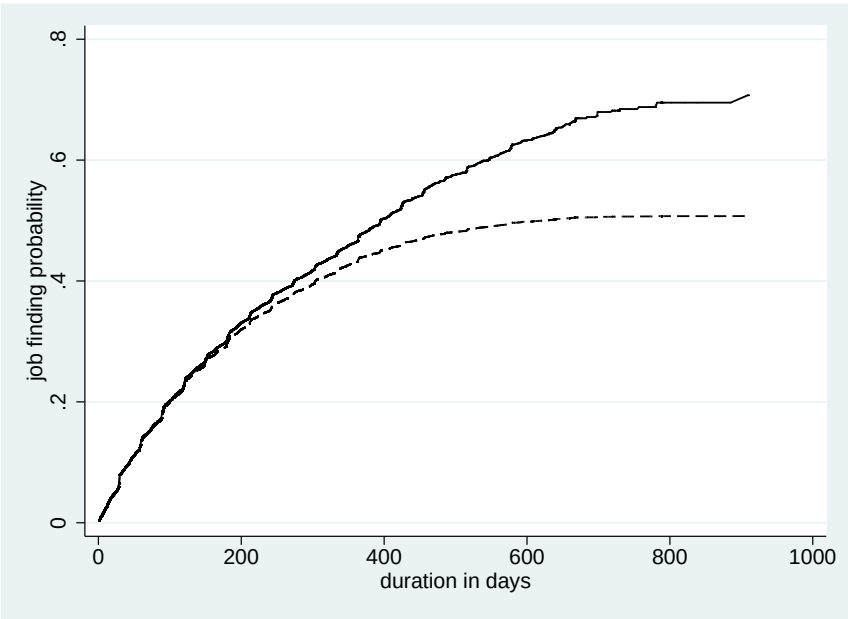


Figure 4.5: Married females, East Germany. Solid line - conditional, dashed line - unconditional job finding probability.

Chapter 5

Spatial Search Theory and Commuting

5.1 Introduction

Interregional mobility has long been regarded as an important adjustment mechanism equilibrating regional disparities in wages and unemployment. Most of the works dealing with mobility concentrate on migration thereby belittling the role of commuting in interregional interaction. Interregional commuting, however, is an important process regulating the functioning of regional labor markets in many modern economies. In Germany, for example, about 30% of employed individuals in 1997 worked at locations other than their place of residence (with locations defined as NUTS-3 regions).

Another problem in studying the commuting process is that most of the works on commuting are purely empirical (for example Raphael (1998)) or use the unilocational search models, i.e. without explicitly allowing for search across locations (e.g. Eliasson, Lindgren, and Westerlund (2003)). It seems natural to build commuting models on search theory.[11] However, models explaining search of the unemployed workers for jobs across spatial units are rare. Some of them like Burda and Profit (1996) use the ”urn” principle in tradition of Pissarides (1979) where the distribution of wage offers is degenerate (see Section 2.1.3). In this setting there is no reservation wage property because all jobs are equally paid. The search strategy is then simple - accept the first offer or do not participate at all. In Burda and Profit (1996), however, agents optimize their search intensity.

Molho (2001) allows non-identical non-degenerate wage distributions at different locations and exogenous separations. The similar approach with minor modifications has been adopted by

Damm and Rosholm (2003) and Arntz (2005). All of those models deal with the search-migrate decision. There are several works on commuting which are based on spatial search theoretical models (see e.g. van Ommeren and van der Straaten (2005)). Unlike in the classical search models, which assume wage dispersion, van Ommeren and van der Straaten (2005) allow for constant wages but dispersion of distances to work. Unemployed individuals in their model solve for maximal acceptable travel distance.

In this chapter, I allow for wage dispersion across regions and travel costs between regions. Hence, reservation wage is not unique to an individual, i.e. agent sets different reservation wages for the ”local” and ”distant” region. Moreover, I allow for arrival rate to be the function of search intensity and therefore unemployed workers receive job offers only in those regions where their search intensity is nonzero. In this setting the maximal acceptable travel distance is determined by the condition of nonzero search intensity. It will be shown that exogenous factors in the home region affect the reservation wage and search intensity in the ”distant” region and vice versa. The model shows that the circle of possible commuting destinations is bounded and the maximum distance a worker is willing to commute is determined by exogenous factors in the model. It is also shown that the maximal travel distance condition involves a selection mechanism which is further addressed in the empirical model.

Most importantly, I address in this chapter the problem of traditional search-theoretical approach when the wage offer distribution is asymmetric and dispersions in the two tails of the distribution may vary independently of one another. The McCall model predicts increase in the reservation wage with the mean and the mean-preserving spread of the wage offer distribution (see for details Mortensen (1986)). However, the mean-preserving spread, i.e. changing the spread holding the mean constant implies a symmetric stretching or compression of the wage distribution which is problematic in the empirical context. If the wage distribution is not symmetric and variances in the left tail and in the right tail are allowed to change independently then the mean-preserving spread is not an adequate measure anymore (asymmetric changes of the dispersion in the left and right tail will change the mean as well). A good solution to this problem could be using median as a location parameter of the distribution and the median-preserving spread in the left tail as a scale parameter for the left tail and the median-preserving spread in the right tail as a scale parameter for the right tail of the wage distribution (see Möller and Aldashev (2006b) who first point to that problem). It will be shown that if the wage distribution is not symmetric and variance in two tails of the wages distribution can change independently of one another, the implications of the search theory slightly change. Namely, the dispersion in the left tail of the wage distribution reduces reservation wage and search intensity, and the dispersion in the right tail increases reservation wage and search intensity. As a consequence, in my empirical model I include dispersion in the left tail and the right tail of the wage distribution as separate additional regressors.

The numerical simulations show that increasing wages in the region of origin and destination by the same amount will increase reservation wages and search intensities in both regions. Therefore, I use wages in destination and origin as separate regressors, not a ratio of wages in the destination to the origin as it is usually done in the tradition of gravity models (see for example Hatton and Williamson (2002), Mitchell and Pain (2003)).

5.2 Bilocational Search

Assume for simplicity that there are only two locations in the economy: place of residence and a distant region (extension to a multilocational model is straightforward). Throughout the paper I will use the terms region A to denote the local labor market and region B to denote the distant labor market. I consider here only job search decisions of a resident of region A as decisions of residents of B are derived in a likewise manner. Offers to work in region A and B arrive to the searcher according to a stationary Poisson process with the arrival rate $\lambda_A(\theta_A)$ and $\lambda_B(\theta_B)$, where θ_A is the intensity with which a resident of region A searches for jobs in the local labor market and θ_B is the search intensity of a resident of region A in distant labor market, which are to be determined endogenously. The arrival rates satisfy the following properties: $\lambda'_A(\theta_A) > 0$; $\lambda''_A(\theta_A) < 0$ and $\lambda'_B(\theta_B) > 0$; $\lambda''_B(\theta_B) < 0$. Searching in each region involves a search cost which is a function of search intensity. The cost functions satisfy the following properties: $c'_A(\theta_A) > 0$; $c''_A(\theta_A) > 0$ and $c'_B(\theta_B) > 0$; $c''_B(\theta_B) > 0$.

The probability that an agent receives n job offers in region A during period τ is given as:

$$q_A(n,\tau) = \frac{e^{-\lambda_A(\theta_A)\tau}(\lambda_A(\theta_A)\tau)^n}{n!}. \tag{5.1}$$

In the same fashion, the probability that an agent receives n job offers in region B can be written as:

$$q_B(n,\tau) = \frac{e^{-\lambda_B(\theta_B)\tau}(\lambda_B(\theta_B)\tau)^n}{n!}. \tag{5.2}$$

The distributions of job offers in region A and B are denoted as $F_A(w)$ and $F_B(w)$ respectively and are exogenously given. If the searcher receives more than one job offer in both regions, he picks the best one. Let $G(w;n,m)$ be the distribution of maximal *net* [12] wages from n offers drawn from $F_A(w)$ and m from $F_B(w)$. The value of search for an individual residing in region A can be given as:

$$\Omega = (b - c_A(\theta_A) - c_B(\theta_B))\tau + \beta(\tau) \times \tag{5.3}$$
$$\Big[\sum_{n=1}^{\infty}\sum_{m=1}^{\infty} q_A(m,\tau)q_B(n,\tau)\int_0^{\infty}\max(W(w),\Omega)dG(w;n,m)+$$
$$\sum_{n=1}^{\infty} q_A(n,\tau)q_B(0,\tau)\int_0^{\infty}\max(W(w),\Omega)dG_A(w;n)+$$
$$\sum_{n=1}^{\infty} q_B(n,\tau)q_A(0,\tau)\int_0^{\infty}\max(W(w)-\delta/r,\Omega)dG_B(w;n)+$$
$$q_B(0,\tau)q_A(0,\tau)\Omega\Big],$$

where $\beta(\tau)$ is the discount factor which is equal to e^{-rt} and δ is the cost of commuting per period. The third term in 5.3 implies that if an agent does not get any offer in region B (with probability $q_B(0,\tau)$, he picks the best offer (out of n) in region A (given that they exceed the value of search), so $G_A(w;n)$ is the distribution of maximal wages from n offers drawn from $F_A(w)$. Using the same logic, the forth term means that if an agent does not get any offer in region A (with probability $q_A(0,\tau)$), he picks the best offer in region B, so $G_B(w;n)$ is the distribution of maximal wages from n offers drawn from $F_B(w)$. The second term in 5.3 implies that if an agent receives more than one offer in each region, he picks the offer with the highest wage. The last term in 5.3 means that the unemployed workers continue the search if they do not receive any offers both in region A and B.

It can be shown that in continuous time ($\tau \to 0$) the third term in 5.3 equals to zero and the value of search equation simplifies to:

$$r\Omega = b - c_A(\theta_A) - c_B(\theta_B) + \lambda_A(\theta_A)\int_0^{\infty}\max(W(w)-\Omega,0)dF_A(w)+ \\ \lambda_B(\theta_B)\int_0^{\infty}\max(W(w)-\Omega-\delta/r,0)dF_B(w). \tag{5.4}$$

It is straightforward to show that $r\Omega = w_A^R = w_B^R - \delta$, where w_A^R and w_B^R are reservation wages in region A and B respectively.

Proposition 1 *Reservation wages in A (B) are decreasing (increasing) in the travel cost between A and B. The corresponding elasticities are less than unity in absolute value.*

Proof See Appendix

The interpretation of Proposition1 is straightforward: a higher travel cost reduces the value of search so the reservation wage at location A goes down, i.e. agents become less picky and

are ready to accept jobs they would not have taken before. Reservation wage in B cannot go down with commuting cost, as part of the wage would have to be sacrificed to cover the travel expenses. It is shown, however, that $0 \leqslant \dfrac{\partial w_B^R}{\partial \delta} \leqslant 1$, so the elasticity is less than unity in absolute value, which means that although the wage aspirations become higher at location B, the net demanded wage is lower than before, so, indeed, agents are less picky.

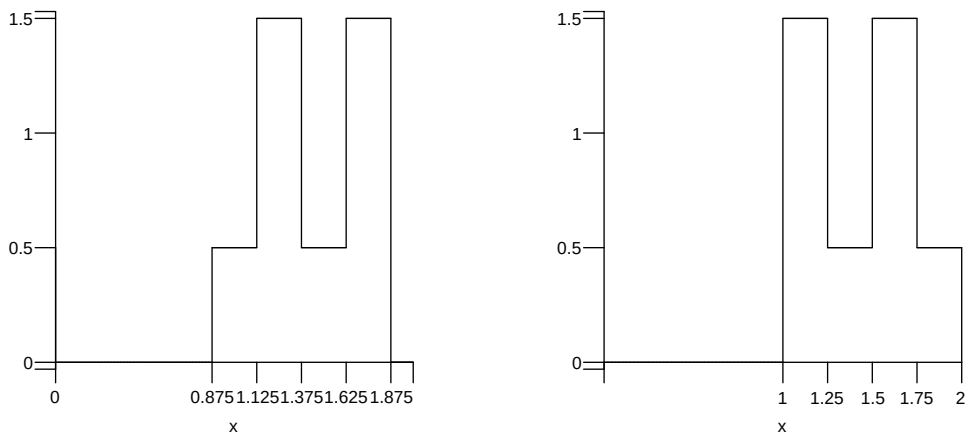
It is interesting to see the effects of changes in the moments of the wage offer distribution on reservation wages. It is common in the search literature to use the mean and the mean-preserving spread to characterize the wage distribution due Rotschild and Stiglitz (1970). However, the classical Rotschild-Stiglitz definition of the mean-preserving spread says that a distribution F_2 with a higher variance is a mean-preserving spread of the distribution F_1 if they have the same mean and

$$\int_0^w F_2(x)dx \geqslant \int_0^w F_1(x)dx.^{13} \tag{5.5}$$

The classical definition gives unambiguous results that a higher mean-preserving spread leads to higher reservation wages. However, the Rotschild-Stiglitz definition is impaired with much inflexibility as it implies a symmetric stretching or compression of the distribution, which is very unlikely to be observed empirically. If one tries to change the spread asymmetrically whilst preserving the mean then the results are not in accord with the theory.

Consider an example below.

Figure 5.1: Mean-preserving spread. An example



both distributions have the mean 1.4375 and the variance 0.079427

Call the distribution to the left on the figure 5.1 F_1 and the distribution to the right F_2. Both distribution have the same mean and the variance: $E(x) = 1.4375$ and var$(x) = 0.079427$. The reservation wage function in a simple unilocational context is given as: $w^R = b - c +$

$\frac{\lambda}{r}\int_{w^R}^{\infty}(w-w^R)dF(w)$. Let $b-c=0$ and $\lambda/r=6$ without the loss of generality. Then the reservation wage of the individual facing the distribution F_1, $w_1^R=1.28324$ and the reservation wage of the person facing F_2, $w_2^R=1.27653$. So despite having the same mean and variance, these two distributions cause different reservation wages. Ambiguity comes from the fact that the condition 5.5 does not hold.

As an alternative to Rotschild-Stiglitz mean-preserving spread I propose using the median-preserving spread. To be more precise, I suggest using the median as a location parameter of the distribution and the median-preserving spread in the left tail as a scale parameter of the left tail and the median-preserving spread in the right tail as a scale parameter of the right tail of the wage distribution (see also Möller and Aldashev (2006b) who address this issue).

Define the median of the wage offer distribution as $\bar{w}$, the median-preserving spread in the right tail as σ_R, and the median-preserving spread in the left tail as σ_L. The spread is median-preserving if for any arbitrary $\sigma_{R1}<\sigma_{R2}$ and $\sigma_{L1}<\sigma_{L2}$:

$$F(\bar{w};\sigma_{R1})=F(\bar{w};\sigma_{R2})=F(\bar{w};\sigma_{L1})=F(\bar{w};\sigma_{L2})=1/2. \tag{5.6}$$

Moreover,

$$\int_0^{\bar{w}}F(w;\sigma_{L1})\,dw<\int_0^{\bar{w}}F(w;\sigma_{L2})\,dw;\int_{\bar{w}}^{\infty}F(w;\sigma_{R1})\,dw>\int_{\bar{w}}^{\infty}F(w;\sigma_{R2})\,dw. \tag{5.7}$$

Proposition 2 *The reservation wage for region A increases with the median wage of region A and region B (but the elasticity is less than unity) and the median-preserving spread in the right tail of the wage distribution of region A and region B. It decreases with the median-preserving spread in the left tail of the wage distribution of region A and region B. The same applies for the reservation wage for region B.*

Proof See Appendix.

Corollary The reservation wage set by a searcher for any location does not only depend on wages in this location, but also on wages in all other locations.

Some ideas in Proposition 2 are not new. For example in classical search models the elasticity of the mean wage is also positive and less than unity. However, in standard search models reservation wage increases with the variance. This is because "...the worker has the option of waiting for an offer in the upper tail of the wage distribution" (Mortensen (1986, p. 865)). The effect of the median-preserving spread in the right tail has the same interpretation. However, increasing the spread in the lower tail of the wage distribution allocates more probability mass

to the jobs with lower wages and less probability mass to the jobs with higher wages. Moreover, some of the probability mass has gone to jobs which pay wages below the reservation wage. To compensate for this loss of probability mass, reservation wage declines with the median-preserving spread in the left tail.

It is assumed that agents optimize their search effort to maximize the returns to search. This implies that search intensity in the region A solves:

$$c'_A(\theta_A) = \frac{\lambda'_A(\theta_A)}{r} \int_{w_A^R}^{\infty} (w - w_A^R) dF_A(w), \tag{5.8}$$

and likewise:

$$c'_B(\theta_B) = \frac{\lambda'_B(\theta_B)}{r} \int_{w_B^R}^{\infty} (w - w_B^R) dF_B(w). \tag{5.9}$$

Proposition 3 *Higher commuting costs make agents search harder in the local labor market and less intensively in a distant labor market.*

Proof See Appendix.

The result of Proposition 3 is important that it establishes interdependency of search intensities. When commuting costs go up, the net expected wage in a distant labor market decreases and wealth-maximizing agents not just reduce their search intensity in the distant region, they *reallocate* this search intensity to search activity in the local labor market.

Proposition 4 *Agents search harder in the local labor market and less intensively in the distant labor market if median wage or the median-preserving spread in the right tail of the wage distribution increase in the home region or when the median-preserving spread in the left tail of the wage distribution decrease in the home region. If median wage or the median-preserving spread in the right tail of the wage distribution increase both in the distant and local labor markets by the same amount, agents search harder in both regions; and if the median-preserving spread in the left tail of the wage distribution increase both in the distant and local labor markets by the same amount, agents search less intensively in both regions.*

Proof See Appendix.

As in Proposition 3 the result establishes reallocation of search intensity to regions where expected wage increases. An important result of Proposition 4 is also that if the median wage in both regions increases by the same amount, search intensities increase in both regions.

