

Gorter, Cees; de Graaff, Thomas; Nijkamp, Peter

**Conference Paper**

## Estimating The Effect Of Labour Market Absorption Of Ethnic Minorities

40th Congress of the European Regional Science Association: "European Monetary Union and Regional Policy", August 29 - September 1, 2000, Barcelona, Spain

**Provided in Cooperation with:**

European Regional Science Association (ERSA)

*Suggested Citation:* Gorter, Cees; de Graaff, Thomas; Nijkamp, Peter (2000) : Estimating The Effect Of Labour Market Absorption Of Ethnic Minorities, 40th Congress of the European Regional Science Association: "European Monetary Union and Regional Policy", August 29 - September 1, 2000, Barcelona, Spain, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<https://hdl.handle.net/10419/114863>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Estimating the Effect of Labor Market Absorption of Ethnic Minorities

Thomas de Graaff\*, Cees Gorter and Peter Nijkamp

June 30, 2000

Preliminary version!

Paper prepared for the 40<sup>th</sup> European Congress in Barcelona

## Abstract

Recent migration studies are increasingly based on a network approach to account for unfavourable labor market and skill acquisition performances of ethnic minorities compared to that of the indigenous population. Socio-economic research emphasises that utilisation of social networks highly favors the probability of success in the labor market and in skill acquisition. Despite the increasing awareness of the importance of networks, empirical literature on the effects of networks is scarce, because of lack of data. Moreover, it is difficult to allow for endogeneity of the chosen network variable in combination with the latent nature of the network variable.

In our empirical application we estimate the network effect using individual data of a large survey among ethnic minorities in the Netherlands in 1994. This data source comprises both information on the labor market performance as well as detailed information on the characteristics of the personal and social networks of individuals. The approach adopted analyses individual labor market performance by using reduced form models in which the emphasis is on network or interaction effects. Preliminary results indicate that ethnic minorities with less developed networks that are not intertwined with that of the indigenous population have lower labor market performances.

**keywords:** Ethnic minorities, clustering, social networks, labor markets

## 1 Introduction

The concept of networks in economics – and in social science in general – has become very popular since the mid 1980s. Starting from real physical networks (Farrell and Saloner, 1986 and Katz and Shapiro, 1985), the theory was quickly extended to social networks (see Montgomery, 1991). In the mid 1990s a growing group of economists (see e.g. Durlauf, 1994; Bénabou, 1996) tried to model the interaction between various economic agents within a given spatial area, e.g. cities or neighborhoods. However, the network approach in the social sciences dates further back (Berkowitz, 1982; Rees, 1966; Granovetter, 1974) and is especially applied to socio-economic processes such as labor market performances.

In the Netherlands, the labor market performance of ethnic groups lags behind that of the indigenous population (see e.g. Tesser et al., 1999). These groups are characterized by strong ties within their community and are often subject to poor socio-economic integration in the country of destination. In 1996, Carrington et al. gave an insightful economic explanation why immigrants have the propensity to cluster, on the basis of a cost benefit analysis. By means of this process, initial conditions for ethnic enclaves are easily established. It is likely that geographical clustering is a determinant for labor market and schooling performance, although van der Klaauw (2000) did

---

\*Corresponding author, email tgraaff@econ.vu.nl. Spatial Economics, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV, Amsterdam. Tel: +31 (0) 20 4446098, fax: +31 (0) 20 4446004.

not find neighborhood effects for the labor market performance of non-Dutch job losers in the city of Rotterdam. However, as Manski (1993) indicated, there are severe problems inherent in measuring neighborhood effects, because, without a control group of individuals, a direct distinction between exogenous, endogenous and correlation effects is not possible. However, with prior information or additional assumptions on the behavior of the sampled individuals, indirect results about the separate effects may be attainable.

In this paper we aim to map out the labor market absorption of ethnic minorities in the Netherlands. We are especially interested in the significance of the following three main characteristics of the probability of success on the Dutch labor market of ethnic minorities.

