A Quantitative Approach to Overlapping Labour Market

ABDUL SULEMAN[†] FATIMA SULEMAN[‡] NÁDIA SIMÕES[‡]

[†] *ISCTE – Lisbon University Institute*

Department of Quantitative Methods, UNIDE

Av. Forcas Armadas,

1649 – 026 Lisbon – Portugal

(e-mail: abdul.suleman@iscte.pt)

[‡] *ISCTE – Lisbon University Institute*

Department of Economy, DINAMIA

Av. Forcas Armadas,

1649 – 026 Lisbon – Portugal

(e-mail: fatima.suleman@iscte.pt, nadia.simoes@iscte.pt)

Abstract: The concept of overlapping labour market involves either two types of workers competing for one type of job or one type of workers filling two types of jobs. These two perspectives are referred to as within jobs or between jobs overlapping, respectively. Although the concept of overlapping labour markets is a more realistic representation of labour market, the question about how overlapping can be measured is rather unanswered. The purpose of this study is to model overlapping and to measure its distribution at firm level. The model is based on the concept of partial membership of fuzzy sets theory. The extent of overlapping is determined statistically. Overlapping distribution is calculated through a ordered partition of sample data. The empirical analysis resorts to a sample of Portuguese LEED of two major banks and yields promising results. In both cases within jobs and between jobs overlapping were found for the stock of banking employees in 2006.

JEL classification numbers:

Keywords: Overlapping labour markets, fuzzy sets, grade of membership, fuzzy partition.

1. Introduction

What significant trends can we glean from examining the data on contemporary labour markets? An analysis of this data shows that similar jobs are held by workers with different characteristics (e.g. education, tenure). Sattinger (2006) suggests that the concept of overlapping labour markets (OLM, hereafter) is a more realistic representation of the actual labour market, as composed by heterogeneous workers and jobs. In this sense, distinction between ordinary labour market and OLM is introduced by the author. In the first one, workers' characteristics match the ones required by jobs, thus leading to frictionless assignments. Each worker is assigned to exactly one class according to his/her specific job characteristics.

In OLM, with heterogeneous workers and jobs, one type of workers competes for more than one type of jobs or otherwise employers fill one type of job with different type of workers. In summing up, “overlapping labour market arises in labour market with heterogeneous workers and jobs when direct interactions between particular types of workers and jobs do not occur” (Sattinger, 2006).

Empirical evidence seems to suggest that labour markets are more realistically described using the concept of overlapping labour markets. Hence, the question about how overlapping can be measured is rather unanswered. The purpose of this study is to present a methodology to measure the kinds and the intensity of overlapping at firm level. It is based on a fuzzy partition of labour market which structures it into job typologies. Fuzzy partitions allow individual heterogeneity representation by his/her grades of membership in different typologies. Thus, individuals with similar attributes may have (partial) membership in two or more typologies at the same time leading to between job overlapping. Or otherwise, as employers fill similar jobs with heterogeneous supply, within jobs overlap could also prevail.

The paper is organized as follows. In the next section (Section 2), we present some theoretical framework and possible explanations for overlapping. In Section 3, some basic concepts of fuzzy sets theory are described. In Section 4 we propose a mathematical model for the labour market and in sequel a methodology to decompose it by overlapping. In Section 5, we present the statistical model used to carry out the empirical analysis. The results achieved are analyzed in Section 6 for the stock of a sample of banks in 2006. The conclusions of this study are drawn in Section 7.

2. Theoretical framework: explaining overlapping

Sattinger (2006) assumes that the concept of overlapping labour market (OLM) is a more accurate representation of actual labour markets. In this sense, he distinguishes between ordinary and OLM. In an ordinary labour market, workers and jobs are the same and the assignments are frictionless. Moreover, ordinary labour market with heterogeneous agents is considered a collection of ordinary markets with homogeneous agents. Overlapping labour markets means two types of workers compete for one type of job or employers fill two types of jobs with one type of workers. In Sattinger’s words: workers of type s1 seek either k1 or k2 jobs, while employers with k2 jobs seek either s1 and s2 workers. In summing up, “overlapping labour market arises in labour market with heterogeneous workers and jobs when direct interactions between particular types of workers and jobs do not occur” (Sattinger, 2006). We can infer from Sattinger’s definition that two kinds of overlapping are suggested. The first leading to within jobs overlapping as heterogeneous supply compete for unique demand; the second allowing between job overlapping as two kinds of jobs are filled by homogeneous supply.

The paper suggests some reasons why the labour markets should be better classified as OLM. The theoretical models suggest different explanations for overlapping labour markets: the job search costs and hiring costs for workers and employers, respectively; skill composition of the workforce (supply-side characteristics); and the rules governing employment relationship (demand-side rules and procedures). The key developments in each of these areas are summarized below.

Several models have tried to analyze how vacancies and jobseekers are matched. These models differ in the type of workers and jobs considered (homogeneous/heterogeneous) and in the existence or absence of frictions, which are a key determinant of the efficiency of the matching process.

If workers and vacancies are homogeneous and the process of assignment of workers to jobs occurs without frictions, the market leads to an efficient outcome since there is a self-selection process (Tinbergen, 1951, 1956; Sattinger, 1975; Teulings, 1995). Highly skilled workers have an absolute advantage in all jobs. However they have a comparative advantage in more complex jobs. If wages reward complexity, a mapping of skills on complexity will arise, generating optimal assignments (Teulings, 1995). Wages will reflect skills differentials. This is classified as an ordinary labour market.

The labour markets can also be described through a two sided search model (Diamond, 1982; Mortensen, 1982; Pissarides, 1984). In this framework, a matching function is used to relate firms and workers searching for a new job (Moen, 1997; Shi, 2001; Mortensen and Wright, 2002). These models show ordinary labour markets with homogeneous agents.

One of the relevant issues to discuss is whether or not the market is able to produce efficient outcomes. The labour market fails to deliver efficient outcomes when the process of job matching is time consuming (Pissarides, 1990; Sattinger, 1995; Shimer and Smith, 1998). Usually information is incomplete and therefore workers have to search in order to identify job opportunities as well as skills requirements, wage, and other terms associated with the available vacancies; and firms also face some difficulties in hiring new workers. To make a new hiring, firms have to publicly announce job offers and then interview job applicants. This process is costly and therefore both sides of the market (workers and firms) weight costs and benefits of further search. When a match is formed, an externality arises since there is a worker and a vacancy no longer available that will condition the following assignments. If firms and workers use this type of job searching strategies, multiple equilibrium are possible, including inefficient outcomes (Sattinger, 1995).

Another issue that arises when we analyse the matching process is whether the outcome is positively assortative (i.e. complex technologies matched with high skilled workers). With frictions, the assignment of skills to jobs is not necessarily positively assortative which increases the difficulty to understand wage differentials between workers with different skills. Shi (2001) examining a framework where agents on the two sides of the market are heterogeneous and the matching process is time consuming concludes that the efficient matching between workers and vacancies is not always positively assortative.

Thus workers can choose jobs and set reservation wage, while employers can select workers from a pool of candidates and set skill requirements. Thus, OLM could be interpreted as adjustment process revealing the flexibility of labour market. In the context of oversupply, workers may switch their search to other markets; accept lower wages; or accept less attractive labour contracts (Wieling and Borghans, 1995). On the other hand, employers reveal their preference for higher over lower educated workers to fill jobs previously occupied by lower educated workers (Ours and Ridder, 1995). Moreover, in a context of high unemployment, employers raise their hiring standards (Okun, 1981 cited by Ours and Ridder, 1995) and contribute to increase the competition between individuals with different levels of education. Overeducation literature cast doubt on whether there is a demand for higher skills revealing the upgrading of job requirements or, otherwise, there is an underutilization of skills taking advantage of labour market conditions (Borghans and Grip, 1999).

After all, labour market reveals to be more flexible and this flexibility is reflected through the prevalence of different types of education or skill in a range of occupations or jobs, rather than in one specific occupation.