5.3 Maximal Acceptable Travel Cost

The next important issue which I would like to address in this chapter is the maximal acceptable commuting cost. The necessary condition that a resident of region A searches in region B is that the returns to search in a distant labor market cover the search costs. It is then possible that after some critical level of commuting cost the returns to search do not cover the search costs anymore. The condition that the returns to search are fully offset by the search costs is called here a *zero search condition* meaning that at this point, a resident of A is indifferent between searching in both regions or in the local labor market only. It is possible to determine the value of the travel cost making the searcher indifferent between investing in search in region B, or search in A only. So far we have a system of four equations with four unknowns:

$$\begin{cases} w_A^R = b - c_A(\theta_A) - c_B(\theta_B) + \\ \frac{\lambda_A(\theta_A)}{r}\int_{w_A^R}^{\infty}(w - w_A^R)dF_A(w) + \frac{\lambda_B(\theta_B)}{r}\int_{w_B^R}^{\infty}(w - w_B^R)dF_B(w) \\ w_A^R + \delta = w_B^R \\ c_A'(\theta_A) = \frac{\lambda_A'(\theta_A)}{r}\int_{w_A^R}^{\infty}(w - w_A^R)dF_A(w) \\ c_B'(\theta_B) = \frac{\lambda_B'(\theta_B)}{r}\int_{w_B^R}^{\infty}(w - w_B^R)dF_B(w). \end{cases} \tag{5.10}$$

Solving this system we can find reservation wages in A and B and search intensities in A and B. In order to find the travel cost which makes an agent indifferent between searching in A and B and searching in A only, the following condition has to be imposed:

$$c_B(\theta_B) \leqslant \frac{\lambda_B(\theta_B)}{r}\int_{w_B^R}^{\infty}(w - w_B^R)dF_B(w). \tag{5.11}$$

Restriction in 5.11 sets the condition that the returns to search in region B should not be less than the costs of the search in B (the Kuhn-Tucker condition). Solving Equation 5.10 and 5.11 simultaneously in five endogenous variables: $w_A^R, w_B^R, \theta_A, \theta_B, \delta$ yields the value of the maximal acceptable commuting cost. If the travel costs exceeds this critical level, then the returns to search in a distant region do not cover search costs. Hence, a resident of A will invest in search only in those regions which lie within the acceptable travel cost. If for the ease of exposition we assume that the commuting cost depends only on the physical distance between regions, then Equation 5.10 and 5.11 enable us to find the maximal radius of possible commuting destinations. This is schematically shown in Figure 5.2.

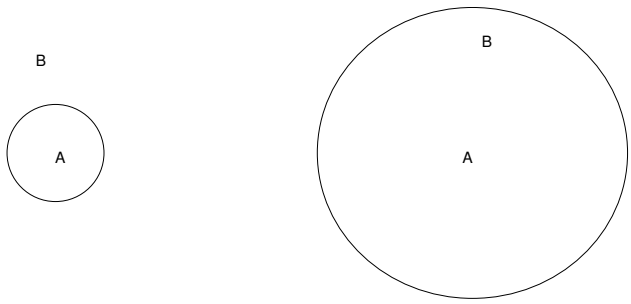


Figure 5.2: Circle of possible commuting destinations

In Figure 5.2 to the left we see that the circle of acceptable commuting distance around region A (small circle) does not include region B, hence, a searcher in region A does not commute to B. But the situation changes, if, for example, wages in both regions increase by the same amount as in Figure A.3, the circle inflates (as on the graph to the right) and a searcher in region A may commute to B. The reverse is also true, i.e. region B is an acceptable commuting destination for a resident of A (situation on the right graph). But if the wages in both regions decrease the circle shrinks (graph to the left) and B is no more a possible commuting destination.

5.4 Participation

The probability that a resident of region A commutes to B can be given by the hazard rate $\lambda(\theta_B)(1 - F(w_B^R))$ (probability that he is offered a job in B and the offered wage exceeds his reservation wage). The commuter flow from A to B is $S_A \cdot \lambda(\theta_B)(1 - F(w_B^R)$, where S_A is the number of active searchers who live in region A. If workers are homogenous the number of active searchers is constant (either all or none participate in the labor market). If wages in both regions go up, the agents would search harder in region B and, hence, the commuter flow would grow. At the same time the probability of a match in the region of residence also increases. This is in contrast to the approach used in gravity models where the *ratio* of average wages in A and B is used. This approach would imply that only changes in the *ratio* would affect the commuter flow and if this ratio does not change the commuter flow also remains unchanged. Proposition 4 shows that equal growth of wages in regions A and B increases the commuter stream between A and B although the ratio of median wages may not necessarily change.

Now suppose that workers differ in the value they attach to leisure (see also Albrecht and Axell (1984), Möller and Aldashev (2006b)). As in Section 4.1.3, I allow for three states: employment, unemployment, and nonparticipation. Individuals do not participate in the labor market if their returns to search are less than the value of not participating in the labor market. Sup-

pose that individuals can stay inactive thereby earning pure leisure which is worth b. If they participate, their reservation wage as previously defined is:

$$w_A^R = b - c_A(\theta_A) - c_B(\theta_B) + \lambda_A(\theta_A)\int_{w_A^R}^{\infty}(w - w_A^R)dF_A(w) + \lambda_B(\theta_B)\int_{w_B^R}^{\infty}(w - w_B^R)dF_B(w). \tag{5.12}$$

The participation condition is thus: $w_A^R \geqslant b$. Ruling out corner solutions, there exists a "marginal individual" for whom $w_A^R = b$. Suppose the values of leisure are distributed across individuals with the distribution function $G(b)$. The participation rate is then $G(\tilde{b})$, where $\tilde{b}$ is defined as $\tilde{b} \equiv w_A^R(\tilde{b})$. It can be shown that the participation rate increases with the median wage (both in origin and destination), median-preserving spread in the right tail (both in origin and destination), but decreases with the median-preserving spread in the left tail (both in origin and destination).

Suppose that the worker with the lowest value of leisure has $b = 0$. [14] The number of commuters from region A to B can be given as:

$$\int_0^{\tilde{b}} \lambda_B(\theta_B)\Big(1 - F_B(w_B^R)\Big)dG(b)POP_A, \tag{5.13}$$

where POP_A stands for the population size of region A. Increase in the median wage in the destination would increase $\tilde{b}$. Moreover, increase in the median in region B would make agents reallocate their intensity to region B (see Section 5.2) and therefore $\lambda_B(\theta_B)$ also increases. The reservation wage in B increases with the median wage in B but the elasticity is less than unity, hence, $\Big(1 - F_B(w_B^R)\Big)$ also increases with the median wage in B. As a consequence, the commuter flow from A to B unambigously increases with the median wage in the destination. If the median wage in the origin increases, the participation rate also goes up ($\tilde{b}$ increases). However, agents would reallocate their search intensity to region A and hence, $\lambda_B(\theta_B)$ declines. This implies that when wages increase in the origin, less commuting is possible because agents start searching harder in the origin and less harder in the destination, but on the other hand, more commuting is possible because overall number of searchers in the origin increases. Hence, the overall effect is ambiguous.

The effect of spreads cannot be analytically determined. For example, increase in the spread in the right tail of the distribution in the destination would increase participation ($\tilde{b}$ is higher) and search intensity ($\lambda_B(\theta_B)$ would increase). However, the term $1 - F_B(w_B^R)$ might decrease with the spread in the right tail (see proof in Mortensen (1986) on the ambiguity of the effect of dispersion). The effect of the increase in the spread in the right tail of the distribution in the

origin is even less clear as $\tilde{b}$ and $\lambda_B(\theta_B)$ will move in the opposite directions with the changes of the spread in the origin.

Chapter 6

Empirical Estimation of a Commuting Model

6.1 Introduction

Consider residents of region A looking for a job and firms located in B with vacancies. If these workers from A are matched with vacancies in B (given that residents of A search in B with positive search intensity) they commute. Hence, we could treat the commuter stream from A to B as successful matches of searchers from region A with vacancies in region B. The expected number of commuters from A to B, as already discussed in Chapter 5 can be obtained as $S_A \cdot \lambda_B(\theta_B)(1 - F(w_B^R))$. Search intensity and reservation wage of a resident of region A searching in region B depend on characteristics of both origin and destination regions: the arrival rate, search cost, median wage, median-preserving spread in the left and right tail, value of leisure (and if we extend it to a multi-region model then also on characteristics of all other regions).

For empirical convenience the matching function can be taken as Cobb-Douglas which in principle can be estimated by OLS after taking logs. This approach is followed by Gorter and van Ours (1994) and Burda and Profit (1996) who analyze matching of unemployed workers with open vacancies. However, in case of commuting, "...the predominance of zeros and the small values and clearly discrete nature of the dependent variable suggest that we can improve on least squares and the linear model with a specification that accounts for these characteristics" (Green (2003, p. 740)). Many works revert to discrete probability process when analyzing commuting streams (see among others Raphael (1998), Flowerdew and Aitkin (1982), Guy (1987), Yun and Sen (1994)). Although it has to be noted that classical gravity models are still to be found

in some works (e.g. Foot and Milne (1984), Hatton and Williamson (2002), Mitchell and Pain (2003)).

6.2 Data and Descriptive Evidence

The commuter stream data used in this paper are produced by the Institute for Labor Research (IAB) from the employment register of the Federal Labor Office with regional information. The data contain the flows of commuters in 1997 between 440 NUTS-3 regions, which makes 193 160 observations. Unfortunately, the data do not differentiate commuters with respect to gender. The dependent variable is then the commuter flow from region i to j. Wage quantiles of the wage distribution were calculated using the IABS-REG microdataset for 1997 (see description of the data in Appendix A.12 and Bender, Haas, and Klose (2000)). Data on population of NUTS-3 regions were taken from the INKAR database of the Federal Office for Building and Regional Planning. The information on travel time was taken from the data of the Institute for Regional Planning of the University of Dortmund (see description of the data in Appendix A.12 and Spiekermann, Lemke, and Schürmann (2000)).

The exogenous parameters used for estimation are:[15]

- $\bar{w}_i$ - the median log wage in region i,
- $\bar{w}_j$ - the median log wage in region j,
- POP_i - log population of region i (as an indicator of the size of the labor market in i),
- POP_j - log population of region j,
- $D8/D5_i$ - the log difference of 8th to 5th decile of the wage distribution in region i (as an indicator of the median-preserving spread in the right tail),
- $D8/D5_j$ - the log difference of 8th to 5th decile of the wage distribution in region j,
- $D5/D2_i$ - the log difference of 5th to 2nd decile of the wage distribution in region i (as an indicator of the median-preserving spread in the left tail),
- $D5/D2_j$ - the log difference of 5th to 2nd decile of the wage distribution in region j,
- t_{ij} - log travel time between regions i and j.

Since in a multi-region model commuting streams depend on exogenous parameters of all other regions, I also include average wage in other regions (except i and j) with inverse travel time

as weights and average population in other regions (except i and j) with inverse travel time as weights. These spatially weighted variables are calculated as:

$$\tilde{W}_{ij} = \sum_{s \neq i, s \neq j} \frac{\bar{w}_s}{t_{is}} \qquad \text{and} \qquad \tilde{P}_{ij} = \sum_{s \neq i, s \neq j} \frac{POP_s}{t_{is}}. \tag{6.1}$$

Table A.2 presents the descriptive statistics of the data used. The average travel time between pairs of regions is about 4.5 hours. However, the distribution of travel times is rather dispersed. The same can be said about the distribution of population.

Table A.3 presents the descriptive statistics separately for regions of West Germany and East Germany. One could see that average daily earnings are substantially higher in the West, but on the other hand, more dispersed. Western regions are also on average larger in population size, however, with very high variation of population size. Eastern regions experience higher unemployment rates. Moreover, Eastern regions are on average farther from each other in terms of travel time than the Western regions. However, the dispersion of travel times in the East is also larger.

This stresses importance of controlling for differences in search behavior in East Germany and West Germany. Figure 6.1 shows the distributions of commuting flows from West Germany (left panel) and East Germany (right panel). The figure shows that in West Germany the share of stayers is larger than in the East which is reflected by a larger probability of zero commuting. The differences with respect to destination are also important. As shown in Figure 6.2, the share of zeros for destination East Germany is larger than for the West, which tells us that job searchers are more likely not to commute to East Germany. Figure 6.3 shows that movements from the East to the West are more likely than commuter movements from West to the East.

To account for these differences, I introduced two dummy variables indicating whether the region of origin is East or West Germany and whether the destination is East or West Germany.

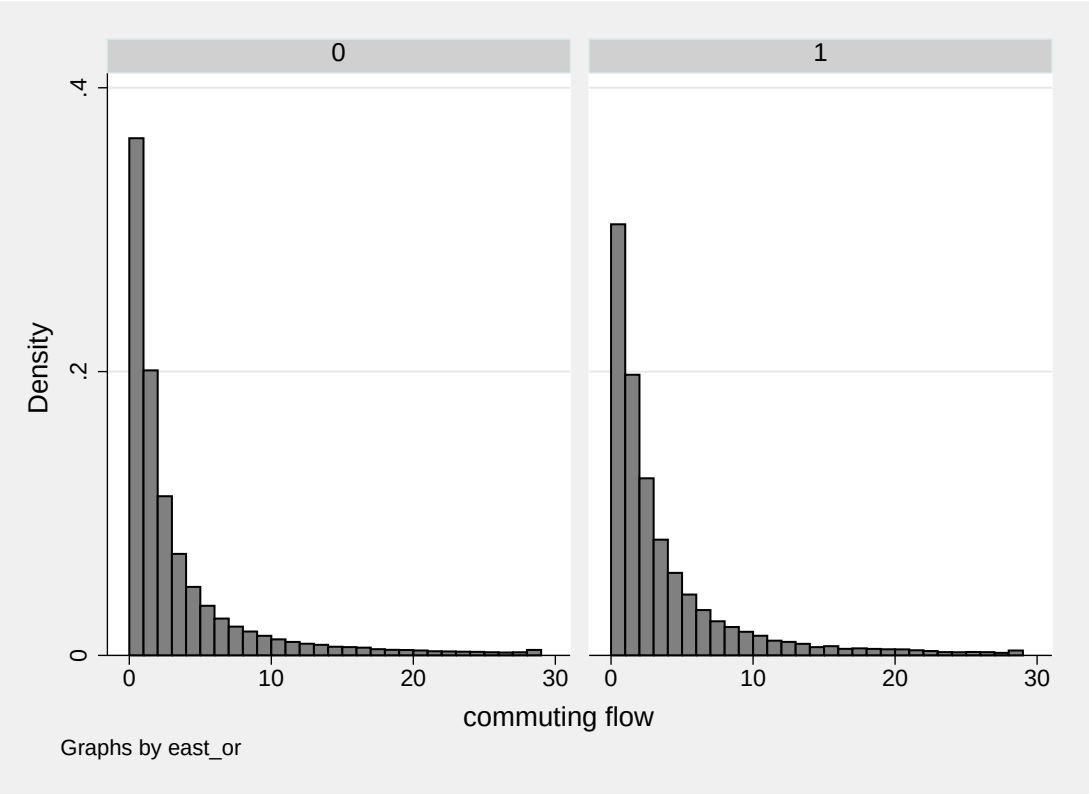


Figure 6.1: Distribution of commuting flows. Origin West Germany (left) and origin East Germany (right).

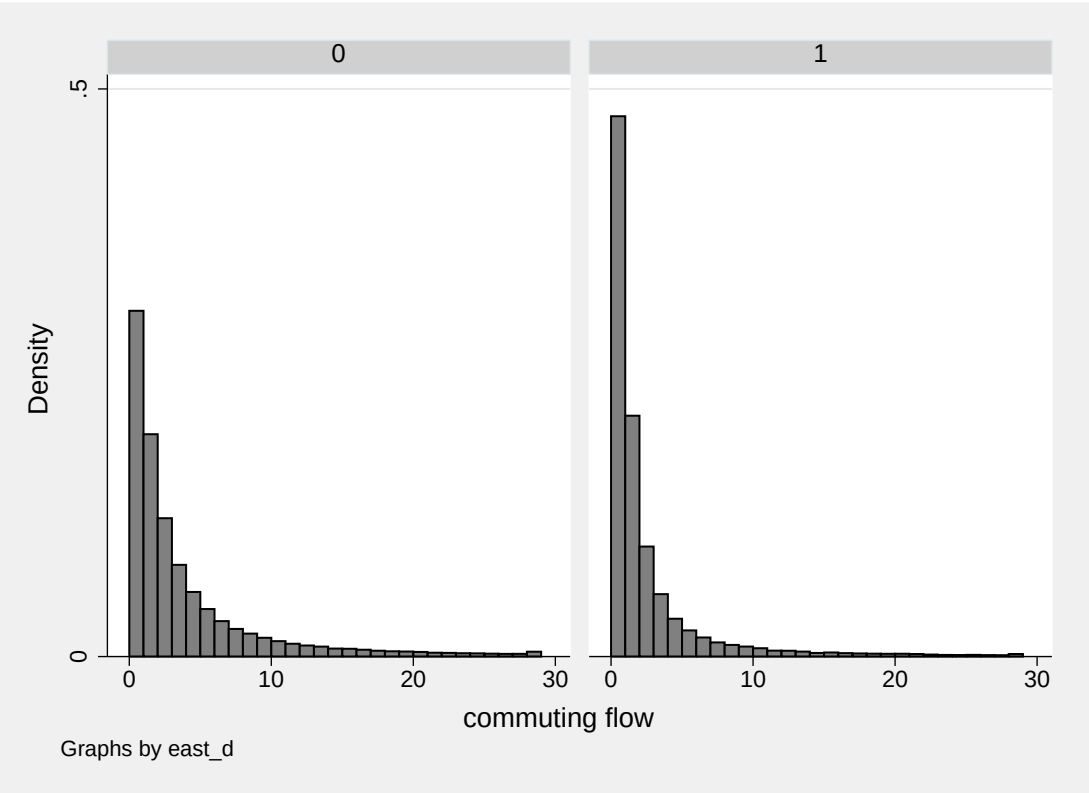


Figure 6.2: Distribution of commuting flows. Destination West Germany (left) and destination East Germany (right).

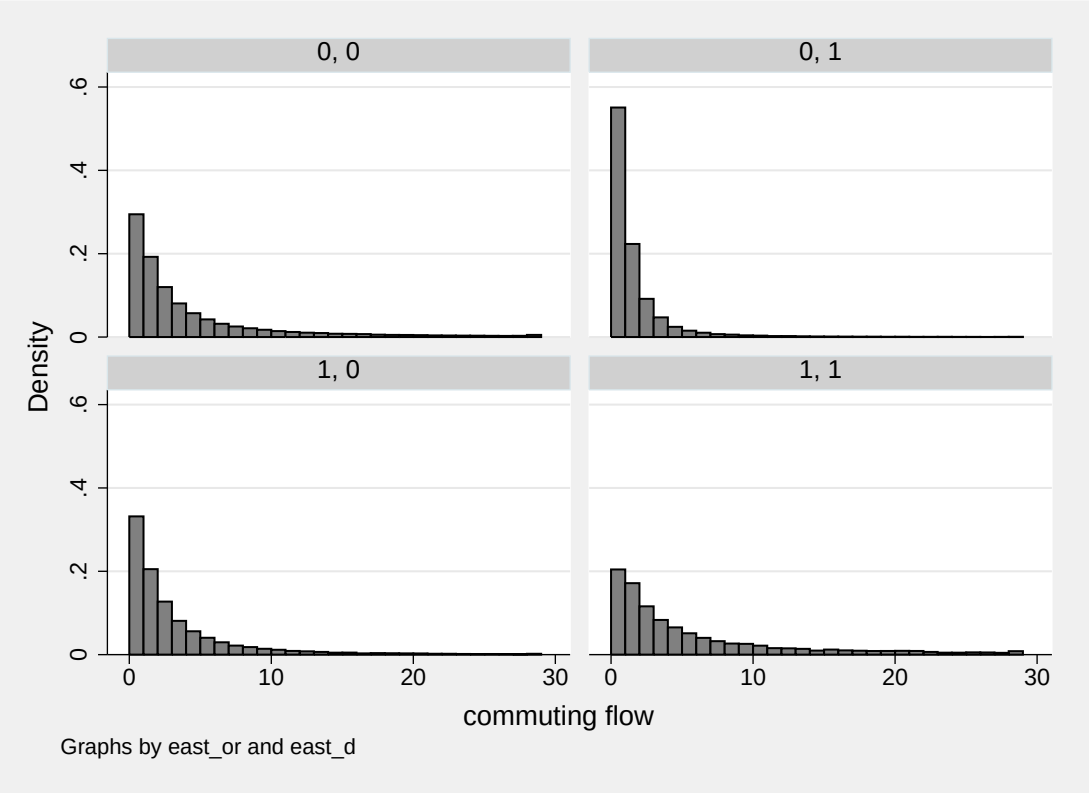


Figure 6.3: Distribution of commuting flows. Origin and destination West Germany (upper left panel), origin West, destination East Germany (upper right panel), origin East, destination West (lower left panel), origin and destination East (lower right panel).