- *Language Ability.* The ability to speak the indigenous language seems highly important for labor market performance of ethnic minorities (see Chiswick and Miller, 1999). Furthermore, with the increasing emphasis on communication proficiencies for higher skilled jobs, this variable seems to be highly important in determining the social segregation of a local ethnic group. We take this variable mainly as exogenously given.
- *Social Clustering.* This concept measures the social integration of ethnic minorities in a society. A strong ethnic network might prohibit the integration of the ethnic population. On the other hand, large social intertwinement between the ethnic and the indigenous population may cause increased labor market and schooling performance, until the labor market characteristics of both the indigenous population and the ethnic minority are the same. In terms of Manski (1993), social network variables may be seen as endogenous effects.
- *Geographical Clustering.* This concept measures directly the physical segregation within a society. If there is a high degree of geographical clustering in a society, neighborhoods or peer effects have more influence on the labor market absorption of ethnic immigrants. Geographical clustering may cause individuals to behave similarly, thus causing correlation, although geographical clustering might also be considered as an endogenous selection effect.

These variables are all – with the exception perhaps of the geographical clustering variable – rather hazy concepts and difficult to measure. Although they can be characterized by proper indicators, these can never fully represent the concept underlying them. That is why we have to construct meaningful latent variables from these indicators, which should be able to capture the true value of the underlying concept.

This paper is organized in the following way. In Section 2, we will characterize our data by using correspondence analysis on our dataset. Section 3 will continue with applying factor analysis on language performance indicators and subsequently indicators on social activities, in order to construct latent variables for language proficiency and social networks. The last subsection in Section 3 will deal with the construction of a variable for geographical clustering, which is mainly based on De Graaff et al. (2000). Section 4 will use the artificial variables to measure labor market absorption. Here, duration models will be developed in order to estimate determinants of employment spells. This Section essentially describes the extent to which persons belonging to an ethnic minority have chances to lose a job and how these chances depend on language and social network variables. Finally, Section 5 offers some tentative conclusions and directions for further research.

## 2 The Data

In this paper we use the SPVA<sup>1</sup> survey from 1994. In this survey four ethnic minorities (Turks, Moroccans, Antilleans and Surinamese) are questioned, together with the indigenous population, as the reference group. In total, 4,096 households are surveyed, which consist of 8,304 respondents

---

<sup>1</sup>The abbreviation stands for "Sociale Positie en Voorzieningengebruik Allochtonen" (in Dutch).

of 12 years and older in 13 cities in the Netherlands<sup>2</sup>. Because the number of questions asked to the heads of the households amounts to 255, we will first use correspondence analysis, followed by factor analysis in Section 3, in order to present the data and to reduce the number of variables substantially. The next subsection will use correspondence analysis as an intuitive tool to gain insight in our large data set.

## 2.1 Ethnic Minorities: a Typology

Working with large data sets could be problematic in the sense that an overview of the relationships is difficult to attain. Furthermore, because large numbers of observations have the tendency to give significant results quickly, proper model specification is troublesome. Correspondence analysis could help in order to make the data surveyable and to give a useful typology of the data. This type of analysis is an inductive method, which can be described as a search strategy for underlying structures within a given data set. The particular strength of correspondence analysis lies in visualizing patterns in large datasets, thereby removing redundant noise without losing essential information (see e.g. Claussen, 1998). The technique is a tool for analyzing the association between two or more categorical variables by representing the two or more variable categories of the variables as points in a low-dimensional space, often only two dimensions. If two categories are very similar, then they will have a small distance in space to each other. Historically seen, correspondence analysis dates far back into the 1930s, but the method became mainly popular in continental Europe, especially in France (for an enlightening application see Bourdieu, 1988) and in the Netherlands (see e.g. van der Heijden and de Leeuw, 1985), and less in English speaking countries.

First, for applying correspondence analysis one should transform the raw data matrix  $\mathbf{X}$  into a contingency table  $\mathbf{A}$ , with row dimension  $I$  and column dimension  $J$ , taking into account the relative weight of each frequency. Next, one is able to calculate the weighted Euclidean distance between each entry  $(i, j)$  in the frequency table in the following way:

$$\delta^2(i, i') = \sum_j \frac{(a_{ij} - a_{i'j})^2}{a_{.j}} \quad (1)$$

$a_{ij}$  and  $a_{i'j}$  are column elements  $j$  in the row  $i$  and  $i'$  respectively, the  $\cdot$  represents summed over elements over the corresponding variable and  $\delta^2(\cdot)$  is the so-called chi-squared distance. Equation (1) is a measure of the difference between row profile  $i$  and row profile  $i'$ .<sup>3</sup> Finding chi-squared distances over column profiles is just the transposed operation.