Besides the differences of view, the measurement of required education is a rather controversial issue (Meer, 2006). The answer demands particular attention on internal working of the firm.

Firms resort to job analysis to set the required attributes and use job evaluation to establish the hierarchy of jobs according to those requirements (for details see Schumann et al., 1994) and finally to define the wage structure. Hierarchies at the workplace have several functions (see Smeets and Warzinsky, 2008), but of particular interest is the assignment policy and earnings distribution as in internal labour market. Generally speaking, when analysing the link between hierarchy levels and wages, the authors found substantial pay variations both within and between levels or jobs (Baker, Gibbs and Holmstrom, 1994; Treble et al., 2001; Grund, 2002). Smeets and Warzinsky (2008) insist that the hierarchies are becoming flatter and firms are moving towards more unequal pay schemes. Seltzer and Merrett (2000) empirical evidence for the bank of Australia reveal that wages are related to jobs but the relationship is weak. This result leads the authors to insist that "wages are related to jobs but are not tied to jobs". At least, there is a remarkable overlapping of wages by job level (Seltzer and Merrett, 2000).

The question left open is how to deal with this overlapping in empirical analysis. If each employee could share the characteristics of two jobs, how to classify him/her in two different jobs to analyse the relationship between hierarchies of jobs and wages? Otherwise, how to define the job requirements if one type of job is filled by most heterogeneous workers? As noted, Sattinger's OLM concept incorporates two perspectives of overlapping. This paper intends to contribute to this analysis by suggesting an overlapping measurement. Our approach should be considered as an explorative and experimental approach to the within and between jobs overlapping measurement.

3. Fuzzy Sets

The notion of fuzzy set was introduced by Zadeh (1965), together with the concept of grade of membership. In the author's own words "a fuzzy set is a class of objects with a continuum grades of membership" with no sharp boundary "between the objects that belong to the class and those that do not" (Bellman and Zadeh, 1970). The basic idea of fuzzy sets theory is to represent real world problems that are characterized by ambiguity. The ambiguity is often a consequence of heterogeneity and may arise in different forms such as in natural language (e.g. pretty woman) or in objects matching hybrid conditions. A formal definition of fuzzy set is stated as follows.

Definition 1 Let $U = \{\boldsymbol{u}\}$ denote a collection of objects (points of $\Re^J$) generically referred by $\boldsymbol{u}$, a *J*-dimensional vector. A fuzzy set *A* of *U* is a set of ordered pairs

$$A = \{ (\boldsymbol{u}; g_A(\boldsymbol{u})),\ \boldsymbol{u} \in U \}$$

where the value $g_A(\boldsymbol{u}) \in [0,1]$ is termed grade of membership (GoM) or degree of compatibility of the element $\boldsymbol{u} \in U$ with the fuzzy set *A*. The mapping $g_A : U \to [0,1]$ is called membership function of the fuzzy set *A*. □

The membership function is a generalization of indicator function of classical sets. If it assumes only two values, 0 and 1, the fuzzy set becomes a crisp set.

The values 0 and 1 represent, respectively, the lowest and the highest grades of membership. When $g_A(\boldsymbol{u}) = 1$, $\boldsymbol{u}$ is an element of fuzzy set *A* in classical sense, i.e., a crisp element of *A*. If $g_A(\boldsymbol{u}) = 0$, $\boldsymbol{u}$ is said incompatible with *A*. If $g_A(\boldsymbol{u}) \in (0,1)$, $\boldsymbol{u}$ is a partial member of *A*. It means that the positive amount $1 - g_A(\boldsymbol{u})$ is part of $\boldsymbol{u}$ that belongs to other(s) fuzzy set(s) or, equivalently, $\boldsymbol{u}$ has partial membership in at least two fuzzy sets.

Definition 2 Suppose that the membership function of the fuzzy set *A* of *U* is g_A. The membership function of the complement of *A*, $\overline{A}$, is $g_{\overline{A}} = 1 - g_A$, i.e.

$$\overline{A} = \{ (\boldsymbol{u}; 1 - g_A(\boldsymbol{u})),\ \boldsymbol{u} \in U \} \quad \square$$

Following our previous comments, if $g_A(\boldsymbol{u}) \in (0,1)$, $\boldsymbol{u}$ belongs partially to the complement of *A*.

Definition 3 A fuzzy set *A* of *U* is said normal if and only if there exists $\boldsymbol{u} \in U$ such that $g_A(\boldsymbol{u}) = 1$ In other words, a normal fuzzy set has at least one crisp element. □

Definition 4 A finite collection of fuzzy sets of *U*, $A_1, A_2, ..., A_K$ with membership functions $g_{A_1}, g_{A_2}, ..., g_{A_K}$, respectively, is a finite fuzzy partition or a fuzzy *K*-partition of *U* if

i) Each A_k is a normal set; that is $g_{A_k}(\boldsymbol{u}) = 1,\ 1 \le k \le K,$ for some $\boldsymbol{u} \in U$;

ii) $\sum_{k=1}^{K} g_{A_k}(\boldsymbol{u}) = 1,\ \forall \boldsymbol{u} \in U$.□

The *K* fuzzy partition sets are often referred to as typologies or extreme profiles (Manton el al., 1994).

A fuzzy partition defines a structure to represent elements or objects of *U*. Each element $\boldsymbol{u} \in U$ may be then represented by the vector,

$$\boldsymbol{g}(\boldsymbol{u}) = (g_{A_1}(\boldsymbol{u}), g_{A_2}(\boldsymbol{u}), ..., g_{A_K}(\boldsymbol{u})) \tag{1}$$

that belongs to so-called unit simplex with *K* extreme points,

$$S_K = \left\{ a = (a_1, a_2, ..., a_K) :\ a_k \ge 0 \wedge \sum_{k=1}^{K} a_k = 1 \right\}. \tag{2}$$

The vector $\boldsymbol{g}(\boldsymbol{u})$ is intended to position the object $\boldsymbol{u}$ in *U*. The vector coordinates are measures of the object $\boldsymbol{u}$ intrinsic heterogeneity, as Nguyen and Walker (2000) point out.

The convex set S_K (2) is an analytical device to represent *U*. Its extreme points are assigned to crisp elements of the fuzzy partition sets $A_1, A_2, ..., A_K$ because they are represented by the canonical base of $\Re^K$, the vectors (1,0,…0), (0,1,…,0), …, (0,0,…,1). Thus, $\boldsymbol{g}(\boldsymbol{u})$ measures the position of object $\boldsymbol{u}$ in S_K relative to crisp elements. In short: the convex set S_K may be used to analyse the distribution of elements in *U* provided that this set is decomposed into a fuzzy *K*-partition.

It is possible to represent S_K graphically if the number of its extreme points is 2, 3 or 4. This special feature allows a richer insight about the heterogeneity of *U*. Recall that S_2 is a line segment, S_3 an equilateral triangle and S_4 a triangular pyramid or tetrahedron. It is worthy noting that the number of extreme points of S_K is, in general, lower than the dimension of objects of *U*, i.e., $K < J$, thus leading to a parsimonious representation of this set.

In the present study, *U* is a finite set as it is related to sample data. Then, we can write $U = \{ \boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_N \}$, where *N* is the number of elements of *U* and in practice it corresponds to the sample size. The objects of this set can be referred to as individuals or, particularly here, employees, and attached with a number. So, when appropriate, the usual notation $\boldsymbol{g}(\boldsymbol{X}_i) = (g_{A_1}(\boldsymbol{X}_i), g_{A_2}(\boldsymbol{X}_i),..., g_{A_K}(\boldsymbol{X}_i))$ for the grade of membership vector of individual *i* may be replaced by the simpler one $\boldsymbol{g}_i = (g_{i1}, g_{i2},..., g_{iK})$, $1 \le i \le N$ (Woodbury et al., 1978). The generic coordinate g_{ik}, $1 \le k \le K$, stands for the grade of membership of individual *i* in fuzzy set A_k or, for short, in *k*-fuzzy set.