6.3 Clustering and Robust Variance Estimation

The vector of commuter flows has the following form:

$$\begin{array}{ll} A & B \\ A & C \\ A & D \\ A & \dots \\ \vdots & \dots \\ B & A \\ B & C \\ B & D \\ B & \dots \\ \vdots & \dots \end{array} \tag{6.2}$$

The first column represents an indicator for the region of origin, the second column is an indicator for the destination. After some permutations 6.2 can be written as:

$$\begin{array}{cc} B & A \\ C & A \\ D & A \\ \dots & A \\ \vdots & \dots \\ A & B \\ C & B \\ D & B \\ \dots & B \\ \vdots & \dots \end{array} \tag{6.3}$$

The above shown representation illustrates that the data are clustered with respect to the origin and the destination, which due to omitted region-specific effects results in correlation of errors within clusters and, hence, biased standard errors. This is called Moulton's bias due to Moulton (1990). This can be easily demonstrated on the following example. Denote y_{ir} as the dependent variable; i is an index for an observation and r is an index for a region. Let the equation of interest be of the form:

$$y_{ir} = x_{ir}\beta + \upsilon_r + \epsilon_{ir}. \tag{6.4}$$

If for a region r we have several observations and the term υ_r is unobserved, then Equation 6.4 can be written as:

$$y_{ir} = x_{ir}\beta + u_{ir}, \tag{6.5}$$

with a compound error term $u_{ir} = \upsilon_r + \epsilon_{ir}$. Error terms within each region would be correlated because they would include the same term υ_r. Therefore, in all further estimations I control for error correlation within clusters and report corrected standard errors (more on clustering and error correlation see in Moulton (1990), Dickens (1990), Blanchflower and Oswald (1995)).

Estimation of the correct variance in the presence of clustering is as follows. Define the score function as: $S = \dfrac{\partial L}{\partial(\mathbf{x}\beta)}$, where L is the log-likelihood function, $\mathbf{x}$ is the matrix of independent variables, and β is the coefficient vector. The Hessian is defined as: $H = \dfrac{\partial^2 L}{\partial(\mathbf{x}\beta)^{\mathbf{2}}}$. Suppose we have R regions within which observations are clustered. The robust standard errors can be calculated as:

$$\text{var}(\beta) = D^{-1}\Big(\sum_{r=1}^{R} U_r' U_r\Big) D^{-1}, \tag{6.6}$$

where $D = \sum_{r=1}^{R} \sum_{i \in r} (H_{ir} x'_{ir} x_{ir})$ and $U_r = \sum_{i \in r} x_{ir} S_{ir}$, which is a contribution of each cluster to the score.

Consider a simple case with two clusters A and B with two observations per cluster. Let the vector of explanatory variable be of the form: $x_{1A}, x_{2A}, x_{3B}, x_{4B}$. If we ignore clustering and consider each observation as independent, the term $\sum_{r=1}^{R} U'_r U_r$ would simplify to $\sum (x_i S_i)^2 = (x_{1A} S_{1A})^2 + (x_{2A} S_{2A})^2 + (x_{3B} S_{3B})^2 + (x_{4B} S_{4B})^2$. If we take into account clustering then $\sum_{r=1}^{R} U'_r U_r = (x_{1A} S_{1A} + x_{2A} S_{2A})^2 + (x_{3B} S_{3B} + x_{4B} S_{4B})^2$, which is larger. This simple illustration shows that ignoring correlation of errors within clusters causes a downward bias in the variance of the estimated coefficients and in many cases coefficients of variables appear to be significant when in reality they are not.

6.4 Estimating the Poisson Model

In its simplest form, the discrete probability process of a variable can be modeled by the Poisson model:

$$\Pr(Y_{ij} = y_{ij}) = \frac{e^{-\lambda_{ij}} \lambda_{ij}^{y_{ij}}}{y_{ij}!}. \tag{6.7}$$

Most commonly λ_{ij} is parametrically specified as: $\lambda_{ij} = e^{x\beta}$ with $\mathbf{x}$ as a matrix of exogenous variables. The problem of the Poisson model is that in the Poisson distribution the mean equals the variance, which in case of overdispersion, i.e. if the variance is in reality greater than the mean, causes a downward bias in the standard errors. More general negative binomial model is used instead of Poisson regression by many authors (the negative binomial specification is discussed in Section 6.5).

Table 6.1: Estimation results of the Poisson model (Model 1 includes two dispersion parameters as regressors, Model 2 only one). Dependent variable - commuting flows.

	Model 1		Model 2	
variable	**coef.**	**robust st. er.**	**coef.**	**robust st. er.**
East origin	-0.12	0.15	0.01	0.16
East destination	**0.43**	0.17	**0.36**	0.17
log travel time	**−2.50**	0.03	**−2.49**	0.03
log median wage (origin)	**−3.00**	0.61	**−3.09**	0.71
log D5/D2 (origin)	**5.74**	0.91	-	-
log D8/D5 (origin)	**−6.93**	1.48	-	-
log D8/D2 (origin)	-	-	**8.63**	2.52
log median wage (destination)	**4.09**	0.54	**3.91**	0.60
log D5/D2 (destination)	**−2.42**	1.03	-	-
log D8/D5 (destination)	2.27	1.72	-	-
log D8/D2 (destination)	-	-	-4.32	3.26
log POP (origin)	**0.80**	0.08	**0.79**	0.09
log POP (destination)	**0.89**	0.06	**0.89**	0.07
spatially weighted POP ($\times 100$)	**−0.40**	0.08	**−0.37**	0.09
spatially weighted median wage	-0.04	0.03	-0.05	0.04
const	3.85	3.39	-0.02	6.42
R^2	0.82		0.83	
N		190 532		

Coefficients significant at least at 5% level are in bold.

Using the results obtained in Section 5.4 we could make certain predictions about the sign of the effects of different variables. For example, the effect of the median wage of the wage distribution in the destination should have positive effect on commuting, whereas the effect of the median wage in the origin is ambiguous. The travel time is expected to have negative influence on commuting. The surrounding regions might also attract searchers to them thereby distracting commuting from destination region. This is usually called intervening opportunities (see the classical work of Stouffer (1940)). Considering that regions with higher wages attract more, $\tilde{W}_{ij}$, ceteris paribus, should have a negative effect on commuting. The effect of dispersion parameters is ambiguous.

Table 6.1 present the results of the Poisson model. The model fits the data well which is shown by a high R^2 value. East dummy is significant and positive. This tells us that controlled for other factors commuter streams to East German regions are by 40% larger than to the West.

This might be partly explained by higher job availability in the West, so for the Western workers it is easier to find a job in the region of residence (although it is not possible to either prove or disprove that within this model). Travel time as expected is negative and highly significant. One percentage increase in the travel time reduces the commuter stream by 2.5 percent. The moments of the wage distribution in the origin are highly significant and have the expected sign. Increase in the median wage in the origin by one percent reduces the number of commuters by three percent. In the light of the theoretical model this implies that the negative effect of the reallocation of the search intensity from the destination to the origin (the agents reallocate their search intensity and search harder in the region of origin and less intensively in the destination) dominates over the positive effect of the wage increase on participation. Increase in the spread in the right tail of the distribution (proxied in the model by the log ratio of the eighth to the second decile) also reduces search in the destination and hence leads to fewer commuters. The spread in the left tail of the wage distribution in the origin (proxied in the model by the log ratio of the eighth to the second decile) increases the outflow of commuters. Increasing the median wage in the destination by one percent would result in four percentage increase of the commuter flow. It is plausible that the coefficient for the effect of the median wage in the destination is greater in magnitude than the coefficient for the origin as participation and search intensity in the destination move in the same direction with the change of the median in the destination, and participation and search intensity in the origin move in the opposite directions with the change of the wages in the origin.

The effects of wages in the origin and destination on commuting enable us to calculate the net commuting effect. If we have two regions, A and B and median wages are only explanatory variables, we could write:

$$\begin{aligned} y_{AB} &= \bar{w}_A\beta_1 + \bar{w}_B\beta_2 \\ y_{BA} &= \bar{w}_B\beta_1 + \bar{w}_A\beta_2. \end{aligned} \tag{6.8}$$

Equation 6.8 shows that if, for example, the median wage in region A increases, the commuter flow from A to B is reduced by β_1 and on the other hand, the commuter flow from B to A increases by β_2. So the effect of an increase in the median wage in A on the commuter saldo between A and B ($y_{AB} - y_{BA}$) according to the results of the Poisson model is: $0.97y_{AB} - 1.04y_{BA}$. If the original commuter saldo is zero ($y_{AB} - y_{BA} = 0$), then increase of the median wage in region A by one percent would reduce the commuter saldo of the region A by 7 percent.

The effects of the population size in the origin and destination are positive and significant (although elasticity is less than unity) which implies that larger regions interact more with each other. The effect of population size on the commuter saldo is negligible (if the original com-

muter saldo is zero). The effect of the population size in the surrounding regions other than destination reduces the commuter flow. This implies that regions that are closer and larger are more likely to be intervening opportunities for potential commuters. The effect of an average wage at locations other than the region of destination is statistically insignificant.

For comparison in Table 6.1 estimates of the Poisson regression are presented where instead of two separate parameters for spreads in the tails, only one spread parameter (the ratio of 8th to 2nd decile) is used.

6.5 Estimating the Negative Binomial Model

Negative binomial can be obtained from the Poisson model by introducing a random component into λ. To be more precise let $\lambda_{ij} = \mu_{ij}\nu_{ij}$. If we specify $\mu_{ij} = e^{x\beta}$ and ν_{ij} follow a Gamma distribution with $E(\nu_{ij}) = 1$ and var$(\nu_{ij}) = \alpha$, then the distribution of y_{ij} can be written as:

$$h(y_{ij}, \alpha, \mu_{ij}) = \frac{\Gamma(\alpha^{-1} + y_{ij})}{\Gamma(\alpha^{-1})\Gamma(1 + y_{ij})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu_{ij}} \right)^{\alpha^{-1}} \left(\frac{\mu_{ij}}{\alpha^{-1} + \mu_{ij}} \right)^{y_{ij}}. \tag{6.9}$$

The first two moments of the negative binomial distribution are: $E(y_{ij}) = \mu_{ij}$ and var$(y_{ij}) = \mu_{ij}(1 + \alpha \cdot \mu_{ij})$. If α is zero then $E(y_{ij}) =$ var(y_{ij}) and negative binomial is identical to the Poisson. Hence, testing $\alpha = 0$ after estimating the negative binomial is identical to testing the negative binomial specification vs. the Poisson.

Table 6.2: Estimation results of the negative binomial model (Model 1 includes two dispersion parameters as regressors, Model 2 only one). Dependent variable - commuting flows.

	Model 1		Model 2	
variable	**coef.**	**robust st. er.**	**coef.**	**robust st. er.**
East origin	**0.62**	0.07	**0.59**	0.07
East destination	**0.75**	0.14	**0.62**	0.14
log travel time	**−2.67**	0.02	**−2.66**	0.02
log median wage (origin)	−0.07	0.21	−0.01	0.23
log D5/D2 (origin)	**0.82**	0.39	-	-
log D8/D5 (origin)	0.44	0.60	-	-
log D8/D2 (origin)	-	-	**2.54**	1.13
log median wage (destination)	**3.99**	0.45	**4.35**	0.51
log D5/D2 (destination)	−0.90	0.83	-	-
log D8/D5 (destination)	**7.41**	1.39	-	-
log D8/D2 (destination)	-	-	**5.71**	2.31
log POP (origin)	**1.01**	0.03	**1.01**	0.03
log POP (destination)	**0.98**	0.05	**0.99**	0.06
spatially weighted POP ($\times 100$)	**−0.14**	0.03	**−0.14**	0.03
spatially weighted median wage	**−0.13**	0.01	**−0.13**	0.01
const	**−13.35**	1.19	**−22.55**	2.28
α	**1.24**	0.02	**1.26**	0.02
N	190 532			

Coefficients significant at least at 5% level are in bold.

Table 6.2 present the results of the negative binomial model. The term α is significant, which implies that the negative binomial specification performs superior to the Poisson. The East dummies for both origin and destination are significant and positive, which means that commuting is more intensive between East German regions. The median wage in the origin, the spread in the right tail of the wage distribution in the origin and the spread in the left tail in the destination appear to be insignificant in this specification. The effect of the spread in the left tail in the origin is significant, however, the magnitude is substantially reduced. The magnitude of the coefficient for the spread in the right tail in the destination on the other hand more than tripled. The effect of the median wage in the destination remains robust. Increasing the median wage in the destination by one percent would result in about four percentage increase of the commuter flow, which is not different from the results of the Poisson model.

One percentage increase in the travel time would reduce the commuter flow by 2.7 percent,

which is not much different from the Poisson model estimate (2.5 percent). The effects of the population size in the origin and destination are positive and significant with unit elasticity which implies that larger regions interact more with each other. The effect of population size on the commuter saldo is negligible (if the original commuter saldo is zero). Larger population in the surrounding regions other than destination reduces the commuter flow (although the magnitude is smaller than in the Poisson model). This implies that regions that are closer and larger are more likely to be intervening opportunities for potential commuters. The effect of an average wage at locations other than the region of destination is statistically significant unlike in the Poisson model. This implies that regions with higher wages are more likely to attract potential commuters thereby reducing the commuter flow to other destinations.

For comparison in Table 6.2 estimates of the negative binomial regression where instead of two separate parameters for spreads in the tails, only one spread parameter (the ratio of 8th to 2nd decile) is used are also presented.

6.6 Zero Inflated Models

If we look at the distribution of commuter flows we see that about one third are zeros (see for example Figures 6.1 - 6.3). This should not be surprising. In the theoretical model the zero search intensity condition implies that if a region lies beyond the circle of acceptable commuting destinations, agents do not search there. This is plausible as the travel time between some regions is about 8-10 hours, making commuting virtually impossible. We could split the decision making process of a resident of A into two stages. At stage one, he solves for the maximal acceptable travel cost (given exogenous variables of all regions) and if a region lies within the acceptable travel cost he allocates his effort into search in this region, i.e. enters stage two. At stage two, he searches for jobs at this location and, if successful, becomes consequently employed. This implies zero and nonzero commuting streams are generated by two different processes. Zeros could mean that agents do not search in these regions - decision taken by workers only, a *choice* outcome. Positive outcomes are generated through a different process - matching, a *random* outcome. To handle models like this a class of hurdle modes has been developed, for example the zero-inflated negative binomial or zero-inflated Poisson approach. In these models zeros and positive outcomes are generated by different processes. Zeros have the density $h_1(\cdot)$, so $\Pr(y_{ij} = 0) = h_1(0)$. Positive outcomes come from the truncated density $h_2(y_{ij}|y_{ij} > 0) = \dfrac{h_2(y_{ij})}{1 - h_2(0)}$. The probability that y_{ij} is drawn from this truncated density is $1 - h_1(0)$. Hence:

$$q(y_{ij}) = \begin{cases} h_1(0) & \text{if } y_{ij} = 0 \\ \dfrac{h_2(y_{ij})}{1 - h_2(0)}(1 - h_1(0)) & \text{if } y_{ij} > 0 \end{cases} \tag{6.10}$$

The likelihood function follows immediately from Equation 6.10.

Testing the zero-inflated Poisson vs. Poisson and zero-inflated negative binomial vs. negative binomial model is not trivial as the models are non-nested. [16] The Vuong test for non-nested hypotheses can be applied in this setting. Let the predicted probability that the random variable Y equals y_i in the standard negative binomial (or in the Poisson) model be $f_2(y_i|x_i)$ and in the zero-inflated negative binomial (or in the zero-inflated Poisson) – $f_1(y_i|x_i)$. We could estimate both models and obtain the log-likelihood functions, L_2 and L_1. Denote $m_i = \ln\Big(\dfrac{f_1(y_i|x_i)}{f_2(y_i|x_i)}\Big)$. The null hypothesis of the Vuong test is:

$$H_0 : E(m_i) = 0, \tag{6.11}$$

which implies that the two specifications are equally close to the true model. Vuong shows that under general conditions:

$$\frac{1}{n} LR \xrightarrow{a.s.} E(m_i), \tag{6.12}$$

where $LR = L_1 - L_2$ is the likelihood ratio test. Vuong demonstrates that:

$$\nu \equiv \frac{LR}{\sqrt{n}\hat{\omega}} \xrightarrow{D} N(0, 1), \tag{6.13}$$

where $\hat{\omega} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} m_i^2 - \Big(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} m_i\Big)^2$.

The result implies that ν has a limiting standard normal distribution. Hence, if $\nu > 1.96$, the test favors model 1, i.e. zero inflated model. If $\nu < -1.96$, the test favors the non-inflated model at 5% significance level. If $-1.96 < \nu < 1.96$, the test favors neither model. The Vuong test is intuitively appealing. If the null hypothesis is true, the log-likelihood ratio should be on average zero. If f_1 is the correct specification then on average the log-likelihood ratio should be significantly greater than zero. If f_2 is the correct model, the log-likelihood ratio should be significantly less than zero. Hence, in principle, the Vuong test is simply the *average* log-likelihood ratio suitably normalized (see also Clark and Signorino (2003)).

Table 6.3: Estimation results of the zero-inflated Poisson model (Model 1 includes two dispersion parameters as regressors, Model 2 only one). Dependent variable - commuting flows.

	Model 1		Model 2	
variable	**coef.**	**robust st. er.**	**coef.**	**robust st. er.**
East origin	−0.15	0.15	−0.02	0.16
East destination	**0.44**	0.17	**0.36**	0.17
log travel time	**−2.46**	0.03	**−2.45**	0.03
log median wage (origin)	**−2.89**	0.59	**−2.97**	0.68
log D5/D2 (origin)	**5.65**	0.89	-	-
log D8/D5 (origin)	**−6.69**	1.46	-	-
log D8/D2 (origin)	-	-	**8.60**	2.47
log median wage (destination)	**4.01**	0.54	**3.82**	0.59
log D5/D2 (destination)	**−2.58**	1.03	-	-
log D8/D5 (destination)	2.19	1.71	-	-
log D8/D2 (destination)	-	-	−4.81	3.24
log POP (origin)	**0.79**	0.08	**0.77**	0.08
log POP (destination)	**0.87**	0.06	**0.87**	0.06
spatially weighted POP (×100)	**−0.41**	0.08	**−0.38**	0.09
spatially weighted median wage	−0.03	0.03	−0.05	0.04
const	3.84	3.31	0.49	6.30
inflate				
East origin	**−1.08**	0.12	**−0.98**	0.12
East destination	**−0.74**	0.20	**−0.57**	0.19
log travel time	**1.69**	0.06	**1.71**	0.06
log median wage (origin)	**−1.66**	0.51	**−1.88**	0.54
log D5/D2 (origin)	0.90	0.68	-	-
log D8/D5 (origin)	**−5.62**	1.19	-	-

continued...