Correspondence analysis is based on (1) and basically decomposes this chi-square metric. First, let  $\mathbf{D}_r$  and  $\mathbf{D}_c$  be diagonal matrices with dimension weights  $a_{i\cdot}$  and  $a_{\cdot j}$  respectively (these are called the row and column masses). Then we are able to construct a matrix  $\mathbf{E} = (\mathbf{D}_r \mathbf{A} \mathbf{D}_c) / n$ , with  $n = \sum_{i,j} a_{ij}$  and  $\mathbf{1}$  a vector of ones (for a detailed analysis, see Greenacre and Blasius, 1994). The elements of  $\mathbf{E}$  now have the following form:

$$e_{ij} = \frac{a_{i\cdot} a_{\cdot j}}{n}, \quad (2)$$

and can be interpreted as the average of the column times the row masses.

Now, we are able to calculate the matrix of standardized residuals or  $\mathbf{D}_r^{-1/2} (\mathbf{A} - \mathbf{E}) \mathbf{D}_c^{-1/2}$  and with the use of Singular Value Decomposition we can find the diagonal eigenvalue matrix  $\mathbf{\Lambda}$ :

$$\mathbf{\Xi} = \mathbf{D}_r^{-1/2} (\mathbf{A} - \mathbf{E}) \mathbf{D}_c^{-1/2} = \mathbf{U} \mathbf{\Lambda}^{1/2} \mathbf{V}'. \quad (3)$$

The diagonal elements of  $\mathbf{\Lambda}^{1/2}$  are the so-called singular values, which are positive numbers and arranged in descending order:  $\lambda_1^{1/2} \geq \lambda_2^{1/2} \geq \dots \geq \lambda_K^{1/2} > 0$ , with  $K$  de rank of  $\mathbf{\Xi}$ . In the context of

<sup>2</sup>The following cities are used in the sample: Amsterdam, Rotterdam, The Hague, Utrecht, Eindhoven, Enschede, Almere, Alphen aan de Rijn, Bergen op Zoom, Hoogezand-Sappemeer, Delft, Dordrecht and Tiel. These cities comprise the four largest cities in the Netherlands.

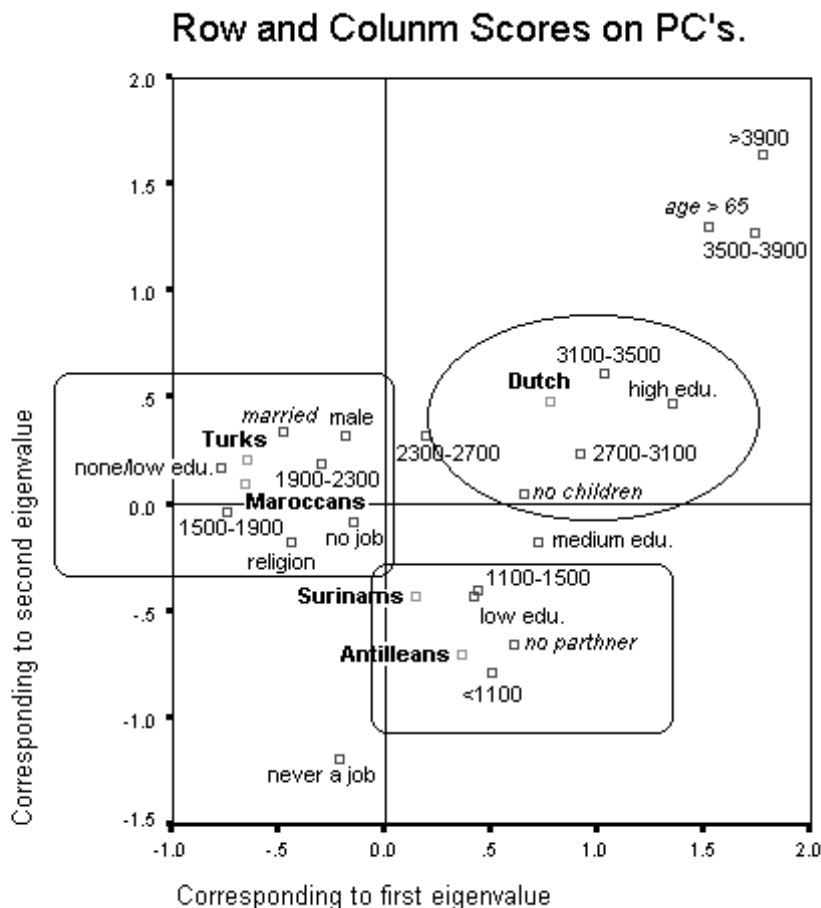
<sup>3</sup>If  $i$  and  $i'$  have the same profiles, then  $\delta^2(\cdot) = 0$  and the matrix  $A$  is singular