4. Decomposition of Overlapping Labour Market

In this study we assume that *U* represents the labour market of interest and it can be decomposed into a fuzzy *K*-partition, $A_1, A_2,..., A_K$. Additionally, we assume that each A_k, $1 \le k \le K$, represents the set of characteristics associated to job category *k* or *k*-job and it will be referred to as *k*-job typology. Our goal is to quantify OLM through the partition sets membership functions.

It is important to note that, if within jobs overlapping exists, two or more job typologies may be assigned with the same job category. In this case, the corresponding typologies should be given additional attention. In what follows we place emphasis on between jobs overlapping yet another form of OLM.

Suppose now that the typologies are known, though in practice they are estimated from the data. We propose a simple method to crisply decompose *U* into ordered by overlapping subsets using a fuzzy partition of *U*. Recall that a crisp partition of *U* is any decomposition of this set into pairwise disjoints nonempty subsets whose union is *U*. Unlike fuzzy partition, in a crisp partition every element of *U* belongs to exactly one of those subsets.

In the context of the labour market, the coordinate g_{ik} of $\boldsymbol{g}_i = (g_{i1}, g_{i2},..., g_{iK})$ is then the amount of *k*-job characteristics shared by individual *i*. If $g_{ik} = 1$ his/her characteristics match exactly the *k*-job typology. This is to say, he/she is a crisp element of the fuzzy set A_k. Thus, no job overlapping occurs. By the definition of fuzzy partition, there are at least *K* crisp employees in *U*. Suppose now $0 < g_{ik} < 1$, for some $1 \le k \le K$. In this case, he/she shares partially job characteristics of A_k and partially the characteristics of the complement of A_k. This means, he/she also shares characteristics of at most *K*-1 remaining sets of the fuzzy partition. Then, there is job overlapping. A formal definition of this concept is stated as follows.

Definition 5 Suppose *U* is a finite subset of $\Re^J$ whose elements are referred to as 1,2,.., *N*. Let $A_1, A_2,..., A_K$ be a fuzzy partition of *U* and $\boldsymbol{g}_i = (g_{i1}, g_{i2},..., g_{iK})$, $1 \le i \le N$, the grades of membership vector of employee *i* relative to that partition. Suppose the fuzzy set $A_k, 1 \le k \le K$, is a k-job typology. Then, there exists between jobs overlapping if and only if $0 < g_{ik} < 1$ for some $1 \le k \le K$.□

To draw an ordered partition of U, the following method is used. Consider a fuzzy partition of dataset U, $A_1, A_2, ..., A_K$. Call $U^{(1)}$ the set of elements of U such that $g_{ik} = 1, 1 \le k \le K$, i.e. the set of crisp elements of the fuzzy partition. Then, $U^{(1)}$ is the set of all employees of U with no job overlapping. These individuals position on the vertices of unit simplex S_K as referred above.

Now construct $U^{(2)}$ with elements of U that have partial membership in exactly two fuzzy partition sets. Formally, $U^{(2)}$ is the set of employees of U represented on the fuzzy partition by vectors with exactly two non-zero coordinates. So, $U^{(2)}$ is the set of employees with two jobs overlapping. This set is geometrically represented by the edges of S_K.

Following this procedure, $U^{(k)}$, $2 \le k \le K$, should be the set of employees with k jobs overlapping. Of course, empty sets, if any, should be removed. For the sake of clarity, assume no set is empty.

Clearly, $U^{(k-1)}$ involves less overlapping than does $U^{(k)}$. Denote this assertion symbolically $U^{(k-1)} \prec U^{(k)}$. Thus, the finite set $\{U^{(1)}, U^{(2)}, \ldots, U^{(K)}\}$ of subsets of U forms an ordered crisp partition of U with respect to $\prec$ symbol.

This subsection can be summarized as follows: U is decomposed beforehand into a fuzzy K-partition to account for individual heterogeneity. The result of this decomposition is the mapping

$$i \to \boldsymbol{g}_i = (g_{i1}, g_{i2}, ..., g_{iK}),\ 1 \le i \le N .$$

Individuals whose vector $\boldsymbol{g}_i$ has only one non-null coordinate are given label 1; the ones whose vector $\boldsymbol{g}_i$ has exactly two non-null coordinates are given label 2; and so forth up to label K. The set $U^{(k)}, 1 \le k \le K$, clusters all individuals with label k. The superscript index k identifies the number of jobs overlapping. The set U is then partioned into K subsets $U^{(1)}$, $U^{(2)}$, …, $U^{(K)}$ verifying the order relation $U^{(1)} \prec U^{(2)} \prec ... \prec U^{(K)}$ and, of course,

$$U = \bigcup_{k=1}^{K} U^{(k)} .$$

The distribution of overlapping can be calculated through the formula

$$\frac{\# U^{(k)}}{\# U},\ 1 \le k \le K.$$

5. Statistical Model

5.1. A brief introduction

In this section we present a latent structure grade of membership (GoM) model introduced by Woodbury and Clive (1974) as a means to analyse multivariate categorical data, particularly in medical context. It is based on fuzzy partitions, and allows the estimation of *K* fuzzy sets or classes, $A_1, A_2, ..., A_K$ and, for each individual his/her grades of membership in those sets. Thus, GoM analysis deals "simultaneously with the dual problem of case clustering and estimation of discriminant coefficients" (Woodbury and Manton, 1982). The number of fuzzy partition sets or classes *K* is a user-defined parameter i.e. *K* is fixed a priori.

To express mathematically the dual nature of the analysis, the model uses two sets of parameters. One set describes how variables are associated with the classes identified by the model and a "second set how individual's observed characteristics relate to each fuzzy class" as Manton et al. (1992) point out. The sets generic elements are referred to as λ_{kjl} and g_{ik}, respectively. The parameters λ_{kjl} are used to account for A_k fuzzy class features and g_{ik} are used to measure the degree of compatibility of individual *i* with this class. By assumption, both sets are latent to observations. In this way, GoM analysis aims to derive collective and individual properties, in form of latent variables, using multivariate measurements (Kovtun et al., 2004).

5.2. Formulation

Let $\boldsymbol{X} = (X_1, X_2, ..., X_J)$ be a random vector whose components are discrete categorical variables with the set of outcomes of j^{th} variable being $\mathcal{L}_j = \{1,2,...,L_j\}, 1 \leq j \leq J$. The L_j categories are assumed exhaustive and mutually exclusive. Continuous variables, if any, should be approximated by categorical variables with comparable distribution. The key problem faced in GoM analysis is the representation of the distribution of $\boldsymbol{X}$, denoted by $\Pr[(X_1, X_2, ..., X_J) = (x_1, x_2, ..., x_J)]$, in a individualized heterogeneity context.

Consider *N* independent realization of the random vector $\boldsymbol{X}$, i.e., a random sample $\boldsymbol{X}_1$, $\boldsymbol{X}_2$,…, $\boldsymbol{X}_N$, where $\boldsymbol{X}_i = (X_{i1}, X_{i2}, ..., X_{iJ})$, $1 \leq i \leq N$. This variable is referred to as measurement vector of individual *i* and its generic coordinate X_{ij} to as outcome of individual *i* in measurement *j*. Assume that $\boldsymbol{X}_i$ carries implicitly information about the position of individual *i* in a fuzzy partition in form of a latent random variable $\boldsymbol{g}_i = (g_{i1}, g_{i2}, ..., g_{iK}) \in S_K$, $1 \leq i \leq N$, where *K* is the number of partition sets. As in (2), S_K is the unit simplex with *K* extreme points.