Table 6.3: Estimation results of the zero-inflated Poisson model (Model 1 includes two dispersion parameters as regressors, Model 2 only one). Dependent variable - commuting flows.

	Model 1		Model 2	
variable	**coef.**	**robust st. er.**	**coef.**	**robust st. er.**
log D8/D2 (origin)	-	-	-3.16^*	1.82
log median wage (destination)	$\mathbf{-3.40}$	0.68	$\mathbf{-3.80}$	0.29
log D5/D2 (destination)	-1.61	1.22	-	-
log D8/D5 (destination)	$\mathbf{-9.62}$	1.89	-	-
log D8/D2 (destination)	-	-	$\mathbf{-14.48}$	0.72
log POP (origin)	$\mathbf{-1.13}$	0.06	$\mathbf{-1.14}$	3.44
log POP (destination)	$\mathbf{-1.17}$	0.07	$\mathbf{-1.18}$	0.07
spatially weighted POP ($\times 100$)	$\mathbf{-0.30}$	0.10	$\mathbf{-0.27}$	0.10
spatially weighted median wage	$\mathbf{0.15}$	0.03	$\mathbf{0.14}$	0.04
const	$\mathbf{29.49}$	2.40	$\mathbf{47.72}$	4.19
Vuong test		**7.57**		**7.46**

Coefficients significant at least at 5% level are in bold.

* Coefficients significant at 10% level.

The results of the zero-inflated Poisson model are presented in Table 6.3. The Vuong test tells that the zero-inflated Poisson model is clearly preferred over the Poisson model. The table presenting the estimates of the zero-inflated models consist of two parts. First part presents the main equation, i.e. the effects of exogenous variables on the size of the commuter stream, given that commuting is nonnegative (nonzero search intensity condition, see Section 5.3). The part of the table with the heading ”inflate” presents the estimates of the inflation or selection equation. The inflation equation estimates the probability that the outcome of the dependent variable is zero. This is a discrete choice equation of zero vs. nonzero outcome, which is estimated by logit method and provides the estimates for $h_1(0)$ in Equation 6.10. The coefficients for East dummy for origin and destination are highly significant and negative, which supports the raw density estimates in Figures 6.1 and 6.2 reporting lower probabilities of zero commuting for East German regions. The effect of travel time is highly significant and positive which implies that it is more likely that no commuting takes place between more distant regions. The estimate

for the median wage in the destination is highly significant and have the expected sign, meaning that it is less likely to observe zero commuting to regions with higher wages. The estimates for the median wage in the origin and the spread in the right tail in the origin are significant but negative. The results imply that higher wages in the origin reduce the probability of zero commuting. However, it does not necessarily mean that people are more likely to leave regions with high wages. It is more likely that the negative sign is driven by the effect of wages in the origin on participation. When wages in the origin increase, more individuals participate in the labor market. This results in more commuting (because more people are active), however, this should also imply that the number of stayers increases. Indeed, the results obtained by Möller and Aldashev (2006b) show that participation rates increase with wages.

Remarkably the results of the main equation in the zero-inflated Poisson model are not much different in magnitude from the estimates of the Poisson model. For comparison the results of the zero-inflated Poisson model where instead of two separate parameters for spreads in the tails, only one spread parameter (the ratio of 8th to 2nd decile) are shown. The spread parameter for the region of origin is highly significant and positive. This implies that more people leave the region when dispersion increases. The spread parameter for the destination is statistically insignificant.

Table 6.4: Estimation results of the zero-inflated negative binomial model (Model 1 includes two dispersion parameters as regressors, Model 2 only one). Dependent variable - commuting flows.

	Model 1		Model 2	
variable	**coef.**	**robust st. er.**	**coef.**	**robust st. er.**
East origin	**0.54**	0.07	**0.51**	0.07
East destination	**0.83**	0.14	**0.70**	0.14
log travel time	**−2.65**	0.02	**−2.65**	0.02
log median wage (origin)	−0.09	0.21	−0.08	0.23
log D5/D2 (origin)	0.68*	0.39	-	-
log D8/D5 (origin)	0.25	0.61	-	-
log D8/D2 (origin)	-	-	1.94	1.15
log median wage (destination)	**3.95**	0.45	**4.37**	0.52
log D5/D2 (destination)	−1.04	0.83	-	-

continued...

Table 6.4: Estimation results of the zero-inflated negative binomial model (Model 1 includes two dispersion parameters as regressors, Model 2 only one). Dependent variable - commuting flows.

	Model 1		Model 2	
variable	**coef.**	**robust st. er.**	**coef.**	**robust st. er.**
log D8/D5 (destination)	**7.86**	1.39	-	-
log D8/D2 (destination)	-	-	**5.81**	2.33
log POP (origin)	**0.99**	0.03	**1.00**	0.03
log POP (destination)	**0.97**	0.05	**0.98**	0.06
spatially weighted POP ($\times 100$)	**−0.17**	0.03	**−0.16**	0.03
spatially weighted median wage	**−0.12**	0.01	**−0.12**	0.01
const	**−12.98**	1.21	**−21.72**	2.36
inflate				
East origin	**−20.89**	0.65	**−21.57**	0.56
East destination	**21.63**	0.85	**19.51**	0.81
log travel time	**0.78**	0.18	**1.01**	0.25
log median wage (origin)	0.11	3.43	−4.00	3.88
log D5/D2 (origin)	−3.07	3.70	-	-
log D8/D5 (origin)	**−15.96**	5.71	-	-
log D8/D2 (origin)	-	-	**−31.48**	8.63
log median wage (destination)	0.75	3.05	−0.53	2.15
log D5/D2 (destination)	−7.22	5.76	-	-
log D8/D5 (destination)	**18.38**	8.20	-	-
log D8/D2 (destination)	-	-	5.92	13.66
log POP (origin)	**−1.11**	0.11	**−1.17**	0.12
log POP (destination)	**−1.74**	0.29	**−1.70**	0.27
spatially weighted POP ($\times 100$)	**−2.76**	0.45	**−2.62**	0.46

continued...

Table 6.4: Estimation results of the zero-inflated negative binomial model (Model 1 includes two dispersion parameters as regressors, Model 2 only one). Dependent variable - commuting flows.

	Model 1		Model 2	
variable	**coef.**	**robust st. er.**	**coef.**	**robust st. er.**
spatially weighted median wage	**0.90**	0.16	**0.86**	0.17
const	−12.94	15.91	38.99	25.67
α	**1.22**	0.02	**1.24**	0.02
Vuong test		**13.74**		**12.06**

Coefficients significant at least at 5% level are in bold.

* Coefficients significant at least at 10% level.

The results of the zero-inflated negative binomial model are presented in Table 6.4. The Vuong test tells that the zero-inflated negative binomial model is clearly preferred over the negative binomial model. Moreover, the significance of α implies that the zero-inflated negative binomial model is preferred over the zero-inflated Poisson.

Unlike in the zero-inflated Poisson model, the coefficient for East dummy for the destination is positive in the inflation equation, which implies that, other things being equal, workers are more likely not to commute to eastern regions. On the other hand the coefficient for East dummy for origin is positive. This would mean that workers in the East are more likely to commute, and given the previous statement, *ceteris paribus*, they are more likely to go to western regions. The effect of travel time is highly significant and positive which implies that it is more likely that no commuting takes place between more distant regions. The same result was obtained in the zero-inflated Poisson model. However, the magnitude is almost halved (Model 1). The results of the main equation in the zero-inflated negative binomial model are not much different in magnitude from the estimates of the negative binomial model. Remarkably, the effect of the travel time, median wage in the destination and population size in both origin and destination are highly significant in all models (Poisson, negative binomial, zero-inflated Poisson, zero-inflated negative binomial). Moreover, the effect of the median wage in the destination is almost identical in all variants in magnitude: a one percentage increase in wages in the destination increases the commuter flow by 4 percent.

Finally, I allow for possible interactions between East dummy and median wage and spread

parameters. The results of the zero-inflated negative binomial model with interaction expansion are provided in Table 6.5.

Table 6.5: Estimation results of the zero-inflated negative binomial model (East dummy (origin and destination) is interacted with parameters of the wage distribution). Dependent variable - commuting flows.

variable	**coef.**	**robust st. er.**
East origin	**5.09**	2.16
East destination	−3.30	4.05
log travel time	**−2.66**	0.02
log median wage (origin)	−0.05	0.25
log D5/D2 (origin)	0.72*	0.44
log D8/D5 (origin)	**1.56**	0.70
log median wage (destination)	**3.00**	0.51
log D5/D2 (destination)	**−2.61**	0.96
log D8/D5 (destination)	**12.68**	1.63
log POP (origin)	**0.99**	0.02
log POP (destination)	**1.00**	0.05
spatially weighted POP (×100)	**−0.17**	0.03
spatially weighted median wage	**−0.12**	0.01
const	**−10.21**	1.45
inflate		
East origin	−10.39	35.80
East destination	101.05	76.20
log travel time	**0.95**	0.25
log median wage (origin)	−1.08	5.16
log D5/D2 (origin)	−4.37	4.78
log D8/D5 (origin)	−13.78*	7.42

continued...

Table 6.5: Estimation results of the zero-inflated negative binomial model (East dummy (origin and destination) is interacted with parameters of the wage distribution). Dependent variable - commuting flows.

variable	**coef.**	**robust st. er.**
log median wage (destination)	24.05	15.09
log D5/D2 (destination)	**−240.64**	113.38
log D8/D5 (destination)	185.36	121.93
log POP (origin)	**−1.16**	0.12
log POP (destination)	**−1.68**	0.27
spatially weighted POP (×100)	**−2.80**	0.56
spatially weighted median wage	**0.91**	0.20
const	−82.53	104.95
α	**1.21**	0.02
Vuong test	12.43	
N	190 532	

Coefficients significant at 5% level are in bold.

* Coefficients significant at 10% level.

Full table with all interactions is given in the Appendix.

The coefficient for the East dummy in the origin becomes ten times greater in magnitude. The coefficient for the East dummy in the destination is insignificant unlike in the models without interaction expansion. This implies that workers are more likely to commute out of East German regions. The effect of travel time remains unchanged. The same can be said about population sizes of the origin and destination. The coefficient for the median wage in the destination is reduced in magnitude. Because median wage was also interacted with East dummy, the estimates show that increase in the median wage in the West German region by one percent would result in increase in commuting to this region by 3 percent. Increase in the East German region would result in about 4.5 percentage increase in commuter flow. Unlike the results of the zero-inflated negative binomial model without interaction expansion, the estimates in TableA.1 show that increase in the spread below the median in the destination would result in a reduction of the commuter flow, which could be explained by the decrease in participation. The increase

in the spread in the right tail of the distribution in a West German region would attract more commuters. However, the increase in the spread in the right tail in an East German region seems to have no significant effect.

6.7 Conclusion

In this chapter I presented a bilocational search model where individuals have an option to commute if offered a job in a region other than their place of residence. It is shown that the minimal wage the worker is willing to accept in a ”distant” region is exactly the ”local” reservation wage plus the travel cost. Reservation wages depend on travel cost between the regions, namely, reservation wage in the origin decreases with the travel cost whereas it increases with the travel cost in the destination. The amount of intensity agents are willing to invest in searching also depends on the commuting cost. It is shown that when travel costs rise, individuals reallocate part of their search intensity from a ”distant” region to a ”local” region. This implies that the farther the region lies, the fewer workers are expected to commute.

The empirical results show that an increase in travel time to a destination region by 1% reduces the number of commuters by about 2.7% on average. Furthermore, this paper justifies the use of two dispersion parameters as regressors in modeling commuting, which has not yet been done in the literature. The empirical results claim that increase in the spread in the left tail in the destination lowers the commuter stream, but increase in the spread in the right tail for West German regions pushes it up (for East German regions no significant effect of the spread above the median is observed). Increasing the median wage in the destination by 1% raises the number of commuters by 3% if the destination is in West Germany and by about 4.5% if the destination is in the East. If median wages both in origin and destination rise by the same amount, the expected number of commuters still goes up, which is supported by numerical simulations. The results also support the role of ”intervening opportunities”. The estimates show that regions with higher wages and population are more likely to be intervening opportunities. Another important issue which is addressed in this paper is the possibility that unemployed workers do not search in destination if the region lies far enough. The theoretical model then treats the positive number of commuters as conditional on searching in the destination. This is consistent with the observed data as about one third of pairs of regions have zero commuter flows. Methods used in the literature to handle these kinds of problems are called ”hurdle” models. However, to the best of my knowledge there were no theoretical models substantiating the use of ”hurdle” models on the base of search theory. The search model presented in this paper gives theoretical justification to the use of ”zero-inflated” or ”hurdle” models for analyzing commuter flows.

Chapter 7

Summary, Potential Drawbacks and Open Questions

Despite being a young actor on the stage of economic analysis, the theory of search has proven to be very useful in bringing more insight into the problem of unemployment, labor force participation, wage inequality, mobility, and many other areas of labor market analysis. Job search theory has also found diverse empirical applications. For example, models of unemployment duration are naturally derived from the theory of job search.[17] The theory of job search is able to give alternative explanations to long-standing problems of economic theory like unemployment persistence and wage inequality across observationally identical workers (major contributions were made by Mortensen and Pissarides (1994), Burdett and Mortensen (1998)). Job search theory has also been successfully applied to the analysis of regional participation rates in Möller and Aldashev (2006a).

Some empirical phenomena, like declining reservation wages, could not have been answered within a basic search model framework. Therefore various extensions to the basic model allowing for nonstationarity of search were developed. Various authors contributed to elucidation of this phenomenon, among others Gronau (1971), Mortensen (1986), Burdett and Wishwanath (1984), van den Berg (1990). Chapter 4.1.3 of this dissertation is devoted to dynamic aspects of search; and as I was able to show, withdrawals from labor force are the logical outcome of the nonstationary job search. In the model, reservation wages decline due to nonstationarity of the arrival rate. An important aspect of the model is that wealth-maximizing agents solve not only for optimal reservation wage but also for optimal withdrawal time, called potential search time in this work, which is the maximum amount of time the unemployed workers are willing to allocate to search. If an agent is unsuccessful during the potential search time he withdraws from the labor market (he could also be considered as a ”discouraged worker”). Moreover, the

model establishes a tradeoff between the probability of finding a job and the length of potential search; hence, a change in a variable which results in reduction in the reservation wage would on the one hand increase the worker's employment chances for a given period of time, but on the other hand it results in decrease in potential search time. This could result in a decline of the overall (during the whole search span) job finding probability. This tradeoff between the probability of finding a job and the length of potential search poses certain problems when one tries to compare performances of different labor markets. Section 4.2 suggests the methods for correcting the estimated job finding rate for the withdrawal probability.

The simulation results presented in Section 3.3 show that reducing the level of unemployment insurance benefits may even worsen the employment situation due to a higher transition rate into nonparticipation. The measures aimed at increasing the arrival rate directly or indirectly seemed to be more effective. This is especially the case if the unemployment insurance benefits are generous compared to the utility of nonparticipation to encourage the unemployed to stay active in the labor market and participate in the active labor market programs if required.

The theory of job search has also been successfully applied to the analysis of labor mobility. The model of locational search is presented in Chapter 5. The model shows that the reservation wage is not unique to an individual, i.e. agents set different reservation wages for the "local" and "distant" region. Moreover, I allow for arrival rate to be the function of search intensity and therefore unemployed workers receive job offers only in those regions where their search intensity is nonzero. In this setting the maximal acceptable travel distance is determined by the condition of nonzero search intensity. It is shown that exogenous factors in "local" region affect the reservation wage and search intensity in "distant" region and vice versa. The model shows that the circle of possible commuting destinations is bounded and the maximum distance a worker is willing to commute is determined by exogenous factors in the model. The hurdle or zero-inflated models which are often applied to count data containing many zeros are a natural choice advised by the theoretical model.

It is common in the search literature to use the mean and the mean-preserving spread to characterize the wage distribution. However, I claim that the mean-preserving spread is not an appropriate measure of dispersion in case of an asymmetric wage offer distribution. In order to control for asymmetric changes in the wage dispersion one has to abandon the concept of the mean-preserving spread as one cannot change the spread in the tails of the distribution separately without affecting the mean. A good solution to this problem could be using the median as a location parameter of the distribution and the median-preserving spread in the left tail as a scale parameter of the left tail and the median-preserving spread in the right tail as a scale parameter of the right tail of the wage distribution. This approach was pioneered by Möller and Aldashev (2006b). The model predicts increase in the reservation wages in response to a

higher median-preserving spread in the right tail and decrease in response to a higher median-preserving spread in the left tail.

The empirical results claim that increase in the spread in the left tail in the destination lowers the commuter stream, but increase in the spread in the right tail for West German regions pushes it up. Increasing the median wage in the destination by 1% raises the number of commuters by 3% if the destination is in West Germany and by about 4.5% if the destination is in the East. The empirical results also support the role of ”intervening opportunities”. The estimates show that regions with higher wages and population are more likely to attract commuters.

There are still many challenges facing a researcher in the field of job search. Analysis of participation behavior and withdrawals from the labor market has received so far little attention. The empirical testing is also rather tricky so far because it is not always possible to identify a ”discouraged worker”. Many problems facing a researcher in job search are data driven: reservation wages are rarely observed, but even so, the measurement error of the self-reported reservation wages is an issue. Moreover, IABS data, for example, contain information on the unemployed receiving the unemployment benefits. Some unemployed are officially registered only to receive the unemployment compensation but do not actively search for work. On the other hand, some of the jobless workers are actively seeking employment but are not entitled to receive unemployment benefits and therefore do not appear in the administrative data. Hence, there is need for data where it is undoubtedly possible to identify the active job search state.

The chapter on commuting provides an alternative search model when the changes in the wage offer distribution are asymmetric. The model derives the sign of the effects of the changes in the spread parameters on the reservation wage unambiguously. However, the theoretical effect on commuting is ambiguous. Therefore, the estimates of the commuting model only indirectly support the theoretical results. To support the theoretical model directly, one needs to estimate the structural parameters of the model, which does not seems to be possible with the aggregate data. To test the data empirically using the micro data requires identification of the place of residence and the place of work in the data, which is not available as a scientific-use version of the IABS. On the other hand, the effects of the changes in the spread parameters of the distribution on participation are unambiguous (see Section 5.4). These results were empirically supported by Möller and Aldashev (2006b). Another potential drawback which has not yet been discussed in the locational search model presented in Section 5.2 is the competition among searchers for jobs. If the number of vacancies in the destination is limited, then the inflow of commuters would reduce the chances of getting a job, hence, reduce the arrival rate. A possible extension to the model would be to endogenize the arrival rate. In a slightly different setting this has already been done in Möller and Aldashev (2006a).