correspondence analysis we refer to  $\lambda_i^{1/2}$  as a principal inertia.<sup>4</sup>  $\mathbf{U}$  is the matrix of the eigenvectors of  $\mathbf{\Xi}\mathbf{\Xi}'$ , while  $\mathbf{V}$  is the matrix of the eigenvectors of  $\mathbf{\Xi}'\mathbf{\Xi}$ . This solution has a dimension equal to  $K = \min\{I - 1, J - 1\}$ . Especially, the matrices  $\mathbf{U}$  and  $\mathbf{V}$  prove to be quite useful. Namely the principal coordinates ( $\mathbf{PC}'$ s) of e.g. the rows are obtained by calculating  $\mathbf{PC}_r = \mathbf{D}_r^{-1/2}\mathbf{U}\mathbf{\Lambda}^{1/2}$ , while the standard coordinates of the rows are the  $\mathbf{PC}'$ s divided by their respective singular values, so  $\mathbf{C}_r = \mathbf{D}_r^{-1/2}\mathbf{U}$ . Both principal coordinates and the standard coordinates can be used to display the variables in a two-dimensional or three-dimensional graph. Higher dimensions are also attainable, but are not considered here because of the graphical difficulties to display them. The procedure for obtaining principal and standard coordinates is exactly the same for the columns.

Correspondence analysis can be seen as a decomposition of the departure from independence in a matrix. The relation between this decomposition and Pearson's goodness-of-fit  $\chi^2$  statistic<sup>5</sup> is:

$$\text{trace } \mathbf{\Lambda}^2 = \sum_i \lambda_i = \frac{\chi^2}{n}. \quad (4)$$

The interpretation of the graphical interpretation of correspondence analysis is now the following. When profiles between row or column points are considerably different, the distance between the two points is quite large, their profiles are rather equal when they are quite near each other.



<sup>4</sup>The sum of the squares of the elements of  $\mathbf{\Xi}$  is called the total inertia and is a measure of the column-row association.

<sup>5</sup>Pearson's goodness of fit  $\chi^2$  statistic is defined as  $\chi^2_{(I-1)(J-1)} = \sum_i \frac{(x_i - \bar{x})^2}{\bar{x}}$ .

Figure 1: A correspondence analysis of five ethnic groups in the Netherlands. Demographic variables are in italics, the ethnic groups in bold and the figures represent household incomes.

When a point is not far from the origin, then the row and column profiles from this point will not differ much from the mean profile. The principal inertia can be used to decompose the total inertia and to measure the contributions of each dimension to the solution. Figure 1 gives a typology of the above mentioned five ethnic groups in the Netherlands using a correspondence analysis in two dimensions also explained above, with the principal coordinates on the axes. The typology is applied on the characteristics on the heads of the households, which can be found in the contingency table A1 in the Appendix. The cumulative explained inertia by the two most important dimensions is 96%, where dimension 1 accounts for 77% and dimension 2 for 19%. Total inertia is about 0.16, which seem to indicate that there is quite some correlation between our column and row dimension. Indeed, Pearson's  $\chi^2$  statistic is equal to more or less 3187, which is significant with 80 degrees of freedom, which indicates large dependence between the row and column variables in table A1 in appendix A.