Denote the conditional probability of the coordinate X_{ij} taking the value *l*, given that *i* is a crisp element of *k*-fuzzy set, as

$$\lambda_{kjl} = \Pr\left[X_{ij} = l \mid g_{ik} = 1\right],\ 1 \le k \le K;\ 1 \le i \le N;\ 1 \le j \le J;\ l \in \mathcal{L}_j$$

In other words, λ_{kjl} represents the probability of fuzzy partition set A_k crisp elements having outcome l in measurement j. Being probabilities, λ_{kjl} verify the following constraints: $0 \le \lambda_{kjl} \le 1$ and $\sum_{l=1}^{L_j} \lambda_{kjl} = 1$ for each j and k. The basic assumption of GoM model is that, given $\boldsymbol{g}_i$, the probability of individual i having outcome l in measurement j is given by

$$p_{ijl} = \Pr\left[X_{ij} = l \mid \boldsymbol{g}_i\right] = \sum_{k=1}^{K} g_{ik}\lambda_{kjl},\ 1 \le k \le K;\ 1 \le i \le N;\ 1 \le j \le J; l \in \mathcal{L}_j\ . \tag{3}$$

In this model, the fuzzy partition that, by assumption, structures the underlying population is identified by crisp elements probabilities vector

$$\Lambda_k = \left(\left(\lambda_{kjl}, l \in \mathcal{L}_j\right), 1 \le j \le J\right),\ 1 \le k \le K\ .$$

5.3. Estimation

Likewise other latent variables analysis, local or conditional independence is a central assumption of GoM model statistical application. Formally, conditional on value of $\boldsymbol{g}_i = \left(g_{i1}, g_{i2}, ..., g_{iK}\right)$, the random variables $X_{i1}, X_{i2}, ..., X_{iJ}$ are assumed mutually independent. This assumption leads to the factorization of the joint probability distribution

$$\Pr\left[\left(X_{i1}, X_{i2}, ..., X_{iJ}\right) = \left(x_{i1}, x_{i2}, ..., x_{iJ}\right) \mid \boldsymbol{g}_i\right] = \prod_{j=1}^{J} \Pr\left[X_{ij} = x_{ij} \mid \boldsymbol{g}_i\right]$$

In presence of N independent observations, the conditional to latent variable $\boldsymbol{g}_i = \left(g_{i1}, g_{i2}, ..., g_{iK}\right)$ likelihood can be written in multinomial form (Manton et al., 1994),

$$\boldsymbol{L}_K = \prod_{i=1}^{N}\prod_{j=1}^{J}\prod_{l=1}^{L_j}\left(\sum_{k=1}^{K} g_{ik}\lambda_{kjl}\right)^{y_{ijl}} \tag{4}$$

where

$$y_{ijl} = \begin{cases} 1 & \text{if } x_{ij} = l \\ 0 & \text{otherwise} \end{cases}$$

Note that when writing conditional likelihood (4), one implicitly assumes that number of fuzzy classes K is fixed a priori. Prior knowledge about the population under study makes the choice of K easier and often leads to a richer analysis.

Estimation strategy of GoM model parameters is based on maximum likelihood method. To obtain estimates of model parameters, an alternate optimization scheme is applied. The likelihood (4) is then optimized iteratively by updating one set of parameters (e.g. λ_{kjl}) while holding the other set (e.g. g_{ik}) fixed and vice-versa. As usual, the process ends as long as convergence criteria are met. Details about numerical implementation of this procedure can be found in Woodbury et al. (1978).

The measurement of model fit is a critical issue in GoM context due to boundary parameters. Often goodness-of-fit is assessed by a likelihood ratio test, comparing the likelihood associated with K fuzzy classes $\boldsymbol{L}_K$, as in (4), to so-called independence model likelihood $\boldsymbol{L}_0$ (Berkman et al., 1989), associated with observed marginal relative frequencies

$$\hat{\pi}_{jl} = \frac{\#\{X_i : X_{ij} = l, 1 \le i \le N\}}{N}, 1 \le j \le J; l \in \mathcal{L}_j .$$

Under null hypothesis, i.e. when $\boldsymbol{L}_0$ fits, the statistic $\chi^2 = -2\ln\left(\boldsymbol{L}_K \big/ \boldsymbol{L}_0\right)$ is approximated to a qui-square distribution with the number of degrees of freedom ν equals the number of estimated parameters (Manton et al., 1994). Due to excessive number of parameters involved, approximation of χ^2 to Normal distribution, provided by the statistic $Z_F = \sqrt{2\chi^2} - \sqrt{2\nu - 1}$, is used alternatively. The empirical value $z_F > z_{(1-\alpha)}$ leads to the rejection of the null hypothesis, where $z_{(1-\alpha)}$ is the (1-α)×100% quantile of standard Normal distribution.

5.4. Typologies construction

The likelihood ratio test referred above gives an overall scalar measure for model acceptance. Having obtained a good fit with $\boldsymbol{L}_K$, the main concern is the characterization of fuzzy classes. In this regard, Marini et al. (1996), for example, suggest $(1+\delta), \delta \in [0,1]$, criterion in that the variable-category pair (j,l) is said to contribute substantively to characterize the fuzzy class A_k if

$$\hat{\lambda}_{kjl} > (1+\delta) \times \hat{\pi}_{jl}, \tag{5}$$

where $\hat{\pi}_{jl}$ is the corresponding sample marginal frequency. The value of δ is set by the user. The referred authors use δ=0.40 in their empirical work excepting the categories "with very high marginal frequencies". In our empirical analysis we subjectively set δ=0.30.

6. Empirical Analysis

6.1. The dataset

The data for this research come from a Portuguese administrative linked employer employee data (LEED) – *Quadros de Pessoal*, and refer to the year of 2006. The data offers longitudinal information matching firms and employees, which cover the population of firms with wage earners of private sectors of Portuguese economy (see Cardoso and Portugal, 2005; Mamede, 2006 for details). The information contained therein includes:

i) From employer side: location, industry, employment, sales, firm tenure, capital stock, number of establishments, and legal and institutional settings;
ii) From workers side: age, gender, education, qualification level, occupation, tenure, wage floor, additional monetary payments, type of labour contracts and working hours.

While employers and employees characteristics are largely described in the dataset, job attributes are generally absent. The challenge of studying the match between jobs and employees is then to ascertain job attributes proxies. The prior task of our study is to identify job attributes from LEED which is mainly focused on employer-employee data. These attributes will be used to identify the fuzzy *K*-partition. By assumption, in this study we identenfy job characteristics by estimated prevalent workers characteristics within jobs using the criterion (5), as referred in Subsection 5.4.

In this experimental study we limit our analysis to a sample of employers and employees. It focuses on a sample of banks and takes advantage of our previous work on skill matching (Sgobbi and Suleman, 2008). Two largest banks constitute the sample and hereafter they will be referred to as bank A and bank B, for confidentiality reasons. The sample sizes are 4,518 for the bank A and 8,286 for the bank B.

6.2. Job attributes variables construction

In LEED style data, job attributes are generally not described besides the type of industry, location and plant size. As referred earlier, this problem is tackled by assuming workers prevalent characteristics within jobs as job attributes and demand-side requirements. To build up job-based clusters, we assume that inside firm each job has several wage and non-wage attributes; and requires a specific type and degree of skills. At least, employers screen employees according to those requirements both at entry level and for career progression.

Proxy variables considered in this analysis focus both on attributes and required skills for core jobs inside the banking industry:

- Proxy of tasks assigned to jobs: core jobs were defined by crossing the occupations from Portuguese dictionary of occupation titles and jobs defined by collective agreement which codes are included in the dataset. Four core jobs are considered in this paper: branch manager, assistant branch manager, retail bankers and bank tellers.
- Proxy of observable required skills: from the dataset three variables allows defining type and degree of skills: education, tenure and qualification levels. This means, we are using an indirect measure of demanded skills which is identified from most prevalent provided skills by employees. At least, we focus on source of skills as measure of skills (Suleman, 2007).
    - o Education variable includes both the schooling and scientific field. The common code refers to: less than basic education; general or vocational basic education;

general or vocational secondary education; bachelor in human sciences; social science; management, technologies and others; and higher education follows the same categories of bachelor. Sixteen levels discriminates the education variable;

 - o Tenure represents the demand for specific skills by banking jobs and was coded in 8 levels;
 - o Qualification levels which define job ranges based on skills and tasks. Three from 8 levels prevail in the banking employees: intermediate executives, highly skilled and skilled (see, Martins 2004 for detailed information).