Notes

[1]In Hall (1979) a case with a linear indifference curve is also presented which results in a corner solution for an efficient separation rate.

[2]Introduction of dividends into the model is not really critical to the model and they do not enter most of the further calculations. It is assumed that unemployment compensation is financed by lump-sum taxation on dividends.

[3]If at a certain wage w' profits could be higher than at any other, all firms would set this wage and the wage distribution would collapse to a one-point distribution.

[4]In the absence of unobserved heterogeneity reservation wages do not have to be necessarily observed, however, reservation wages must be observed in order to identify the distribution of unobservables (see Frijters and van der Klaauw (2006) and Flinn and Heckman (1982))

[5]Frijters and van der Klaauw (2006) also come to this conclusion.

[6]Blanchard (1996) finds that in tight labor markets, firms will not discriminate against the long-term unemployed, but in more depressed markets they will.

[7]Lancaster (see Lancaster (1979)) calls the hazard specification a ”second best” to studying reservation wage itself.

[8]Exponential specification was called by Kiefer (1988) ”natural choice”.

[9]In case of censoring and ties the framework should be slightly modified. See Kiefer for details.

[10]I am indebted to Bernd Fitzenberger for this comment.

[11]Search theory is not the only candidate to explain interregional mobility. An interesting example of the efficiency wage theory in a locational context can be found in Zenou (2002)

[12]Commuting to region B involves a commuting or travel cost δ , so to compare wages in A and B one needs to subtract the commuting cost from the wages in B.

[13]see Mortensen (1986)

[14]The actual value is not critical; one could also assume that the lowest value of leisure is $b = b_{min}$.

[15]This is the preferred model. Alternative specifications can be found in the Appendix

[16]Testing zero-inflated Poisson vs. zero-inflated negative binomial is straightforward as the two are nested.

[17] It has to be noted, however, that the statistical theory behind duration models was developed long before durations models found applications in labor market analysis. Duration models have been successfully applied in many other sciences, for example, biology, medicine, engineering, and etc.

Bibliography

ALBRECHT, J. W., AND B. AXELL (1984): "An Equilibrium Model of Search Unemployment," *Journal of Political Economy*, 92(51), 824–840.

ARNTZ, M. (2005): "The Geographical Mobility of Unemployed Workers: Evidence from West Germany," Discussion Paper 05-34, ZEW.

BENDER, S., A. HAAS, AND K. KLOSE (2000): "IAB Employment Subsample 1975-1995: Opportunities for Analysis Provided by the Anonymised Subsample," Discussion Paper 117, IZA.

VAN DEN BERG, G. J. (1990): "Nonstationarity in Job Search Theory," *Review of Economic Studies*, 57, 255–277.

——— (1994): "The Effects of Changes of the Job Offer Arrival Rate on the Duration of Unemployment," *Journal of Labor Economics*, 12(3), 478–498.

VAN DEN BERG, G. J., A. G. C. VAN LOMWEL, AND J. C. VAN OURS (2003): "Nonparametric Estimation of a Dependent Competing Risk Model for Unemployment Durations," Discussion Paper 898, IZA.

VAN DEN BERG, G. J., AND G. RIDDER (1998): "An Empirical Equilibrium Search Model of the Labor Market," *Econometrica*, 66(5), 1183–1221.

BLANCHARD, O. (1996): "The Plight of the Long-Term Unemployed," mimeo, MIT.

BLANCHFLOWER, D. G., AND A. J. OSWALD (1995): "An Introduction to the Wage Curve," *Journal of Economic Perspectives*, 9(3), 153–167.

BONTEMPS, C., J.-M. ROBIN, AND G. VAN DEN BERG (1999): "An Empirical Equilibrium Job Search Model with Search on the Job and Heterogeneous Workers and Firms," *International Economic Review*, 40(4), 1039–1074.

——— (2000): "Equilibrium Search with Continuous Productivity Dispersion: Theory and Nonparametric Estimation," *International Economic Review*, 41(2), 305–358.

BOWLUS, A., AND L. GROGAN (2001): "Equilibrium Job Search and Gender Wage Differentials in the UK," Working Paper 2001-06, IRISS.

BOWLUS, A., N. M. KIEFER, AND G. R. NEUMANN (1995): "Estimation of Equilibrium Wage Distributions with Heterogeneity," *Journal of Applied Econometrics*, 10(Suppl.), 119–131.

BURDA, M., AND S. PROFIT (1996): "Matching Across Space: Evidence on Mobility in the Czech Republic," *Labour Economics*, 3, 255–278.

BURDETT, K. (1978): "A Theory of Employee Job Search and Quit Rates," *American Economic Review*, 68(1), 212–220.

BURDETT, K., N. KIEFER, AND G. NEUMANN (1984): "Earnings, Unemployment, and the Allocation of Time over Time," *Review of Economic Studies*, 51(4), 559–578.

BURDETT, K., AND D. T. MORTENSEN (1998): "Wage Differentials, Employer Size and Unemployment," *International Economic Review*, 39(May), 257–273.

BURDETT, K., AND T. WISHWANATH (1984): "Declining Reservation Wages and Learning," *Review of Economic Studies*, 5, 655–665.

CAHUC, P., AND A. ZYLBERBERG (2005): *Labor Economics*. MIT Press, Cambridge, Massachusetts.

CLARK, K. A., AND C. S. SIGNORINO (2003): "Discriminating Methods: Tests for Nonnested Discrete Choice Models," mimeo, Department of Political Science, University of Rochester, Harkness Hall.

CLARK, K. B., AND L. H. SUMMERS (1982): "Labour Force Participation: Timing and Persistence," *Review of Economic Studies*, 49, 825–844.

CLEVES, M., W. GOULD, AND R. GUTIERREZ (2004): *An Introduction to Survival Analysis using Stata*. Stata Press, College Station, Texas.

COX, D. R. (1972): "Regression Models and Life-Tables (with Discussion)," *Journal of Royal Statistical Society*, B34, 187–220.

——— (1975): "Partial-Likelihood," *Biometrika*, 62(2), 269–276.

DAMM, L. P., AND M. ROSHOLM (2003): "Employment Effects of Dispersal Policies on Refugee Immigrants, Part I: Theory," Discussion Paper 924, IZA.

DICKENS, W. T. (1990): "Error Components in Grouped Data: Is it Ever Worth Weighting?," *Review of Economics and Statistics*, 72(2), 328–333.

ECKSTEIN, Z., AND K. I. WOLPIN (1990): "Estimating a Market Equilibrium Search Model from Panel Data on Individuals," *Econometrica*, 58(4), 783–808.

ELIASSON, K., U. LINDGREN, AND O. WESTERLUND (2003): "Geographical Labour Mobility: Migration or Commuting?," *Regional Studies*, 37(8), 827–837.

FITZENBERGER, B., AND R. A. WILKE (2004): "Unemployment Durations in West Germany Before and After the Reform of the Unemployment Compensation System During the 1980s," Discussion Paper 04-24, IZA.

FLINN, C. J., AND J. J. HECKMAN (1982): "New Methods for Analyzing Structural Models of Labor Force Dynamics," *Journal of Econometrics*, 18(1), 115–168.

FLOWERDEW, R., AND M. AITKIN (1982): "A Method of Fitting the Gravity Model Based on the Poisson Distribution," *Journal of Regional Science*, 22, 191–202.

FOOT, D. K., AND W. J. MILNE (1984): "Net Migration Estimation in an Extended Multiregional Gravity Model," *Journal of Regional Science*, 24, 119–133.

FRANZ, W. (2006): *Arbeitsmarktökonomik*. 6. Auflage, Springer.

FRIJTERS, P., AND B. VAN DER KLAAUW (2006): "Job Search with Nonparticipation," *The Economic Journal*, 116(January), 45–83.

GALAMBOS, J. (1984): *Introductory Probability Theory*. Marcel Dekker, Inc.

GONZALO, M. T., AND J. SAARELA (2000): "Gender Differences in Exit Rates from Unemployment: Evidence from a Local Finnish Labour Market," *Finnish Economic Papers*, 13(2), 129–139.

GORDON, S. (2002): "Stochastic Dependence in Competing Risks," *American Journal of Political Science*, 46(1), 200–217.

GORTER, C., AND J. VAN OURS (1994): "Matching Unemployment and Vacancies in Regional Labor Markets: An Empirical Analysis for the Netherlands," *Papers in Regional Science*, 73, 153–167.

GREEN, W. H. (2003): *Econometric Analysis*. Pearson Education, New Jersey.

GRONAU, R. (1971): "Information and Frictional Unemployment," *American Economic Review*, 61, 290–301.

GUY, C. M. (1987): "Recent Advances in Spatial Interaction Modelling: An Application to the Forecasting of Shopping Travel," *Environment and Planning A*, 19, 173–186.

HALL, R. E. (1979): "A Theory of Natural Unemployment Rate and the Duration of Employment," *Journal of Monetary Economics*, 5, 153–169.

HATTON, T. J., AND J. G. WILLIAMSON (2002): "What Fundamentals Drive World Migration?," Working Paper 9159, NBER.

HAUSMAN, J., AND A. HAN (1990): "Flexible Parametric Estimation of Duration and Competing Risk Models," *Journal of Applied Econometrics*, 5(1), 1–28.

HOUGAARD, P. (1987): "Modeling Multivariate Survival," *Scandinavian Journal of Statistics*, 14, 291–304.

HUJER, R., K. MAURER, AND M. WELLNER (1997): "The Impact of Training on Unemployment Duration in West Germany - Combining a Discrete Hazard Rate Model with Matching Techniques," Working Paper, J. W. Goethe University, Frankfurt am Main.

JOVANOVIC, B. (1979): "Job Matching and the Theory of Turnover," *Journal of Political Economy*, 87(5), 972–990.

KALBFLEISCH, J. D., AND R. L. PRENTICE (2005): *The Statistical Analysis of Failure Time Data*. Second Edition, Wiley, New York.

KAPLAN, E. L., AND P. MEIER (1958): "Nonparametric Estimation from Incomplete Observations," *Journal of American Statistical Association*, 53, 457–481.

KASPER, H. (1967): "The Asking Price of Labour and the Duration of Unemployment," *The Review of Economics and Statistics*, 49, 165–172.

KATZ, L. (1986): "Layoffs, Recall, and the Duration of Unemployment," Working Paper 1825, NBER.

KIEFER, N. M. (1988): "Economic Duration Data and Hazard Functions," *Journal of Economic Literature*, 26, 646–679.

KIEFER, N. M., AND G. R. NEUMANN (1979): "An Empirical Job-Search Model with a Test of the Constant Reservation-Wage Hypothesis," *Journal of Political Economy*, 87(1), 89–108.

KOPKA, H. (1988): *LaTeX: Eine Einführung*. Addison-Wesley.

KREMER, N. S. (2001): *Probability Theory and Mathematical Statistics*. Unity, Moscow (in Russian).

——— (2002): *Advanced Mathematics for Economists*. Unity, Moscow (in Russian).

KUGLER, A. D., AND G. SAINT-PAUL (2000): "Hiring and Firing Costs, Adverse Selection and Long-Term Unemployment," Working Paper, Universitat Pompeu Fabra.

LANCASTER, T. (1979): "Econometric Methods for the Duration of Unemployment," *Econometrica*, 47(4), 939–956.

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

MCCALL, J. J. (1970): "Economics of Information and Job Search," *Quarterly Journal of Economics*, 84, 113–126.

MEYER, B. D. (1995): "Semiparametric Estimation of Hazard Models," Working Paper, Department of Economics and Center for Urban Affairs and Policy Research, Northwestern University.

MINCER, J. (1974): *Schooling, Experience and Earnings*. National Bureau of Economic Research, New York.

MITCHELL, J., AND N. PAIN (2003): "The Determinants of International Migration into the UK: A Panel Based Modelling," Working Paper, National Institute of Economic and Social Research.

MÖLLER, J., AND A. ALDASHEV (2006a): "Interregional Differences in Labor Market Participation," *Jahrbuch für Regionalwissenschaft (Review of Regional Research)*, 26(1), 25–50.

——— (2006b): "Wage Inequality, Reservation Wages, and Labour Market Participation," Working Paper (forthcoming in *International Regional Science Review*), University of Regensburg.

MOLHO, I. (2001): "Spatial Search, Migration and Regional Unemployment," *Economica*, 68, 269–283.

MORTENSEN, D. T. (1970): "Job Search, the Duration of Unemployment, and the Philips Curve," *American Economic Review*, 60, 847–862.

——— (1977): "Unemployment Insurance and Job Search Decisions," *Industrial and Labor Relations Review*, 30(4), 505–517.

——— (1986): "Job Search and Labour Market Analysis," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and R. Layard, vol. 2, chap. 15, pp. 849–919. Elsevier Science B.V., Amsterdam.

——— (2003): *Wage Dispersion: Why Are Similar Workers Paid Differently?* Cambridge, Massachusetts.

MORTENSEN, D. T., AND C. A. PISSARIDES (1994): "Job Creation and Job Destruction in the Theory of Unemployment," *Review of Economic Studies*, 61(3), 397–415.

MOULTON, B. R. (1990): "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units," *Review of Economics and Statistics*, 72(2), 334–338.

VAN OMMEREN, J., AND W. VAN DER STRAATEN (2005): "Identification of 'Wasteful Commuting' Using Search Theory," Discussion Paper TI 2005-088/3, Tinbergen Institute.

OSTASZEWSKI, A. (1990): *Advanced Mathematical Methods*. Cambridge University Press.

PISSARIDES, C. A. (1979): "Job Matchings with State Employment Agencies and Random Search," *Economic Journal*, 89, 818–833.

——— (1990): *Equilibrium Unemployment Theory*. Basil Blackwell, Oxford.

QIAN, L., AND J. A. CORREA (2003): "Estimation of Weibull Parameters for Grouped Data with Competing Risks," *Journal of Statistical Computation and Simulation*, 73(4), 261–275.

RAPHAEL, S. (1998): "Intervening Opportunities, Competing Searchers, and the Intrametropolitan Flow of Male Youth Labor," *Journal of Regional Science*, 38, 43–59.

RØED, K., AND T. ZHANG (2003): "Does Unemployment Compensation Affect Unemployment Duration," *The Economic Journal*, 113, 190–226.

ROGERSON, R., R. SHIMER, AND R. WRIGHT (2005): "Search-Theoretic Models of the Labor Market: A Survey," *Journal of Economic Literature*, 43(4), 959–988.

ROTSCHILD, M., AND J. STIGLITZ (1970): "Increasing Risk I: A Definition," *Journal of Economic Theory*, 2, 225–243.

SHAPIRO, C., AND J. STIGLITZ (1984): "Equilibrium Unemplyment as a Worker Discipline Device," *American Economic Review*, 74, 433–444.

SHATZKIN, A., AND E. SLUD (1989): "Competing Risks Bias Arising from an Omitted Risk Factor," *American Journal of Epidemiology*, 129(4), 850–856.

SPIEKERMANN, K., M. LEMKE, AND C. SCHÜRMANN (2000): "Trans-European Transport Networks, Digital Elevation Models and Emission Modelling," Working Paper, Institute of Spatial Planning, University of Dortmund (IRPUD).

STIGLER, G. (1961): "The Economics of Information," *Journal of Political Economy*, 69, 213–225.

——— (1962): "Information in the Labor Market," *Journal of Political Economy*, 70, 94–105.

STIGLITZ, J. (2002): "Information and the Change in the Paradigm in Economics," in *Les Prix Nobel; The Nobel Prizes 2001*, ed. by T. Frangsmyr, pp. 472–540. The Nobel Foundation.

STOUFFER, S. A. (1940): "Intervening Opportunities: A Theory Relating Mobility and Distance," *American Sociological Review*, 5, 845–867.

TAYLOR, A. E., AND W. R. MANN (1982): *Advanced Calculus*. third edition, John Wiley & Sons, Inc.

WILKE, R. A. (2004): "New Estimates of the Duration and Risk of Unemployment for West Germany," *Jorunal of Applied Social Science Studies (Schmollers Jahrbuch)*, 128(2), 207–237.

WOLFF, J. (2003): "Unemployment Benefits and the Duration of Unemployment in East Germany," Sonderforschungsbereich 386 Discussion Paper 344, LMU.

——— (2004): "The Duration of New Job Matches in East and West Germany," Discussion Paper 2004-10, LMU.

WOLFF, J., AND R. TRÜBSWETTER (2003): "The Speed of Leaving the Old Job: The Study on Job Changes and Exit into Unemployment During the East German Transition Process," Sonderforschungsbereich 386 Discussion Paper 343, LMU.

YUN, S., AND A. SEN (1994): "Computation of Maximum Likelihood Estimates of Gravity Model Parameters," *Journal of Regional Science*, 34, 199–216.

ZENOU, Y. (2002): "How Do Firms Redline Workers?," *Journal of Urban Economics*, 52, 391–408.

Appendix A

Formulae Derivation

A.1 Formulae from Section 2.1.3

Derivation of Equation 2.21.

$$\max_{m_j} \phi(m_j) = [1 - (1 - f_j)^{m_j}]\, w_j/r - m_j(c + a \cdot D_{ij}). \tag{A.1}$$

From the first order condition:

$$\frac{\partial \phi(m_j)}{\partial m_j} = -(1 - f_j)^{m_j} \ln(1 - f_j) w_j/r - (c + a \cdot D_{ij}) = 0. \tag{A.2}$$

It follows that (ignoring the subscript j to save notation):

$$m = \frac{\ln\left(-\dfrac{c + a \cdot D_{ij}}{\ln(1 - f) w/r}\right)}{\ln(1 - f)}. \tag{A.3}$$

Since $\ln(1 - f) = -f$ for small f, one obtains:

$$m = \ln\left(-\frac{c + a \cdot D_{ij}}{-f \cdot w/r}\right)/(-f) = -f^{-1} \ln\left(\frac{c + a \cdot D_{ij}}{f \cdot w/r}\right) = f^{-1} \ln\left(\frac{f \cdot w/r}{c + a \cdot D_{ij}}\right) \tag{A.4}$$

A.2 Formulae from Section 2.2.1

Conditional distribution of wage offers above the reservation wage can be written as:

$$f(w|w \geqslant w^R) = \frac{f(w)}{\Pr(w \geqslant w^R)}, \tag{A.5}$$

with $\Pr(w \geqslant w^R) = \int_{w^R} f(w)dw$.