Figure 1 reveals that there are significant socio-economic differences between the five ethnic groups, although the characteristics of the Turks and the Moroccans seems rather similar. Also the Surinamese and Antillean household heads lie close to each other in the two dimensional subspace. That Dutch households incomes and education of the households heads are higher comes not as a surprise. Furthermore, the variable '*age > 65*', which is in the Netherlands the retirement age, lies far from the Turks, Moroccans, Surinamese and Antilleans, indicating the still young demographic profile for these groups. Typical for the Turks and Moroccans is that the head of the household has a higher probability of being a male, having no job, being married and being religious. On the other hand, Surinamese and Antilleans household heads are quite likely to have no partner and low incomes. These typologies seem to be justified for the Netherlands, according to e.g. Tesser et al. (1999). Somewhat strange is the fact that Surinamese and Antillean household heads have higher education but lower income on average than Moroccan and Turkish household heads. However, this could be explained by the knowledge that Surinamese and Antillean households are more often single mother with children households. The following Section will try to find latent variables for language and social network variables, that are able to explain the differences found in this subsection and also a geographical measure suitable for capturing physical networks between ethnic minorities.

### 3 Determinants of Integration

In comparison with correspondence analysis, the main purpose of factor analysis is not to gain insight in the structure of the data. Instead, it want to reduce the number of variables, in order to make the data more manageable and tractable. Factor analysis is essentially a set of techniques that are able to make a number of new (latent) variables, usually smaller than the number of original variables, which are able to capture most of the correlation between the original variables. Where correspondence analysis uses a two-dimensional variable structure, factor analysis uses a one-dimensional variable structure, though they essentially make use of the same matrix algebra, namely matrix decomposition. In this section we will explain the basic mechanics of factor analysis, but for more information we refer to Jackson (1991). First, we want our new variables to ideally have the following two properties:

1. *Independency.* The new variables you create should be independent from each other (in other words, orthogonal to each other) which facilitates further analysis, as regression analysis or in our case duration analysis.
2. *Interpretability.* The new latent variables should also have an easy to understand interpretation, so that coefficients in further analysis are easier to discuss.

Principal Component Analysis (PCA) as the first step is the most common practice in factor analysis. Just as in subsection 2.1 we use the data matrix  $\mathbf{X}$  with  $m$  variables and now we transform it in its correlation matrix  $\mathbf{\Sigma}$ , which has size  $(m \times m)$  and is symmetric.<sup>6</sup> Now we may state that with Schur decomposition:

$$\mathbf{\Sigma} = \mathbf{S}'\mathbf{\Lambda}\mathbf{S} \quad (5)$$

, where  $\mathbf{S}$  is an orthogonal matrix of eigenvectors of  $\mathbf{\Sigma}$  and  $\mathbf{\Lambda}$  a matrix with the eigenvalues of  $\mathbf{\Sigma}$  on the diagonal, again in a descending order. Note the similarity with (3). The vectors of  $\mathbf{S}$  (also called the factor loadings) can now be used to weight the original variables to form  $m$  new variables (the principal components). The results are that the created principal components are orthogonal on each other (they are independent) and that the  $m$  eigenvectors span the subspace  $\mathbb{R}^m$  (nothing from the original information has been lost).

The first eigenvector can be interpreted as that direction in  $\mathbb{R}^m$ , in which the original data show the largest variance. The second eigenvector is that direction in  $\mathbb{R}^m$  with the second largest variance and so on. In order to reduce the number of variables we only have to take the  $k \leq m$  eigenvectors and transform them in principal components. The  $m - k$  omitted variables may be regarded then as noise. There are a large number of criteria in how to choose the value of  $k$ , but most common in practice is to only take those eigenvectors, whose corresponding eigenvalues are larger than 1.