- Proxy of unobservable skills: as known, some skills are attributed to young people as well as to women or men. So we used 9 levels age-groups and gender as proxy of some of non-observable required skills;
- Proxy of non-wage attributes: working time, labour contract and promotions define the other features of jobs. The dataset discriminates standard (147 hours per month) and over working time. Total working time includes both standard and overtime which are coded in three levels according to the normal time in less than, equal and higher than 147. The labour contracts are discriminated in stable and flexible patterns. The career is analysed from the latest promotion.

Finally, wage attributes are analysed from contractual and total earnings. In both cases, hourly wage was calculated and 10 levels wage structure is used to compare whether received earnings are close to collective rules defining wage by job.

6.3. GoM model application output

As referred above, to apply GoM model to dataset of Portuguese 2006 LEED, the number of fuzzy partition sets K is fixed a priori. Often, this model variable is set according to some prior knowledge of the universe under study. As we target on job characterization the value of K was set to 4, in accordance with equal number of job categories: branch manager, assistant branch manager, retail bankers and bank tellers. An empirical measure of model goodness-of-fit leads to the value $z_F = 120.9$ for the bank A sample and $z_F = 151.9$ for the bank B sample. Both values are much higher than 1.645, the 95% quartile of the standard Normal distribution. Thus, the fuzzy sets theory based model, with K=4 fuzzy sets, seems to be congruent with both dataset.

Now we are concerned with the characterization of each fuzzy partition. As noted earlier, the estimates of prototypes response probabilities λ_{kjl} allow the configuration of fuzzy K-partition while the estimates of individual grades of membership are used to crisply decompose datasets into ordered by overlapping partitions.

6.4. Job typologies characterization

Tables A1 and A2, in Appendix A, contain pertinent data summaries of job characterization in bank A and bank B, respectively. In both tables, all values are in x100% format.

Marginal frequencies (column 2) are presented together with maximum likelihood estimates $\hat{\lambda}_{\mathrm{kjl}}$ of prototypes conditions (columns 3 to 6). Sample frequencies in different categories may be used to infer about population characteristics as a whole, whilst the estimates $\hat{\lambda}_{\mathrm{kjl}}$ help to understand the prevalence of each outcome for job characterization purpose. Distinguished features are in boldface figures and, excepting high observed frequencies, they are selected according to the criterion defined in (5), setting $\delta = 0.30$ as already referred. High observed frequencies are tackled by additionally taking the outcome *l* of variable *j* as a distinguish feature of job typologies whenever the conditions $\hat{\lambda}_{kjl} > \hat{\pi}_{jl} \wedge \hat{\lambda}_{kjl} > 0.90$ hold.

Table 1: Summary of estimated job typologies in bank A.

Fuzzy Partition Set →	A_1	A_2	A_3	A_4
Job Category	Branch manager or Assistant branch manager	Bank teller	Retail banker	Bank teller
Gender	Male	Male	Female	Female
School	Basic	Secondary or lower	Higher educ. management	Higher education
Tenure	10-24	5-14	5-9	≤ 4
Age	≥ 41	≥ 36	31-35	≤ 30
Qualification	Intermediate executive	Skilled employee	Higher skilled employee	Skilled employee
Total working time	Standard	Standard	Standard	> Standard
Type of contract	Open-ended	Open-ended	Open-ended	Fixed-term
Basic hour wage (€)	≥ 10.0	7.5 - 10.0	5.0 - 7.5	≤ 5.0
Total hour wage (€)	≥ 12.5	10.0 – 15.0	10.0 – 12.5	≤ 7.5
Last Promotion	2003	On or before 2002	2003-2004	2005-2006

Note: The entries in this table are based on GoM model output (see Table A1, in Appendix A)

Tables 1 and 2 summarize the prevalent conditions of banking employees in each of the four job typologies estimated using GoM model. The results indicate that the job attributes and requirements are quite similar in the two banks. A look at the estimated typologies shows an alignment with the main features of employment in the banking industry as described by Lima et al. (2008). The banking industry is becoming increasingly dominated by female employees, young and most educated workers.

The estimated job clusters reveal that low position jobs, as the bank tellers, are the most heterogeneous. Indeed, GoM model leads to two typologies of bank tellers with deeply differentiate attributes thus highlighting the prevalence of within job overlapping. While one subgroup of bank tellers is characterized as mid-age male with secondary or lower education, relatively high tenure, open-ended contract, and not recently promoted, the other follows the pattern described as new dynamics of the banking labour market. Among the new employees, the proportion of females, youngsters and individuals with high formal education is higher than in the previous.

At the ports of entry jobs, the findings suggest competition between graduates holding different field of education. This competition occurs essentially between graduates of human science and social science (bank A) and also management (bank B). The banks of the sample seem to take advantage of the investments in human capital made by young job searchers revealing the preference of employers on higher rather than lower education employees.

Table 2: Summary of estimated job typologies in bank B.

Fuzzy Partition Set →	A_1	A_2	A_3	A_4
Job Category	Branch manager	Assistant branch manager or retail banker	Bank teller	Bank teller
Gender	Male	Male	Female	(undistinguished feature)
School	Basic or lower	Secondary or higher	Higher education	General secondary
Tenure	≥ 25	10-24	< 10	10-24
Age	≥ 51	36-50	≤ 35	36-50
Qualification	Intermediate executive	Higher skilled employee	Skilled employee	Skilled employee
Total working time	Standard	Standard	> Standard	> Standard
Type of contract	Open-ended	Open-ended	Fixed term	Open-ended
Basic hour wage (€)	12.5 - 17.5	10.0 - 12.5	5.0 - 7.5	7.5 - 10.0
Total hour wage (€)	≥ 17.5	15.0 – 20.0	5.0 - 10.0	12.5 - 15.0
Last Promotion	Before 2001or 2002	2003 or 2005	2004-2006	2001

Note: The entries in this table are based on GoM model output (see Table A2, in Appendix A).

Furthermore, the prevalence of senior employees in teller-level jobs, as in the one of the typologies, indicates that they probably were not selected for higher position jobs. The conclusion arising from these findings is that the competition at lower levels jobs evolves around candidates with general skills as well as with specific skills. Moreover, younger employees' flexible labour contracts and "fast track" promotions seem to be two instruments used by banks to increase the efficiency in the matching process.

At the higher position jobs, the estimated typologies allow additional picture of the jobs attributes from the stock of banking employees. Branch manager and assistant branch manager jobs are mainly occupied by senior males with low education and high tenure. Moreover, the prevalence such attributes offers signs of hierarchical model within the banks of the sample (see Suleman and Sgobbi, paper).

In summing up, within jobs overlapping found at the entry positions offers a picture of differentiated job attributes and skill requirements which lead to two typologies of bank tellers. Next we examine the between jobs overlapping as to ascertain the extent in which banking employees share the attributes and/or skill requirements of two or more job estimated typologies, the job tellers ones inclusive.

6.5. Analysis of between jobs overlapping

Table 3 displays the estimated between job overlapping in bank A and bank B. Substantial amount of overlapping occurs in the set of employees that share exactly two job typologies, i.e. $U^{(2)}$, roughly by the same amount in either bank. A non negligible overlapping also occurs in $U^{(3)}$ although the same can not be said about highest heterogeneous employees set $U^{(4)}$.

Table 3: Between jobs overlapping distribution.

Bank ↓	U	$U^{(1)}$	$U^{(2)}$	$U^{(3)}$	$U^{(4)}$
A	4,518	1,990 (44.0%)	1,877 (41.5%)	646 (14.3%)	5 (0.1%)
B	8,286	3,345 (40.4%)	3,631 (43.8%)	1,229 (14.8%)	81 (1.0%)

The distribution of banking employees in the respective fuzzy partitions is given graphically in Figure 1, by picturing on unit simplex S4 the estimated grade of membership vectors. As noted earlier, individuals with no job overlapping position on the vertices while the ones having jobs overlapping position on the edges ($U^{(2)}$), faces ($U^{(3)}$) or in the interior ($U^{(4)}$) of S4.