The expected wage, given that the offers exceed the reservation wage can be given as:

$$\begin{aligned} &E(w|w \geqslant w^R) = \\ &\int_{w^R} wq(z,w)f(w|w \geqslant w^R)dw = \int_{w^R} w\frac{q(z,w)f(w)}{\int_{w^R} q(z,w)f(w)dw}dw = \\ &\frac{1}{\int_{w^R} q(z,w)f(w)dw}\int_{w^R} wq(z,w)f(w)dw = \frac{\int_{w^R} wq(z,w)f(w)dw}{\int_{w^R} q(z,w)f(w)dw}. \end{aligned} \tag{A.6}$$

Suppose that at $t = 0$ the job offer is not accepted, but accepted in next period, i.e. at $t = 1$, then the present value of the returns to search in the next period are given by:

$$b - c + \sum_{t=1}^{\infty} \frac{p(z,w^R)E(w|w \geqslant w^R)}{(1+r)^t}, \tag{A.7}$$

where b are the unemployment benefits and c is the search cost.

An agent continues the search in period 2 if unsuccessful in period 1, thus the present value of the returns to search in period 2 is conditional upon being unsuccessful in period 1:

$$\left(1 - p(z,w^R)\right)\left[\frac{b-c}{1+r} + \sum_{t=2}^{\infty} \frac{p(z,w^R)E(w|w \geqslant w^R)}{(1+r)^t}\right]. \tag{A.8}$$

For period 3 in the same fashion:

$$\left(1 - p(z,w^R)\right)^2\left[\frac{b-c}{(1+r)^2} + \sum_{t=3}^{\infty} \frac{p(z,w^R)E(w|w \geqslant w^R)}{(1+r)^t}\right]. \tag{A.9}$$

To sum up the returns to search across time periods, it would be helpful to separate the two terms in the square brackets. For the first term we get the following series:

$$b - c + \left(1 - p(z,w^R)\right)\frac{b-c}{1+r} + \left(1 - p(z,w^R)\right)^2\frac{b-c}{(1+r)^2} + ..., \tag{A.10}$$

or:

$$(b-c)\left(1 + \frac{1 - p(z,w^R)}{1+r} + \frac{\left(1 - p(z,w^R)\right)^2}{(1+r)^2} + ...\right). \tag{A.11}$$

The second in Equation A.11 is the diminishing geometric series, which sums up to:

$$1+\frac{1-p(z,w^R)}{1+r}+\frac{\Big(1-p(z,w^R)\Big)^2}{(1+r)^2}+...=\frac{1}{1-\dfrac{1-p(z,w^R)}{1+r}}=\frac{1+r}{r+p(z,w^R)}. \tag{A.12}$$

Thus, the expression in Equation A.10 equals:

$$(b-c)\frac{1+r}{r+p(z,w^R)}. \tag{A.13}$$

The sum of expected wages across periods can be given as the following series:

$$\begin{gathered}\Big[\frac{p(z,w^R)E(w|w\geqslant w^R)}{1+r}+\frac{p(z,w^R)E(w|w\geqslant w^R)}{(1+r)^2}+\cdots\Big]+\\ \Big(1-p(z,w^R)\Big)\Big[\frac{p(z,w^R)E(w|w\geqslant w^R)}{(1+r)^2}+\frac{p(z,w^R)E(w|w\geqslant w^R)}{(1+r)^3}+\cdots\Big]+\\ \Big(1-p(z,w^R)\Big)^2\Big[\frac{p(z,w^R)E(w|w\geqslant w^R)}{(1+r)^3}+\frac{p(z,w^R)E(w|w\geqslant w^R)}{(1+r)^4}+\cdots\Big]+\\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ \Big(1-p(z,w^R)\Big)^\infty[\cdots].\end{gathered} \tag{A.14}$$

This could be simplified to:

$$\frac{p(z,w^R)E(w|w\geqslant w^R)}{1+r}\left[\begin{array}{c}\Big(1+\dfrac{1}{1+r}+\dfrac{1}{(1+r)^2}+\cdots\Big)+\\ \dfrac{1-p(z,w^R)}{1+r}\Big(1+\dfrac{1}{1+r}+\dfrac{1}{(1+r)^2}+\cdots\Big)+\\ \Big(\dfrac{1-p(z,w^R)}{1+r}\Big)^2\Big(1+\dfrac{1}{1+r}+\dfrac{1}{(1+r)^2}+\cdots\Big)+\\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\end{array}\right]. \tag{A.15}$$

The term in square brackets is also a diminishing geometric progression and its sum can be

given as:

$$1 + \frac{1 - p(z, w^R)}{1 + r} + \left(\frac{1 - p(z, w^R)}{1 + r}\right)^2 + \cdots = \frac{1 + r}{r + p(z, w^R)}. \tag{A.16}$$

Thus, Equation A.15 equals:

$$\frac{p(z, w^R)E(w|w \geqslant w^R)}{1 + r}\left(\frac{1 + r}{r}\right)\left(\frac{1 + r}{r + p(z, w^R)}\right) = \frac{p(z, w^R)E(w|w \geqslant w^R)(1 + r)}{r(r + p(z, w^R)}. \tag{A.17}$$

Summing up A.17 and A.13 gives:

$$\frac{(b - c)(1 + r)}{r + p(z, w^R)} + p(z, w^R) \cdot E(w|w \geqslant w^R)\frac{1 + r}{r(r + p(z, w^R))}. \tag{A.18}$$

A.3 Formulae from Section 2.2.2

In continuous time case, $t \to 0$, one obtains:

$$\begin{gathered}
\lim_{\tau \to 0} \frac{q(1, \tau)}{\tau} = \lim_{\tau \to 0} \frac{e^{-\lambda\tau}\lambda\tau}{\tau} = \lim_{\tau \to 0} e^{-\lambda\tau}\lambda = \lambda \\
\lim_{\tau \to 0} \frac{q(m, \tau)}{\tau} = \lim_{\tau \to 0} \frac{e^{\lambda\tau}(\lambda\tau)^m}{m!\tau} = \lim_{\tau \to 0} \frac{e^{-\lambda\tau}\lambda^m\tau^{m-1}}{m!}; \;\; m > 1 \to \tau^{m-1} = 0, \lim_{\tau \to 0} \frac{q(m, \tau)}{\tau} = 0 \\
\lim_{\tau \to 0} \frac{1 - \beta(\tau)}{\tau} = \lim_{\tau \to 0} \frac{1 - e^{-r\tau}}{\tau} = \lim_{\tau \to 0}(re^{-r\tau}) = r \\
\lim_{\tau \to 0} \beta(\tau) = 1.
\end{gathered} \tag{A.19}$$

The Bellman equation in 2.28 considerably simplifies:

$$\begin{gathered}
\Omega(1 - \beta(\tau)) = (b - c)\tau + \beta(\tau)\Big[\textstyle\sum_{n=1}^{\infty} q(n, \tau) \int\limits_0^{\infty} \max\left[0, W(w) - \Omega\right] g(\tilde{w}_n)dw\Big] \to \\
\to \frac{\Omega(1 - \beta(\tau))}{\tau} = (b - c) + \beta(\tau)\Big[\sum_{n=1}^{\infty} \frac{q(n, \tau)}{\tau} \int\limits_0^{\infty} \max[0, W(w) - \Omega]g(\tilde{w}_n)dw\Big].
\end{gathered} \tag{A.20}$$

A.4 Formulae from Section 2.3.2

The change in the share of workers earning w or less can be given as:

$$dG(w) = \lambda F(w)u - \Big(\delta + \lambda(1 - F(w))\Big)(1 - u)G(w). \tag{A.21}$$

The first term denotes the share of unemployed workers (denoted by u) who find jobs (at a rate λ) offering wages w or less (given by $F(w)$). The second term denotes the employed workers ($1-u$) who become unemployed for exogenous reasons (at a rate δ) or who move to higher paid jobs (at a rate $\lambda(1-F(w))$). In the steady state $dG(w)=0$ and hence:

$$G(w;F)=\frac{\delta F(w)}{\delta+\lambda(1-F(w))}. \tag{A.22}$$

The share of firms offering the lowest wage is $F(b)=0$. A firm offering the lowest wage would then have a profit of:

$$\pi(w;F)=\frac{\delta\lambda(p-b)}{M[\delta+\lambda]^2}. \tag{A.23}$$

A firm offering a wage w would have a profit of:

$$\pi(w;F)=\frac{\delta\lambda(p-w)}{M[\delta+\lambda(1-F(w))]^2}. \tag{A.24}$$

Since in equilibrium all firms are equally profitable, equation A.23 and A.24 would yield:

$$F(w)=\frac{\delta+\lambda}{\lambda}\left[1-\sqrt{\frac{p-b}{p-w}}\right]. \tag{A.25}$$

Given the relationship between the offered and observed wages in Equation 2.53, one obtains:

$$G(w)=\frac{\delta}{\lambda}\left[\sqrt{\frac{p-b}{p-w}}-1\right]. \tag{A.26}$$

By equating $G(w)$ in A.26 to 0 and to 1, one can find that the lowest wage in the economy is b and the highest is $p-(p-b)\left(\frac{\delta}{\delta+\lambda}\right)^2$.

The expected wage in the economy is:

$$E(w)=\int_{w_L}^{w_H} w dG(w), \tag{A.27}$$

where $w_L = b$ and $w_H = p - (p-b)\Big(\frac{\delta}{\delta+\lambda}\Big)^2$.

Integrating A.27 by parts yields:

$$E(w) = wG(w)\Big\|_b^{w_H} - \int_b^{w_H} G(w)dw. \tag{A.28}$$

The first term in A.28 is equal to w_H.

$$\begin{aligned}\int_b^{w_H} G(w)dw = -2(p-w_H)^{1/2}\frac{\delta}{\lambda}\sqrt{p-b} - \frac{\delta w_H}{\lambda} + \\ 2(p-b)^{1/2}\frac{\delta}{\lambda}\sqrt{p-b} + \frac{\delta b}{\lambda}.\end{aligned} \tag{A.29}$$

Hence,

$$\begin{aligned}E(w) = w_H + 2(p-w_H)^{1/2}\frac{\delta}{\lambda}\sqrt{p-b} + \frac{\delta w_H}{\lambda} - \\ 2(p-b)^{1/2}\frac{\delta}{\lambda}\sqrt{p-b} - \frac{\delta b}{\lambda}.\end{aligned} \tag{A.30}$$

Substituting $w_H = p - (p-b)\Big(\frac{\delta}{\delta+\lambda}\Big)^2$ into A.30 yields:

$$E(w) = \frac{\delta}{\delta+\lambda}b + \frac{\lambda}{\delta+\lambda}p. \tag{A.31}$$

A.5 Formulae from Section 3.1

Collecting terms in Equation 3.2 yields:

$$\begin{aligned}\Omega(t) - \Omega(t+\tau) = (b-c)\tau + \beta(\tau)\textstyle\sum_{m=1}^{\infty} q(m,\tau)\int\limits_0^{\infty}\max[0, W(w) - \Omega(t)]g(\tilde{w}_m)dw + \\ +\beta(\tau)\textstyle\sum_{m=1}^{\infty} q(m,\tau)\Omega(t) + \beta(\tau)q(0,\tau,t)\Omega(t+\tau) - \Omega(t+\tau).\end{aligned} \tag{A.32}$$

Note that $\beta(\tau) = e^{-r\tau}$ and $q(0,\tau,\lambda) = e^{-\lambda(\tau)}$, hence one can simplify Equation A.32:

$$\Omega(t) - \Omega(t+\tau) = (b-c)\tau + \beta(\tau)\sum_{m=1}^{\infty} q(m,\tau)\int_0^{\infty} \max[0, W(w) - \Omega(t)]g(\tilde{w}_m)dw$$
$$+\beta(\tau)\sum_{m=1}^{\infty} q(m,\tau)\Omega(t) - \Omega(t+\tau)\left(1 - e^{-r\tau}e^{-\lambda(t)\tau}\right). \tag{A.33}$$

Moreover, in continuous time:

$$\lim_{\tau\to 0}\left[\frac{\Omega(t+\tau)-\Omega(t)}{\tau}\right] = \frac{d\Omega(t)}{dt};$$
$$\lim_{\tau\to 0}\frac{q(1,\tau,\lambda)}{\tau} = \lambda(t); \qquad \lim_{\tau\to 0}\frac{q(m,\tau,\lambda)}{\tau} = 0, \text{ for } m > 1 \tag{A.34}$$
$$\lim_{\tau\to 0}\frac{\left(1 - e^{-r\tau}e^{-\lambda(t)\tau}\right)}{\tau} = r + \lambda(t).$$

Hence, dividing A.33 by τ and collecting terms yields:

$$-\frac{d\Omega(t)}{dt} = b - c + \lambda(t)\int_0^{\infty} \max\{0, W(w) - \Omega(t)\}\, dF(w) +$$
$$+\lambda(t)\Omega(t) - \Omega(t)(\lambda(t) + r) = \tag{A.35}$$
$$= b - c + \lambda(t)\int_0^{\infty} \max\{0, W(w) - \Omega(t)\}\, dF(w) - \Omega(t)r.$$

Remembering that $\Omega(t)r = w^R(t)$ we can rewrite Equation A.35:

$$w^R(t) = b - c + \frac{\lambda(t)}{r}\int_{w^R(t)}^{\infty}\left(w - w^R(t)\right) dF(w) + \frac{dw^R(t)}{rdt}, \tag{A.36}$$

or alternatively:

$$\frac{dw^R(t)}{dt} = rw^R(t) - r(b-c) - \lambda(t)\int_{w^R(t)}^{\infty}\left(w - w^R(t)\right) dF(w). \tag{A.37}$$

A.6 Formulae from Section 4.1

The probability of becoming employed before time t can be written as:

$$P(T < t) = \sum_{m=0}^{\infty} \left\{ \frac{(\lambda t)^m e^{-\lambda t}}{m!} \left(1 - F^m \left(w^R\right)\right) \right\}. \tag{A.38}$$

By the definition of the Poisson distribution, $\sum_{m=0}^{\infty} \frac{(\lambda t)^m e^{-\lambda t}}{m!} = 1$, and, hence, $\sum_{m=0}^{\infty} \frac{(\lambda t)^m}{m!} = e^{\lambda t}$. One may rewrite Equation A.38 as:

$$P(T < t) = 1 - e^{-\lambda t} \sum_{m=0}^{\infty} \frac{\left(\lambda t F \left(w^R\right)\right)^m}{m!}. \tag{A.39}$$

Since $\sum_{m=0}^{\infty} \frac{\left(\lambda t F \left(w^R\right)\right)^m}{m!} = e^{\lambda t F(w^R)}$ then $P(T < t) = 1 - e^{-\lambda t(1-F(w^R))}$.

A.7 Formulae from Section 5.2

The value of search function in the bilocational model is given as:

$$\begin{aligned}
&\Omega = \Big(b - c_A(\theta_A) - c_B(\theta_B)\Big)\tau + \\
&\beta(\tau)\Big\{\textstyle\sum_{n=1}^{\infty}\sum_{m=1}^{\infty} q_A(m,\tau) q_B(n,\tau) \int_0^{\infty} \max(W(x),\Omega) dG(x) + \\
&+ \textstyle\sum_{n=1}^{\infty} q_A(n,\tau) q_B(0,\tau) \int_0^{\infty} \max(W(x),\Omega) dG_A(x;n) + \\
&+ \textstyle\sum_{n=1}^{\infty} q_B(n,\tau) q_A(0,\tau) \int_0^{\infty} \max\left(W(x) - \frac{\delta}{r}, \Omega\right) dG_B(x;n) + q_B(0,\tau) q_A(0,\tau)\Omega\Big\}.
\end{aligned} \tag{A.40}$$

We can rewrite A.40 as:

$$\Omega = \Big(b - c_A(\theta_A) - c_B(\theta_B)\Big)\tau + \beta(\tau)\Big\{\sum_{n=1}^{\infty}\sum_{m=1}^{\infty} q_A(m,\tau)q_B(n,\tau)\int_0^{\infty}\max(W(x)-\Omega,0)dG(x)+$$
$$+\sum_{n=1}^{\infty} q_A(n,\tau)q_B(0,\tau)\int_0^{\infty}\max(W(x)-\Omega,0)dG_A(x;n)+$$
$$+\sum_{n=1}^{\infty} q_B(n,\tau)q_A(0,\tau)\int_0^{\infty}\max\left(W(x)-\frac{\delta}{r}-\Omega,0\right)dG_B(x;n)+q_B(0,\tau)q_A(0,\tau)\Omega+$$
$$+\sum_{n=1}^{\infty}\sum_{m=1}^{\infty} q_A(m,\tau)q_B(n,\tau)\Omega+\sum_{n=1}^{\infty} q_A(n,\tau)q_B(0,\tau)\Omega+\sum_{n=1}^{\infty} q_B(n,\tau)q_A(0,\tau)\Omega\Big\}. \tag{A.41}$$

Consequently:

$$\Omega = \Big(b - c_A(\theta_A) - c_B(\theta_B)\Big)\tau + \beta(\tau)\Big\{\sum_{n=1}^{\infty}\sum_{m=1}^{\infty} q_A(m,\tau)q_B(n,\tau)\int_0^{\infty}\max(W(x)-\Omega,0)dG(x)+$$
$$+\sum_{n=1}^{\infty} q_A(n,\tau)q_B(0,\tau)\int_0^{\infty}\max(W(x)-\Omega,0)dG_A(x;n)+$$
$$+\sum_{n=1}^{\infty} q_B(n,\tau)q_A(0,\tau)\int_0^{\infty}\max\left(W(x)-\frac{\delta}{r}-\Omega,0\right)dG_B(x;n)+\Omega\Big\}. \tag{A.42}$$

Since $q_A(m,\tau)q_B(m,\tau) = \dfrac{e^{-\lambda_A(\theta_A)\tau}(\lambda_A(\theta_A)\tau)^m}{m!} \times \dfrac{e^{-\lambda_B(\theta_B)\tau}(\lambda_B(\theta_B)\tau)^n}{n!}$ it is straightforward to show that $\lim_{\tau=0}\dfrac{q_A(m,\tau)q_B(n,\tau)}{\tau} = 0$, when both $m \geqslant 1$ and $n \geqslant 1$. Hence, the term $\dfrac{1}{\tau}\displaystyle\sum_{n=1}^{\infty}\sum_{m=1}^{\infty} q_A(m,\tau)q_B(n,\tau)\int_0^{\infty}\max(W(x)-\Omega,0)dG(x) = 0$ in the limit. Moreover, $q_A(n,\tau)q_B(0,\tau) = \dfrac{e^{-\lambda_A(\theta_A)\tau}(\lambda_A(\theta_A)\tau)^n}{n!}e^{-\lambda_B(\theta_B)\tau}$ and, hence,
$\lim_{\tau=0} q_A(n,\tau)q_B(0,\tau) = \lambda_A(\theta_A)$ if $n = 1$ and $\lim_{\tau=0} q_A(n,\tau)q_B(0,\tau) = 0$ for $n > 1$. In the same fashion $\lim_{\tau=0} q_B(n,\tau)q_A(0,\tau) = \lambda_B(\theta_B)$ if $n = 1$ and $\lim_{\tau=0} q_B(n,\tau)q_A(0,\tau) = 0$ for $n > 1$.