This representation affords a different insight about employees' distribution in the labour market. While bank B edges are visibly populated in bank A the edge connecting two extreme positions jobs, that is top managers and recent tellers, is almost empty. Furthermore, scarse density is observed on the edge connecting two teller typologies. These outcomes appeal for a wider analysis of the labour market particularly concerning the sources of between jobs overlapping to better understand its contemporary features.

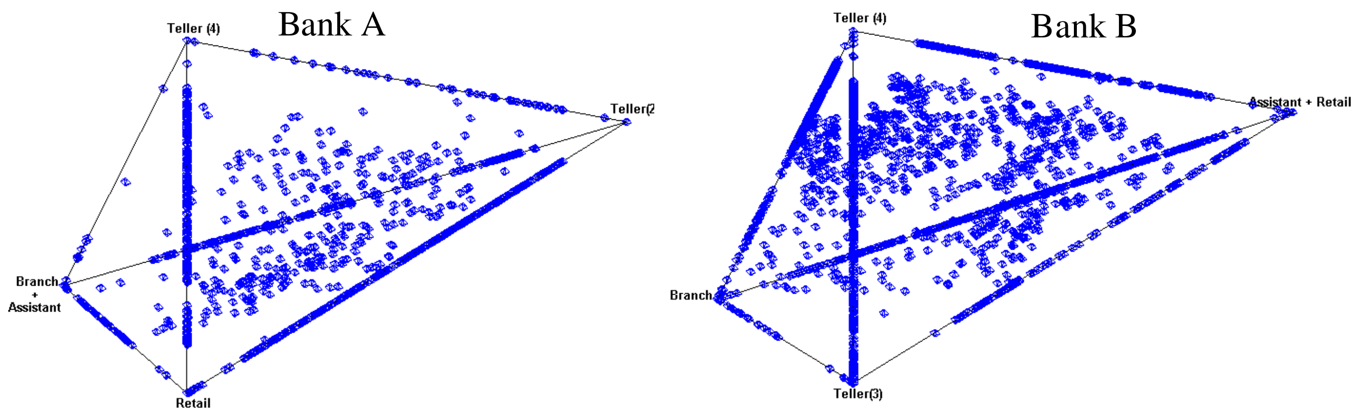

Figure 1: Representation of banks A and B estimated fuzzy partitions through unit simplex S4.

7. Concluding remarks

Workers assigned to a specific type of jobs are far from having the same characteristics. Therefore, overlapping labour markets are a more realistic representation of the labour market.

This study explores the concept of overlapping labour market as stated in Sattinger (2006) and goes further by introducing a methodology to measure the degree of overlapping in a particular labour market. It is based on fuzzy sets theory and assumes that the labour market can be decomposed into a fuzzy partition.

Job typologies as well as the grades of membership to these typologies of workers holding different jobs were estimated under GoM statistical model (Woodbury and Clive, 1974). The typological features are based upon subjective criteia defining prevalent job conditions for the estimated probabilities $\hat{\lambda}_{kjl}$. The question about these criteria or whether those conditions represent each job requirements is open for debate.

Using Portuguese administrative linked employer employee data (LEED) for the stock of employees in the two largest banks in 2006, the empirical analysis suggests that overlapping is significant in each firm considered in the sample. Empirical findings lead to differentiate both within and between jobs overlapping. The estimated typologies indicates essentially two generations of low positions jobs, one clustering senior and less educated employees with favourable labour contracts, and the other filled by higher educated youngsters, and less attractive contracts.

Focusing in the between jobs overlapping, we may suggest that the degree of jobs overlapping does not substantively vary across banks. Sligthly more than 40% of employees share characteristics of two job typologies (set $U^{(2)}$), about 14% share characteristics of three job typologies (set $U^{(3)}$), and a negligible number shares characteristics of four typologies, though it is much higher in bank B (Table 3).

Given the size of the estimated degree of between jobs overlapping, further research is needed to ascertain explanations for this evidence. The next issue of our study concerns the sources of between jobs overlapping to understand the features of the contemporary labour market. The current debate suggests three possible sources: the job search and hiring costs for workers and employees, respectively; skill composition of the workforce, so overlapping results from supply-side characteristics; or overlapping is caused by the rules governing employment relationship being clearly generated by demand-side rules and procedures.

Acknowledgment

The authors thank the Ministry of Labour and Social Solidarity (MTSS) for providing the Portuguese LEED "Quadros de Pessoal". They are also most indebted to MTSS staff for their prompt assistance on additional information about the dataset.

References

Baker, G., Gibbs, M., Holmstrom, B. (1994). The Internal Economics of the Firm Level: Evidence from Personnel Data, *Quarterly Journal of Economics*, November, pp. 882-919.

Bellman, R.E. and Zadeh, L.A. (1970). Decision-making in fuzzy environment, *Management Science*, Vol. 17(4), B141-B164.

Berkman, L., Singer, B. and Manton, K. (1989). Black/White Differences in Health Status and Mortality among Eldery, *Demography*, Vol. 26(4), pp. 661-678.

Borghans, L. and Grip, A. (1999), Skills and low pay: upgrading or overeducation?, ROA-RM-1999/5E

Cardoso, A.R. and Portugal, P. (2005). Contractual Wages and the Wage Cushion under Different Bargaining Settings, *Journal of Labor Economics*, Vol. 23(4), pp. 875-901.

Diamond, P. (1982), A Model of Price Adjustment, *Journal of Economic Theory*, 3, pp. 156-168.

Döring, C, Lesot, M-J and Kruse, R. (2006). Data analysis with fuzzy clustering methods, *Computation Statistics and Data Analysis*, Vol. 51, pp. 192-214.

Grund, C. (2002). The Wage Policy of Firms – Comparative Evidence for the U.S. and Germany from Personnel Data, IZA Discussion Paper nº 605.

Kovtun, M., Akushevich, I., Manton, K.G. and Tolley, H.D. (2004). Grade of Membership Analysis: One Possible Approach to Foundations. Available in http://arXiv:math/0403373v1.

Mamede, R. (2006). Labour Mobility and Firm Survival, paper presented at Workshop "Quadros de Pessoal e Investigação em Economia", University of Minho, Portugal, 15 September 2006.

Lima, M. P., Guerreira, A., Kolarova, M., and Nunes, C. (2008). Globalização e Relações Laborais: Análise dos Sectores Textil, Automóvel, Bancário, Telecomunicações e Hotelaria e Restauração, 6º *Congresso Português de Sociologia*, Universidade Nova de Lisboa, 25-28 de Junho de 2008.

Manton, K.G., Woodbury, M.A. and Tolley, H.D. (1994). *Statistical Applications Using Fuzzy Sets*, John Wiley & Sons, Inc.

Manton, K.G., Woodbury, M.A., Stallard, E. and Corder, L.S. (1992). The Use of Grade of Membership Techniques to estimate Regression Relationships, in P. Marsden (ed.), *Sociological Methodology*, pp. 321-281, Basil Blackwell, Oxford, England.

Marini, M.M., Li, X. and Fan, P.L. (1996). Characterizing Latent Structures: Factor Analytic and Grade of Membership Models, *Sociological Methodology*, Vol. 26, pp. 133-164.

Moen, E. (1997), Competitive Search Equilibrium, *Journal of Political Economy*, Vol. 105(2), pp. 385-411..

Mortensen (1982), Property Rights and Efficiency in Mating, Racing and Related Games, *American Economic Review,* Vol. 72, pp. 968-979.

Mortensen, D. and Wright, R. (2002), Competitive Pricing and Efficiency in a Search Equilibrium, *International Economic Review*, Vol. 43(1), pp. 1-20.

Nguyen, H.T, Prasad, N.R., Walker, C.L. and Walker, E.A. (2003). *A First Course in Fuzzy and Neural Control*, Chapman & Hall/CRC.

Nguyen, H.T. and Walker, E.A. (2000). *A First Course in Fuzzy Logic*, (2nd Edition), Chapman & Hall/CRC.

Pissarides, C. (1984), Efficient Job Rejection, *Economic Journal*, Vol. 94, pp. 97-108.

Pissarides, C. (1990), Equilibrium Unemployment Theory, Basil Blackwell, Cambridge, MA.

Sattinger, M. (1975), Comparative Advantage and the Distribution of Earnings and Abilities, *Econometrica*, Vol. 43, pp. 455-468.

Sattinger, M. (1995), Search and the Efficient Assignment of Workers to Jobs, *International Economic Review*, Vol. 36, pp. 283-302.

Sattinger, M. (2006), Overlapping Labour Markets, *Labour Economics*, Vol. 13, pp. 237-257.

Schumann, P, Ahlberg, D. and Mahoney, C. (1994). The Effects of Human Capital and Job Characteristics on Pay, *Journal of Human Resources*, Vol. 29(2), pp. 481-503.

Seltzer A. and Merrett, D. (2000). Personnel Policies at the Union Bank of Australia: Evidence from 1888-1900 Entry Cohorts, *Journal of Labor Economics*, Vol. 18 (4), 573-613.

Sgobbi, F. and Suleman, F. (2008). A Methodological Contribution to the Measurement of Skill (Mis)match, Paper presented at International Workshop– Performance, Skills and Competences in the 21st Century, Lisbon, ISCTE, 11-12 December 2008.

Shi, S. (2001), Frictional Assignments, Part I: Efficiency, *Journal of Economic Theory*, Vol. 98, pp. 232-260.

Shi, S. (2002), A Directed Search Model of Inequality with Heterogeneous Skills and Skill-Based Technology, *Review of Economic Studies*, Vol. 69, pp. 467-491.

Shimer, R. and Smith, L. (1998), Assortative Matching and Search, manuscript, Princeton University.

Singer, B. (1989). Grade of Membership Representations: Concept and Problems, in T.W. Anderson, K.B. Athreya and D. Iglechardt (eds), *Probability, Statistics and Mathematics: Essays in Honor of Samuel Karlin*, pp. 317-334, Academic Press, San Diego.

Smeets, V. and Warzynski, F. (2008). Too Many Theories, Too Few Facts? What the Data tell us about the Link between Span of Control, Compensation and Careers Dynamics, *Labour Economics*, Vol. 15, pp. 688-704.

Suleman, F. (2007). *O Valor das Competências no Mercado de Trabalho*, Lisbon, Livros Horizonte.

Teulings, C. (1995), The Wages Distribution in a Model of the Assignment of Skills to Jobs, *Journal of Political Economy*, Vol. 103, pp. 280-315.

Tinbergen, J. (1951), Some Remarks on the Distribution of Labour Outcome, *International Economic Papers*, Vol. 1, pp. 195-207.

Tinbergen, J. (1956), Competition for Jobs in a Growing Economy and the Emergence of Dualism, *The Economic Journal*, Vol. 109, pp. 349-371.

Treble, J., Gameren, E., Bridges, S. and Barmby, T. (2001). The Internal Economics of the Firm: Further Evidence From Personnel Data, *Labour Economics*, Vol. 8, pp. 531-552.

Van der Meer, P. H. (2006), The Validity of two Education Requirement Measures, *Economics of Education Review*, Vol. 25 (2), pp. 211-219.

Van Ours, J.C. and G. Ridder (1995). Job Matching and Job Competition: Are Lower Educated Workers at the Back of Job Queues?, *European Economic Review*, Vol. 55, pp. 125-154.

Wieling, M. and Borghans, L. (1995). Discrepancies between Demand and Supply and the Adjustment Processes on the Labour Market, ROA, Research Memorandum.

Woodbury, M.A. and Clive, J. (1974). Clinical Pure Types as a Fuzzy Partition', *Journal of Cybernetics*, Vol. 4, pp. 111-121.

Woodbury, M.A. and Manton, K.G. (1982). A new Procedure for Analysis of Medical Classification, *Methods of Information in Medicine*, Vol. 21, pp. 210-220.

Woodbury, M.A., Clive, J. and Garson Jr., A. (1978). Mathematical Typology: a Grade of Membership Technique for obtaining Disease Definition, *Computer and Biomedical Research*, Vol. 11, pp. 277-298.

Woodbury, M.A., Manton, K.G. and Tolley, H.D. (1997). Convex Models of High Dimensional Discrete Data, *Annals of the Institute of Statistical Mathematics*, Vol. 49, pp. 371-393.
Zadeh, L. (1965). Fuzzy Sets, *Information and Control*, Vol. 8, pp. 338-353.

APPENDIX A

Table A1: Estimates of crisp elements outcome probabilities $\hat{\lambda}_{kjl}$ (bank A)

Variable	$\hat{\pi}_{jl}$	$\hat{\lambda}_{1jl}$	$\hat{\lambda}_{2jl}$	$\hat{\lambda}_{3jl}$	$\hat{\lambda}_{4jl}$
Gender					
Male	53.19	**81.00**	**70.39**	35.12	36.45
Female	46.81	19.00	29.61	**64.88**	**63.55**
Job category					
Branch manager	9.36	**54.79**	0.00	0.00	0.00
Assistant branch managers	7.72	**45.21**	0.00	0.00	0.00
Retail banker	17.33	0.00	0.00	**100.00**	0.00
Bank teller	65.58	0.00	**100.00**	0.00	**100.00**
School					
< Basic	1.68	0.00	**5.82**	0.00	0.00
General basic	6.42	**13.05**	**14.70**	0.00	0.00
Vocational basic	1.73	**4.22**	**3.55**	0.00	0.00
General secondary	51.02	56.98	**75.42**	37.41	35.24
Vocational secondary	0.64	0.39	0.00	0.00	**1.99**
Bachelor – human sciences	0.22	0.00	0.00	0.00	**0.76**
Bachelor – social sciences	0.24	0.00	**0.51**	0.00	**0.33**
Bachelor – management	2.74	0.00	0.00	**6.91**	3.39
Bachelor – technologies	0.27	0.00	0.00	**0.61**	**0.38**
Bachelor – others	0.42	0.00	0.00	**1.65**	0.00
Higher education – human sciences	5.31	3.11	0.00	5.13	**12.02**
Higher education – social sciences	9.65	7.07	0.00	11.76	**18.89**
Higher education – management	17.35	14.38	0.00	**35.10**	20.76
Higher education – technologies	1.64	0.80	0.00	0.94	**4.36**
Higher education – others	0.66	0.00	0.00	0.49	**1.86**
Age group					
≤ 25	7.06	0.00	0.00	0.00	**25.50**
26—30	20.63	0.00	0.00	0.00	**74.50**
31—35	28.20	0.00	0.00	**92.66**	0.00
36—40	15.23	19.34	**39.28**	7.34	0.00
41—45	10.18	**30.65**	**19.84**	0.00	0.00
46—50	12.37	**34.67**	**25.92**	0.00	0.00
51—55	4.94	**11.57**	**11.94**	0.00	0.00
56—60	1.28	**3.39**	**2.83**	0.00	0.00
≥ 61	0.11	**0.37**	**0.19**	0.00	0.00
Tenure					
< 01	9.16	0.00	0.00	0.00	**29.38**
01—04	22.02	0.00	0.00	0.00	**70.62**
05—09	53.47	47.19	**81.11**	**100.00**	0.00
10—14	6.73	**8.95**	**18.89**	0.00	0.00
15—19	8.30	**42.28**	0.00	0.00	0.00
20—24	0.31	**1.58**	0.00	0.00	0.00
25—29	9.16	0.00	0.00	0.00	**29.38**
≥ 30	22.02	0.00	0.00	0.00	**70.62**
Qualification					
Intermediate executive	17.09	**100.00**	0.00	0.00	0.00
Higher skilled employee	17.33	0.00	0.00	**100.00**	0.00
Skilled employee	65.58	0.00	**100.00**	0.00	**100.00**
Total working time (month)					
< 147 hours	0.38	0.34	**0.89**	0.00	0.26
= 147 hours	96.75	**98.11**	**97.48**	**98.53**	93.61
> 147 hours	2.87	1.55	1.63	1.47	**6.13**