Reservation wages are chosen to maximize the value of search. Knowing that it is easy to show that $W(w_A^R) = \Omega$ or $w_A^R = r\Omega$ and similarly $w_B^R = r\Omega - \delta$.

Dividing Equation A.42 by τ and taking the limit at $\tau = 0$ yields Equation in 5.4.

A.8 Proof of Proposition 1

The reservation wage equation for the region A is given by:

$$w_A^R = b - c_A(\theta_A) - c_B(\theta_B) + \frac{\lambda_A(\theta_A)}{r} \int_{w_A^R}^{\infty} (w - w_A^R) dF_A(w) + \\ + \frac{\lambda_B(\theta_B)}{r} \int_{w_B^R}^{\infty} (w - w_B^R) dF_B(w). \tag{A.43}$$

Taking the derivative with respect to δ yields:

$$\frac{\partial w_A^R}{\partial \delta} = -\frac{\partial w_A^R}{\partial \delta} \frac{\lambda_A(\theta_A)}{r} \left(1 - F_A\left(w_A^R\right)\right) - \left(\frac{\partial w_A^R}{\partial \delta} + 1\right) \frac{\lambda_B(\theta_B)}{r} \left(1 - F_B\left(w_B^R\right)\right) = zinb_w \\ = \frac{-\lambda_B(\theta_B)\left(1 - F_B\left(w_B^R\right)\right)}{r + \lambda_A(\theta_A)\left(1 - F_A\left(w_A^R\right)\right) + \lambda_B(\theta_B)\left(1 - F_B\left(w_B^R\right)\right)}. \tag{A.44}$$

It immediately follows that $-1 \leqslant \frac{\partial w_A^R}{\partial \delta} \leqslant 0$. Having $w_A^R + \delta = w_B^R$ one obtains:

$$\frac{\partial w_B^R}{\partial \delta} = \frac{r + \lambda_A(\theta_A)\left(1 - F_A\left(w_A^R\right)\right)}{r + \lambda_A(\theta_A)\left(1 - F_A\left(w_A^R\right)\right) + \lambda_B(\theta_B\left(1 - F_B\left(w_B^R\right)\right)}, \tag{A.45}$$

and therefore, $0 \leqslant \frac{\partial w_B^R}{\partial \delta} \leqslant 1$.

A.9 Proof of Proposition 2

To derive the effect of the change of the median on the reservation wage one needs to introduce the notion of the translation of the distribution. Changing the median holding the shape of the distribution constant is simply a parallel shift of the distribution. If we increase the median of the distribution $F(x)$ by the value μ, the resulting distribution would be a translation of the original c.d.f. $F(x)$. The distribution $G(x)$ is a translation of $F(x)$ if $G(x + \mu) = F(x)$ and, hence, $G(x) = F(x - \mu)$.

Given that $w_B^R = w_A^R + \delta$, the reservation wage in the region A is given as:

$$w_A^R = b - c_A(\theta_A) - c_B(\theta_B) + \frac{\lambda_A(\theta_A)}{r}\int_{w_A^R}^{\infty}(w - w_A^R)dF_A(w) + \frac{\lambda_B(\theta_B)}{r}\int_{w_A^R+\delta}^{\infty}(w - w_A^R - \delta)dF_B(w). \tag{A.46}$$

Increase of the median in region A by μ would result in a new reservation wage:

$$\begin{aligned} &w_A^R(\mu) = b - c_A(\theta_A) - c_B(\theta_B) + \frac{\lambda_A(\theta_A)}{r}\int_{w_A^R(\mu)}^{\infty}(w - w_A^R(\mu))dF_A(w-\mu) + \\ &\frac{\lambda_B(\theta_B)}{r}\int_{w_A^R(\mu)+\delta}^{\infty}(w - w_A^R(\mu) - \delta)dF_B(w). \end{aligned} \tag{A.47}$$

Subtracting A.47 from A.46 we obtain:

$$\begin{aligned} &w_A^R(\mu) - w_A^R = \\ &\frac{\lambda_A(\theta_A)}{r}\Big[\int_{w_A^R(\mu)}^{\infty}(w - w_A^R(\mu))dF_A(w-\mu) - \int_{w_A^R}^{\infty}(w - w_A^R)dF_A(w)\Big] + \\ &\frac{\lambda_B(\theta_B)}{r}\Big[\int_{w_A^R(\mu)+\delta}^{\infty}(w - w_A^R(\mu) - \delta)dF_B(w) - \int_{w_A^R+\delta}^{\infty}(w - w_A^R - \delta)dF_B(w)\Big] \end{aligned} \tag{A.48}$$

By integration by parts one obtains:

$$\int_{w_A^R(\mu)}^{\infty}(w - w_A^R(\mu))dF_A(w-\mu) = E_A(w) + \mu - w_A^R(\mu) + \int_0^{w_A^R(\mu)} F_A(w-\mu)dw \tag{A.49}$$

$$\int_{w_A^R}^{\infty}(w - w_A^R)dF_A(w) = E_A(w) - w_A^R + \int_0^{w_A^R} F_A(w)dw \tag{A.50}$$

$$\int_{w_A^R(\mu)+\delta}^{\infty}(w - w_A^R(\mu) - \delta)dF_B(w) = E_B(w) - w_A^R(\mu) - \delta + \int_0^{w_A^R(\mu)+\delta} F_B(w)dw \tag{A.51}$$

$$\int_{w_A^R+\delta}^{\infty}(w - w_A^R - \delta)dF_B(w) = E_B(w) - w_A^R - \delta + \int_0^{w_A^R+\delta)} F_B(w)dw. \tag{A.52}$$

Hence,

$$
\begin{aligned}
&w_A^R(\mu) - w_A^R = \\
&\frac{\lambda_A(\theta_A)}{r}\Big(\mu - w_A^R(\mu) + w_A^R + \int_0^{w_A^R(\mu)} F_A(w-\mu)dw - \int_0^{w_A^R} F_A(w)dw\Big) + \\
&\frac{\lambda_B(\theta_B)}{r}\Big(-w_A^R(\mu) + w_A^R + \int_0^{w_A^R(\mu)+\delta} F_B(w)dw - \int_0^{w_A^R+\delta} F_B(w)dw\Big).
\end{aligned} \tag{A.53}
$$

Dividing the expression in A.53 by μ and taking the limit at $\mu = 0$ with the help of the results obtained in A.49 - A.52 one gets:

$$
\begin{aligned}
&\frac{\partial w_A^R(\mu)}{\partial \mu} = \\
&\frac{\lambda_A(\theta_A)}{r}\Big(1 - \frac{\partial w_A^R(\mu)}{\partial \mu}(1 - F_A(w_A^R)) - \int_0^{w_A^R} f_A(w)dw\Big) + \\
&\frac{\lambda_B(\theta_B)}{r}\Big(-\frac{\partial w_A^R(\mu)}{\partial \mu}(1 - F_B(w_A^R + \delta))\Big)
\end{aligned} \tag{A.54}
$$

Therefore:

$$
\frac{\partial w_A^R(\mu)}{\partial \mu} = \frac{\lambda_A(\theta_A)(1 - F_A(w_A^R))}{r + \lambda_A(\theta_A)(1 - F_A(w_A^R)) + \lambda_B(\theta_B)(1 - F_B(w_A^R + \delta))}. \tag{A.55}
$$

Obviously $0 \leqslant \frac{\partial w_A^R(\mu)}{\partial \mu} \leqslant 1$. Knowing that $w_A^R + \delta = w_B^R$ one obtains $\frac{\partial w_A^R}{\partial \mu} = \frac{\partial w_B^R}{\partial \mu}$.

In the same fashion, increasing the median in B by μ:

$$
\frac{\partial w_B^R}{\partial \mu} = \frac{\lambda_B(\theta_B)(1 - F_B(w_B^R))}{r + \lambda_B(\theta_B)(1 - F_B(w_B^R)) + \lambda_A(\theta_A)(1 - F_A(w_A^R))}. \tag{A.56}
$$

Again, $0 \leqslant \frac{\partial w_B^R}{\partial \mu} \leqslant 1$.

For the effect of the spreads, assume that the reservation wage is below the median. Denote $\Lambda = \int_{w^R}^{\infty}(w - w^R)dF(w)$. One could rewrite:

$$
\begin{aligned}
\Lambda &= \int_{w^R}^{\bar{w}} (w - w^R)dF(w) + \int_{\bar{w}}^{\infty}(w - w^R)dF(w) = \\
&= \tfrac{\bar{w}}{2} - \int_{w^R}^{\bar{w}} F(w)dw + \int_{\bar{w}}^{\infty} wdF(w) - w^R.
\end{aligned} \tag{A.57}
$$

Note that $\frac{\partial}{\partial \bar{w}} \int_{\bar{w}}^{\infty} wdF(w) > 0$. This result is intuitively clear – truncated mean increases if you move the truncation point to the right. Moreover, $\frac{\partial}{\partial \sigma_R} \int_{\bar{w}}^{\infty} wdF(w) > 0$ – truncated mean

increases if you increase the variance to the right of the truncation point. Hence, $\frac{\partial \Lambda}{\partial \sigma_R} > 0$.

The effect of the spread in the left tail is: $\frac{\partial \Lambda}{\partial \sigma_L} = -\int_{w^R}^{\bar{w}} \frac{\partial}{\partial \sigma_L} F(w) dw < 0$. The logic here is straightforward – increasing the spread in the left tail moves some of the probability mass away to the left of the reservation wage (fewer jobs become attractive). As a result, the reservation wage declines to compensate for the loss of the probability mass.

Hence, $\frac{\partial w_A^R}{\partial \sigma_{AR}} > 0$ and $\frac{\partial w_A^R}{\partial \sigma_{AL}} < 0$. In the same fashion one obtains $\frac{\partial w_B^R}{\partial \sigma_{BR}} > 0$ and $\frac{\partial w_B^R}{\partial \sigma_{BL}} < 0$.

Given the relationship $w_A^R + \delta = w_B^R$ one also obtains: $\frac{\partial w_B^R}{\partial \sigma_{AR}} > 0$ and $\frac{\partial w_B^R}{\partial \sigma_{AL}} < 0$, and $\frac{\partial w_A^R}{\partial \sigma_{BR}} > 0$ and $\frac{\partial w_A^R}{\partial \sigma_{BL}} < 0$.

A.10 Proof of Proposition 3

Differentiate $c_A'(\theta_A) = \frac{\lambda_A'(\theta_A)}{r} \int_{w_A^R}^{\infty} (w - w_A^R) dF_A(w)$ with respect to δ:

$$c_A''(\theta_A) \frac{\partial \theta_A}{\partial \delta} = \frac{\lambda_A''(\theta_A)}{r} \frac{\partial \theta_A}{\partial \delta} \int_{w_A^R}^{\infty} (w - w_A^R) dF_A(w) - \frac{\partial w_A^R}{\partial \delta} \Big(1 - F_A(w_A^R)\Big) \frac{\lambda_A'(\theta_A)}{r}. \quad \text{(A.58)}$$

Hence,

$$\frac{\partial \theta_A}{\partial \delta} = \frac{-\frac{\partial w_A^R}{\partial \delta} \Big(1 - F_A(w_A^R)\Big) \frac{\lambda_A'(\theta_A)}{r}}{c_A''(\theta_A) - \frac{\lambda_A''(\theta_A)}{r} \int_{w_A^R}^{\infty} (w - w_A^R) dF_A(w)} > 0, \quad \text{(A.59)}$$

since $\frac{\partial w_A^R}{\partial \delta} < 0$, $c_A''(\theta_A) > 0$, and $\lambda_A''(\theta_A) < 0$.

For search intensity in region B we can write:

$$c_B'(\theta_B) = \frac{\lambda_B'(\theta_B)}{r} \int_{w_B^R}^{\infty} (w - w_B^R) dF_B(w). \quad \text{(A.60)}$$

And hence:

$$\frac{\partial \theta_B}{\partial \delta} = \frac{-\frac{\partial w_B^R}{\partial \delta}\Big(1 - F_B(w_B^R)\Big)\frac{\lambda_B'(\theta_B)}{r}}{c_B''(\theta_B) - \frac{\lambda_B''(\theta_B)}{r}\int\limits_{w_B^R}^{\infty}(w - w_B^R)dF_B(w)} < 0. \tag{A.61}$$

A.11 Proof of Proposition 4

Differentiate $c_A'(\theta_A) = \frac{\lambda_A'(\theta_A)}{r}\int_{w_A^R}^{\infty}(w - w_A^R)dF_A(w)$ with respect to Λ_A:

$$c_A''(\theta_A)\frac{\partial \theta_A}{\partial \Lambda_A} = \frac{\lambda_A''(\theta_A)}{r}\frac{\partial \theta_A}{\partial \Lambda_A}\Lambda_A + \frac{\lambda_A'(\theta_A)}{r}, \tag{A.62}$$

and

$$\frac{\partial \theta_A}{\partial \Lambda_A} = \frac{\lambda_A(\theta_A)}{rc_A''(\theta_A) - \lambda_A''(\theta_A)\Lambda_A} > 0. \tag{A.63}$$

From $\frac{\partial \theta_A}{\partial \bar{w}_A} = \frac{\partial \theta_A}{\partial \Lambda_A}\cdot\frac{\partial \Lambda_A}{\partial \bar{w}_A}$ it immediately follows that $\frac{\partial \theta_A}{\partial \bar{w}_A} > 0$, $\frac{\partial \theta_A}{\partial \sigma_{AR}} > 0$, and $\frac{\partial \theta_A}{\partial \sigma_{AL}} < 0$.

Differentiate $c_A'(\theta_A) = \frac{\lambda_A'(\theta_A)}{r}\int_{w_A^R}^{\infty}(w - w_A^R)dF_A(w)$ with respect to $\bar{w}_A$:

$$c_B''(\theta_B)\frac{\partial \theta_B}{\partial \bar{w}_A} = \frac{\lambda_B''(\theta_B)}{r}\frac{\partial \theta_B}{\partial \bar{w}_A}\Lambda_B - \frac{\partial w_B^R}{\partial \bar{w}_A}\Big(1 - F_B(w_B^R)\Big)\frac{\lambda_B'(\theta_B)}{r}, \tag{A.64}$$

and hence, $\frac{\partial \theta_B}{\partial \bar{w}_A} = -\frac{\frac{\partial w_B^R}{\partial \bar{w}_A}(1 - F_B(w_B^R))\lambda_B'(\theta_B)}{rc_B''(\theta_B) - \lambda_B''(\theta_B)\Lambda_B} < 0.$

In the same fashion: $\frac{\partial \theta_B}{\partial \sigma_{AR}} = -\frac{\frac{\partial w_B^R}{\partial \sigma_{AR}}(1 - F_B(w_B^R))\lambda_B'(\theta_B)}{rc_B''(\theta_B) - \lambda_B''(\theta_B)\Lambda_B} < 0$

and $\frac{\partial \theta_B}{\partial \sigma_{AL}} = -\frac{\frac{\partial w_B^R}{\partial \sigma_{AL}}(1 - F_B(w_B^R))\lambda_B'(\theta_B)}{rc_B''(\theta_B) - \lambda_B''(\theta_B)\Lambda_B} > 0.$

If we want to see how search intensities react to changes in wages in both regions simultaneously, we simply let the wage distributions in both regions be identical. Then increase in wages

in region A would mean the same increase in region B. Then, $c'_A(\theta_A) = \frac{\lambda'_A(\theta_A)}{r} \int_{w_A^R}^{\infty} (w - w_A^R) dF(w)$ and $c'_B(\theta_B) = \frac{\lambda'_B(\theta_B)}{r} \int_{w_B^R}^{\infty} (w - w_B^R) dF(w)$. It is then easy to show that $\frac{\partial \theta_A}{\partial \bar{w}} > 0$, $\frac{\partial \theta_A}{\partial \sigma_R} > 0$, $\frac{\partial \theta_A}{\partial \sigma_L} < 0$ and $\frac{\partial \theta_B}{\partial \bar{w}} > 0$, $\frac{\partial \theta_B}{\partial \sigma_R} > 0$, $\frac{\partial \theta_B}{\partial \sigma_L} < 0$.

A.12 Data Used

The description of the IABS data set is taken from Möller and Aldashev (2006a). The data on wages and wage dispersion were calculated from IABS-REG. IABS-REG is a 2% random sample from the employment register of the Federal Labor Office with regional information. The data set includes all workers, salaried employees and trainees obliged to pay social security contributions and covers more than 80% of all employment. Excluded are public servants, minor employment and family workers (see Bender, Haas, and Klose (2000) for an extensive description of the data). Because of legal sanctions for misreporting, the earnings information in the data is highly reliable. Among others, IABS-REG contains variables on individual earnings and skills. The regional information is based on the employer. For the empirical analysis the data were restricted to full-time workers of the intermediate skill group (apprenticeship completed without a university-type of education). All male and female workers were selected that were employed on June 30th, 1997. For all regions the median wage and the second and eighth decile of daily earnings were calculated.

INKAR database of the Federal Office for Building and Regional Planning contains basic geographic and demographic indicators on regional level. Regional population used in estimation in Chapter 5 were taken from the INKAR dataset.

The data on commuting time was produced by the Institute for Regional Planning of the University of Dortmund (IRPUD). Using the data on daily commuters who report their travel time from home to workplace, they calculate average travel time between each pair of regions (more see in Spiekermann, Lemke, and Schürmann (2000) and citations therein).