Type of contract					
Open-ended contract	91.12	**100.00**	**100.00**	**100.00**	65.35
Fixed-term contract	8.88	0.00	0.00	0.00	**34.65**
Basic hourly wage (€)					
< 5.	6.19	0.00	0.00	0.00	**22.88**
5.0 —7.5	66.38	0.00	58.29	**100.00**	77.04
7.5 —10.0	12.63	0.00	**41.71**	0.00	0.00
10.0 —12.5	7.11	**48.11**	0.00	0.00	0.00
12.5 —15.0	6.71	**45.40**	0.00	0.00	0.00
15.0 —17.5	0.89	**6.03**	0.00	0.00	0.00
20.0 —22.5	0.07	**0.45**	0.00	0.00	0.00
≥ 25.0	0.02	0.00	0.00	0.00	**0.08**
Total hourly wage (€)					
≤ 5.0	0.31	0.00	0.00	0.00	**1.59**
5.0 —7.5	19.30	0.00	0.00	0.00	**98.41**
7.5 —10.0	34.65	0.00	0.00	**87.66**	0.00
10.0 —12.5	20.88	0.00	**70.53**	12.34	0.00
12.5 —15.0	10.47	**20.82**	**29.47**	0.00	0.00
15.0 —17.5	7.40	**40.69**	0.00	0.00	0.00
17.5 —20.0	4.46	**24.51**	0.00	0.00	0.00
20.0 —22.5	1.94	**10.66**	0.00	0.00	0.00
22.5 —25.0	0.45	**2.45**	0.00	0.00	0.00
≥ 25.0	0.16	**0.86**	0.00	0.00	0.00
Last promotion					
2006	30.74	23.88	0.00	31.55	**59.12**
2005	26.36	23.23	15.09	28.71	**35.35**
2004	12.84	16.40	8.90	**22.89**	5.53
2003	9.36	**13.43**	10.24	**16.85**	0.00
2002	5.00	4.79	**16.43**	0.00	0.00
2001	4.58	5.69	**14.16**	0.00	0.00
< 2001	11.11	12.58	**35.18**	0.00	0.00

Table A2: Estimates of crisp elements outcome probabilities $\hat{\lambda}_{kjl}$ (bank B)

Variable	$\hat{\pi}_{jl}$	$\hat{\lambda}_{1jl}$	$\hat{\lambda}_{2jl}$	$\hat{\lambda}_{3jl}$	$\hat{\lambda}_{4jl}$
Gender					
Male	47.92	**87.17**	**65.60**	21.30	44.98
Female	52.08	12.83	34.40	**78.70**	55.02
Job category					
Branch manager	7.00	**100.00**	0.00	0.00	0.00
Assistant branch managers	8.88	0.00	**48.84**	0.00	0.00
Retail banker	9.30	0.00	**51.16**	0.00	0.00
Bank teller	74.81	0.00	0.00	**100.00**	**100.00**
School					
< Basic	5.44	**31.17**	0.00	0.00	0.00
General basic	16.44	16.28	13.14	0.00	**27.42**
Vocational basic	4.22	**24.19**	0.00	0.00	0.00
General secondary	41.89	14.31	**67.84**	11.72	**65.50**
Vocational secondary	3.43	**8.17**	1.79	1.18	3.34
Bachelor – human sciences	1.04	0.53	**2.12**	0.00	**1.63**
Bachelor – social sciences	0.54	0.51	**2.66**	0.67	0.00
Bachelor – management	4.84	4.54	**6.60**	**12.21**	0.00
Bachelor – technologies	0.52	0.29	**0.87**	0.47	0.56
Bachelor – others	0.07	0.00	**0.72**	0.00	0.00
Higher education – human	2.18	0.00	**2.97**	**4.32**	1.54
Higher education – social sciences	6.01	0.00	0.00	**21.69**	0.00
Higher education – management	12.50	0.00	0.00	**45.12**	0.00
Higher education – technologies	0.63	0.00	**1.29**	**1.80**	0.00
Higher education – others	0.23	0.00	0.00	**0.83**	0.00
Age group					
≤ 25	1.77	0.00	0.00	**6.54**	0.00
26—30	10.68	0.00	0.00	**39.40**	0.00
31—35	14.65	0.00	0.00	**54.05**	0.00
36—40	19.77	0.00	**44.10**	0.00	**43.70**
41—45	11.55	0.00	**27.11**	0.00	**25.24**
46—50	16.80	10.66	**28.79**	0.00	**31.06**
51—55	15.98	**57.62**	0.00	0.00	0.00
56—60	8.24	**29.72**	0.00	0.00	0.00
≥ 61	0.56	**2.00**	0.00	0.00	0.00
Tenure					
< 01	1.06	0.00	0.00	**4.57**	0.00
01—04	9.29	0.00	0.00	**40.00**	0.00
05—09	12.88	0.00	0.00	**55.43**	0.00
10—14	18.24	0.00	**37.64**	0.00	**35.07**
15—19	15.00	0.00	**28.04**	0.00	**29.52**
20—24	21.63	14.00	**34.32**	0.00	**35.41**
25—29	12.82	**50.32**	0.00	0.00	0.00
≥ 30	9.09	**35.68**	0.00	0.00	0.00
Qualification					
Intermediate executive	7.00	**100.00**	0.00	0.00	0.00
Higher skilled employee	18.19	0.00	**100.00**	0.00	0.00
Skilled employee	74.81	0.00	0.00	**100.00**	**100.00**
Total working time (month)					
< 147 hours	1.68	1.30	**3.38**	1.05	1.58
= 147 hours	83.14	**98.70**	**96.62**	75.97	78.09
> 147 hours	15.18	0.00	0.00	**22.98**	**20.33**
Type of contract					
Open-ended contract	94.90	**100.00**	**100.00**	75.02	**100.00**

Fixed-term contract	5.10	0.00	0.00	**24.98**	0.00
Basic hourly wage					
5.0 —7.5	14.18	0.00	0.00	**65.52**	0.00
7.5 —10.0	63.07	0.00	0.00	34.48	**100.00**
10.0 —12.5	16.99	0.00	**100.00**	0.00	0.00
12.5 —15.0	5.46	**94.76**	0.00	0.00	0.00
15.0 —17.5	0.30	**5.24**	0.00	0.00	0.00
Total hourly wage					
5.0 —7.5	1.86	0.00	0.00	**8.17**	0.00
7.5 —1	12.14	0.00	0.00	**53.37**	0.00
10.0 —12.5	31.12	0.00	0.00	38.46	38.37
12.5 —15.0	35.93	0.00	0.00	0.00	**61.63**
15.0 —17.5	9.12	0.00	**75.86**	0.00	0.00
17.5 —20.0	4.80	**27.45**	**24.14**	0.00	0.00
20.0 —22.5	2.34	**33.83**	0.00	0.00	0.00
22.5 —25.0	1.15	**16.57**	0.00	0.00	0.00
≥ 25.0	1.53	**22.15**	0.00	0.00	0.00
Last promotion					
2006	14.33	2.98	12.79	**23.37**	14.04
2005	22.05	0.00	**35.98**	**48.23**	13.62
2004	14.60	5.04	17.39	**19.39**	14.56
2003	10.85	7.04	**14.49**	3.70	14.07
2002	9.30	**18.24**	10.33	5.31	8.24
2001	21.36	14.51	9.02	0.00	**35.47**
< 2001	7.51	**52.19**	0.00	0.00	0.00