A.13 Maximal Acceptable Commuting Distance, Simulations

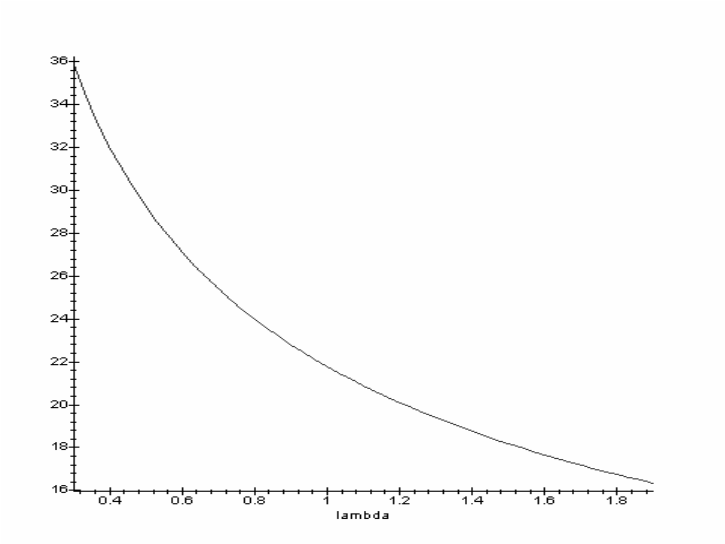


Figure A.1: Effect of a change in the arrival rate on maximal acceptable commuting distance

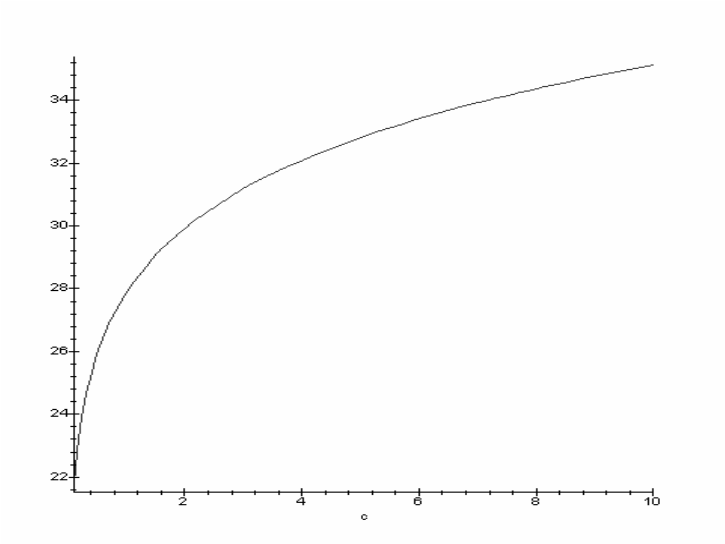


Figure A.2: Effect of a change in the search cost on maximal acceptable commuting distance

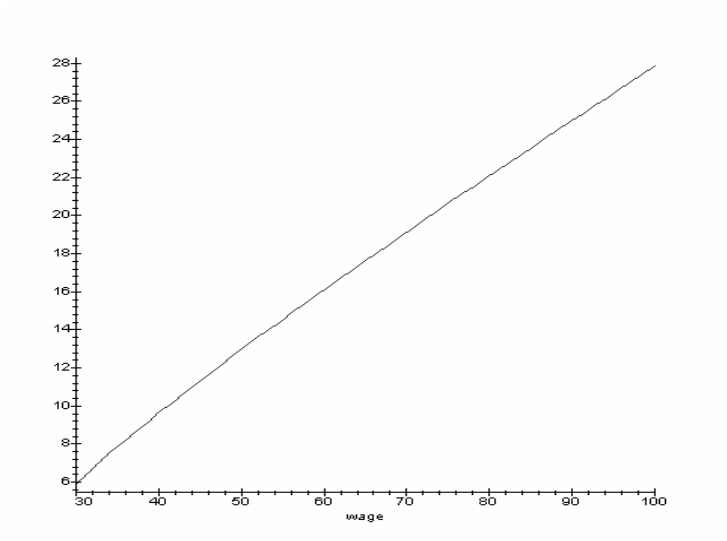


Figure A.3: Effect of a change in the median wage (both in origin and destination) on maximal acceptable commuting distance

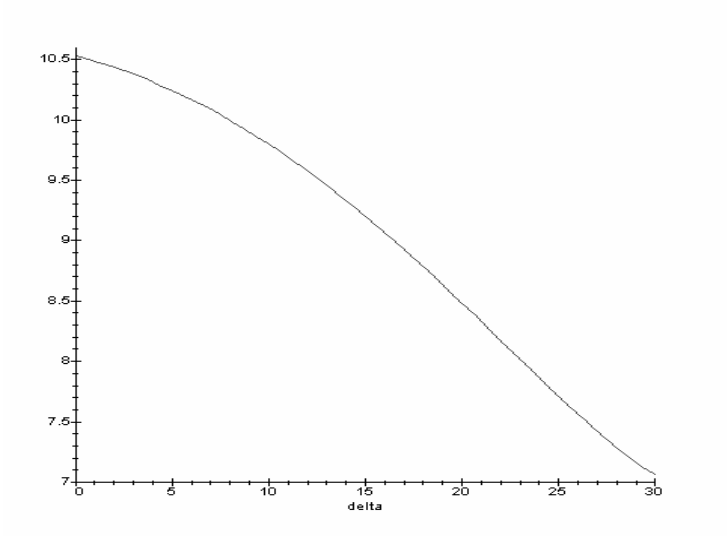


Figure A.4: Effect of a change in the travel cost on joint search intensity in both regions

A.14 Zero-inflated negative binomial estimation

Table A.1: Estimation results of the zero-inflated negative binomial model (East dummy (origin and destination) is interacted with parameters of the wage distribution). Dependent variable - commuting flows.

variable	**coef.**	**robust st. er.**
East origin	**5.09**	2.16
East destination	−3.30	4.05
log travel time	**−2.66**	0.02
log median wage (origin)	−0.05	0.25
log D5/D2 (origin)	0.72*	0.44
log D8/D5 (origin)	**1.56**	0.70
log median wage (destination)	**3.00**	0.51
log D5/D2 (destination)	**−2.61**	0.96
log D8/D5 (destination)	**12.68**	1.63
log POP (origin)	**0.99**	0.02
log POP (destination)	**1.00**	0.05
spatially weighted POP (×100)	**−0.17**	0.03
spatially weighted median wage	**−0.12**	0.01
log median wage (East origin)	−0.69	0.43
log D5/D2 (East origin)	−0.85	0.93
log D8/D5 (East origin)	**−4.12**	1.37
log median wage (East destination)	1.55*	0.82
log D5/D2 (East destination)	2.42	2.00
log D8/D5 (East destination)	**−13.53**	2.70
const	**−10.21**	1.45
inflate		

continued...

Table A.1: Estimation results of the zero-inflated negative binomial model (East dummy (origin and destination) is interacted with parameters of the wage distribution). Dependent variable - commuting flows.

variable	**coef.**	**robust st. er.**
East origin	-10.39	35.80
East destination	101.05	76.20
log travel time	$\mathbf{0.95}$	0.25
log median wage (origin)	-1.08	5.16
log D5/D2 (origin)	-4.37	4.78
log D8/D5 (origin)	-13.78^{*}	7.42
log median wage (destination)	24.05	15.09
log D5/D2 (destination)	$\mathbf{-240.64}$	113.38
log D8/D5 (destination)	185.36	121.93
log POP (origin)	$\mathbf{-1.16}$	0.12
log POP (destination)	$\mathbf{-1.68}$	0.27
spatially weighted POP ($\times 100$)	$\mathbf{-2.80}$	0.56
spatially weighted median wage	$\mathbf{0.91}$	0.20
log median wage (East origin)	-3.71	7.85
log D5/D2 (East origin)	8.39	8.69
log D8/D5 (East origin)	19.90^{*}	13.87
log median wage (East destination)	-23.63	14.53
log D5/D2 (East destination)	$\mathbf{235.78}$	112.12
log D8/D5 (East destination)	-184.89	120.12
const	-82.53	104.95
α	$\mathbf{1.21}$	0.02
Vuong test	12.43	
N	190 532	

Coefficients significant at 5% level are in bold.

* Coefficients significant at 10% level.

A.15 Descriptive Statistics

Table A.2: Selected descriptive statistics of the data

variable	**mean**	**st. dev.**
travel time by car (in minutes)	259.44	114.08
median wage (in logs)	4.63	0.10
population (in thousands)	179.53	148.63
D8/D5	0.31	0.03
D5/D2	0.38	0.05

Total number of stayers : 18,226,796.
Total number of commuters : 8,948,362.
Number of zeros : 62,775.

Table A.3: Selected descriptive statistics (West and East Germany).

	West			East		
variable	mean	st. dev.	N	mean	st. dev.	N
wage	107.34	7.78	327	90.66	5.91	110
population	197.40	164.32	327	126.97	63.53	110
unemployment rate	8.02	8.49	327	10.70	5.02	110
travel time (minutes)	256.80	39.52	327	267.08	49.16	110

Table A.4: Commuter flows between West and East German regions

	Means		Sums	
	West	**East**	**West**	**East**
West	66.3	3.2	$7,112,845$	$117,992$
East	12.3	101.7	$453,397$	$1,264,128$

Table A.5: Pairwise correlation coefficients

	1	2	3	4	5	6	7	8	9
(1) log D8/D5 (origin)	1.00								
(2) log D8/D5 (destination)	−0.00	1.00							
(3) log travel time	−0.03	−0.03	1.00						
(4) East origin	−0.80	0.00	0.05	1.00					
(5) East destination	0.00	−0.80	0.05	−0.00	1.00				
(6) log D5/D2 (origin)	0.72	−0.00	−0.01	−0.77	0.00	1.00			
(7) log D5/D2 (destination)	−0.00	0.72	−0.01	0.00	−0.77	−0.00	1.00		
(8) log median wage (destination)	−0.00	0.53	−0.06	0.00	−0.72	−0.00	0.47	1.00	
(9) log median wage (origin)	0.53	−0.00	−0.06	−0.72	0.00	0.47	−0.00	−0.00	1.00

A.16 Withdrawals from the Labor Market

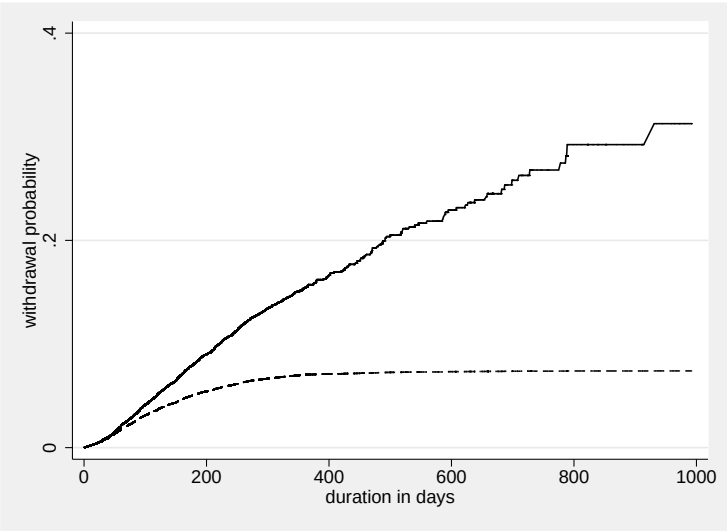


Figure A.5: Single females, West Germany. Solid line - conditional, dashed line - unconditional withdrawal probability.

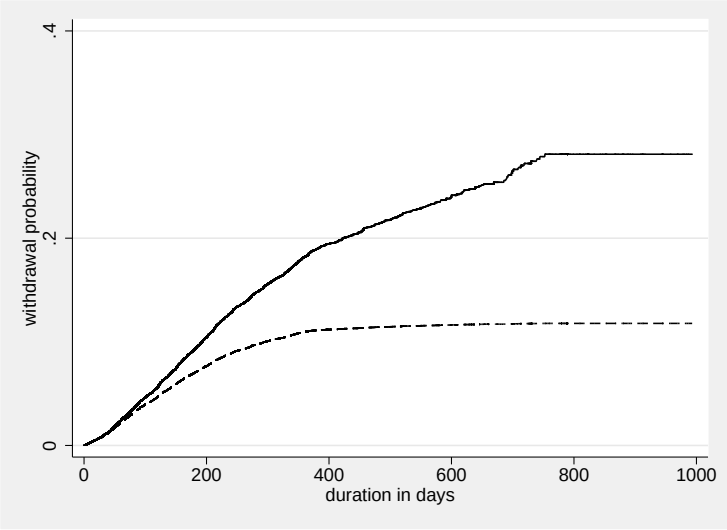


Figure A.6: Married females, West Germany. Solid line - conditional, dashed line - unconditional withdrawal probability.

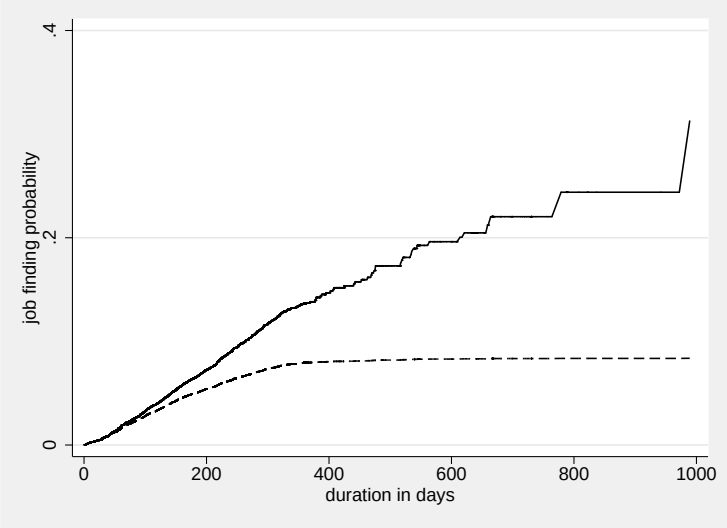


Figure A.7: Single females, East Germany. Solid line - conditional, dashed line - unconditional withdrawal probability.

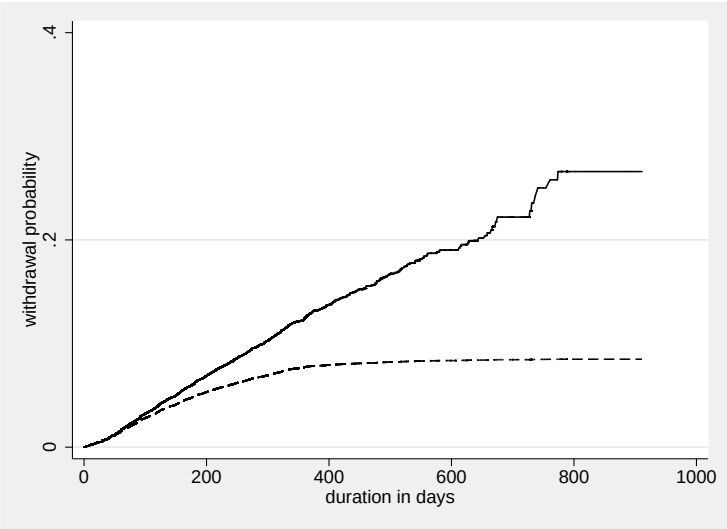


Figure A.8: Married females, East Germany. Solid line - conditional, dashed line - unconditional withdrawal probability.

A.17 Alternative Specifications

Table A.6: Estimation results of the Poisson model. Dependent variable - commuting flows.

variable	**coef.**	**robust st. er.**
East origin	**0.39**	0.15
East destination	0.16	0.19
travel time	−0.02	0.00
log median wage (origin)	−0.77*	0.40
log D5/D2 (origin)	**3.15**	0.70
log D8/D5 (origin)	**−4.39**	1.24
log median wage (destination)	**5.08**	0.53
log D5/D2 (destination)	**−3.77**	1.06
log D8/D5 (destination)	**3.61**	1.42
POP (origin)	**1.80**	0.20
POP (destination)	**3.23**	0.32
spatially weighted POP (×100)	**−0.29**	0.07
spatially weighted median wage	−0.05	0.04
const	**−10.07**	4.46
R^2	0.76	
N	190 532	

Table A.7: Estimation results of the negative binomial model. Dependent variable - commuting flows.

variable	**coef.**	**robust st. er.**
East origin	**0.52**	0.10
East destination	**0.67**	0.10
travel time	**−0.02**	0.00
log median wage (origin)	0.07	0.31
log D5/D2 (origin)	**2.12**	0.56
log D8/D5 (origin)	0.08	0.91
log median wage (destination)	**4.67**	0.56
log D5/D2 (destination)	−0.33	0.97
log D8/D5 (destination)	**7.47**	1.62
POP (origin)	**3.62**	0.46
POP (destination)	**3.23**	0.32
spatially weighted POP ($\times 100$)	−0.09	0.06
spatially weighted median wage	**−0.20**	0.02
α	**2.23**	0.03
R^2	0.23	
N	190 532	

Table A.8: Estimation results of the zero-inflated negative binomial model (Origin West Germany). Dependent variable - commuting flows.

variable	**coef.**	**robust st. er.**
East destination	0.10	0.20
travel time $\times 100$	**−1.39**	0.03
log median wage (origin)	0.17	0.37
log D5/D2 (origin)	**2.48**	0.66

continued...

Table A.8: Estimation results of the zero-inflated negative binomial model (Origin West Germany). Dependent variable - commuting flows.

variable	**coef.**	**robust st. er.**
log D8/D5 (origin)	−0.21	1.26
log median wage (destination)	4.45	0.58
log D5/D2 (destination)	−1.80*	1.07
log D8/D5 (destination)	**9.22**	1.76
POP (origin)	**3.05**	0.43
POP (destination)	**2.93**	0.28
spatially weighted POP (×100)	−0.08	0.05
spatially weighted median wage	**−0.17**	0.03
inflate		
East destination	−0.07	0.49
travel time ×100	**0.50**	0.04
log median wage (origin)	1.64	2.14
log D5/D2 (origin)	−0.07	2.24
log D8/D5 (origin)	**−10.19**	3.17
log median wage (destination)	−1.89	2.45
log D5/D2 (destination)	−1.71	3.60
log D8/D5 (destination)	−5.02	5.61
POP (origin)	**−34.33**	2.42
POP (destination)	**−23.86**	3.29
spatially weighted POP (×100)	0.04	0.33
spatially weighted median wage	0.02	0.12
Vuong test		17.02

* Coefficients significant at 10% level.

Coefficients significant at 5% level are in bold.

Table A.9: Estimation results of the zero-inflated negative binomial model (Origin East Germany). Dependent variable - commuting flows.

variable	**coef.**	**robust st. er.**
East destination	**1.37**	0.31
travel time times 100	**−0.01**	0.00
log median wage (origin)	−0.08	0.46
log D5/D2 (origin)	−1.05	0.78
log D8/D5 (origin)	2.78	2.10
log median wage (destination)	**3.92**	0.65
log D5/D2 (destination)	1.51	1.26
log D8/D5 (destination)	2.78	2.10
POP (origin)	**5.22**	0.55
POP (destination)	**3.34**	0.50
spatially weighted POP (×100)	**−1.11**	0.20
spatially weighted median wage	0.12	0.07
inflate		
East destination	−1.65	4.18
travel time	0.00	0.00
log median wage (origin)	5.17	7.94
log D5/D2 (origin)	−6.25	9.44
log D8/D5 (origin)	9.43	7.38
log median wage (destination)	−11.79*	7.29
log D5/D2 (destination)	−6.76	18.34
log D8/D5 (destination)	−6.93	36.11
POP (origin)	−25.23	5.67
POP (destination)	−40.50*	20.87

continued...

Table A.9: Estimation results of the zero-inflated negative binomial model (Origin East Germany). Dependent variable - commuting flows.

variable	**coef.**	**robust st. er.**
spatially weighted POP ($\times 100$)	-0.08	2.27
spatially weighted median wage	-0.02	0.86
Vuong test		7.71

* Coefficients significant at 10% level.
Coefficients significant at 5% level are in bold.

A.18 Software Used

The body text was compiled in MiKTeX 2.5 freeware using WinEdt v. 5.4, courtesy of WinEdt Inc. Tables were produced using Gnumeric v. 1.7.6 freeware and converted into LaTeX. Simulations and Figures 3.1, 3.2, 3.3, A.1, A.2, A.3 and A.4 were produced in Maple v. 10, courtesy of Waterloo Maple Inc. Estimations of duration and count models and Figures 4.1, 4.2, 4.3, 4.4, 4.5, A.5, A.6, A.7 and A.8 were produced in Stata v. 9.2, courtesy of StataCorp LP.