Other problems which arises is that the eigenvalues of all  $m$  vectors should be different from each other, in order to properly define the subspace. Two eigenvalues almost being equal could heavily inflate the standard errors of the factor loading. In econometric terms we may say that we estimate the orientation consistently, but that we do not know the efficiency. In most applications the eigenvalues of the first  $k$  vectors are quite large and observable different from each other. However, the eigenvalues of the last  $m - k$  vectors are quite small and mostly of the same order of magnitude. We use a test developed by Bartlett (see Jackson, 1991) to test for the equality of the eigenvalues.<sup>7</sup>

Interpreting these  $k$  new independent principal components is quite difficult. We ideally want our factor loadings to have values like -1, 0 or 1, so that each principal component is entirely constructed out of a (small) number of variables. In reality, however, our principal components are quite likely to be constructed out of all variables. An obvious approach in this case is matrix rotation. There are basically two rotation methods. The first one leaves the principal components orthogonal, while the other does not retain orthogonality. The most popular rotation method, which leaves the principal components orthogonal, is the so-called Varimax rotation procedure. As already mentioned at the beginning of this Section, there is a kind of trade-off between the two desirable properties: independency and interpretability. Using non-orthogonal rotation schemes make the new factors dependent of each other but increases ease of their interpretation..

Let  $b_{ij}$  be an element of  $\mathbf{B} = \mathbf{R}\mathbf{S}_k$ , with  $\mathbf{R}$  a rotation matrix and  $\mathbf{S}_k$  the matrix of the first  $k$  factor loadings, then the varimax procedure looks like the following:<sup>8</sup>

$$Q = \sum_{j=1}^k \left( \sum_{i=1}^m b_{ij}^2 - \frac{1}{k} \left( \sum_{i=1}^m b_{ij}^2 \right)^2 \right), \quad (6)$$

where, the variable  $Q$  is maximized by varying the rotation matrix  $\mathbf{R}$ , with size equal to  $(k \times k)$ . This procedure keeps the orthogonality of the principal components and maximizes the sums of

<sup>6</sup> Instead of using the correlation matrix, we may also use the covariance matrix. However, the principal components of the covariance matrix are difficult to interpret, when the data are measured in different units. This problem does not occur when working with the correlation matrix, because its elements are unitless and scaled to 1.

<sup>7</sup> Bartlett's test is defined by Jackson (1991) as:  $\chi^2_{\frac{1}{2}(m-k-1)(m-k+2)} = -v \sum_{i=k+1}^m \ln \lambda_i + v(m-k) \ln \left[ \frac{\sum_{i=k+1}^m \lambda_i}{(m-k)} \right]$ ,

with  $v$  the number of degrees of freedom of the used covariance matrix and  $\lambda_i$  the corresponding eigenvalue. Note that this test-statistic may be used for each null model that assumes some of the eigenvalues to be equal.

<sup>8</sup> Most orthogonal rotation procedures, like e.g. orthomax, can be derived from (6).

squares of  $\mathbf{B}$  columnwise.

In the following two subsections we will use factor analysis to construct two new sets of latent variables, namely language and social network variables. In the last subsection we will give a useful geographical clustering variable, which is capable of taking into account the various city sizes occurring in the SPVA survey.

### 3.1 Language Variables

In determining the success of the absorption of immigrants, language skills are key variables, as already argued by e.g. Chiswick and Miller (1999). Language proficiency does not only facilitate finding a job, but in most cases also increases productivity, due to the increasing importance of communication. If immigrants have low-skilled jobs, complementary to those of the indigenous (high-skilled) population then communication between those two groups is highly preferable and higher returns to language skills are given. In this paper we try to estimate language skills directly from the data, although there can be argued that language skills are a function of the characteristics of the immigrant. According to Chiswick and Miller (1999) the following functional form may be constructed (with the expected sign of the effect in parentheses):

$$\text{Language} = \{ \text{duration of residence (+), marital status (-), age (-), education (+), ethnic density (-), children (?) } \}.$$

Language proficiency will increase with the duration of residence in the country of destination. If the length of the residence increases, immigrants tend to adjust to the country of destination. This situation may also be encountered in the Netherlands (see Tesser et al., 1999). There may be argued that a nonlinear relationship is present in the form of decreasing returns of duration of residence, but this problem can be easily be encountered by squaring the duration of residence.

Marital status can be quite important at the time of migration. If the migrant is married then the probability of speaking his native tongue at home is quite high, thereby reducing the need for acquiring proficiency in the language of destination. From the linguistics literature we know that the ability to learn a new language is quite superior at a younger age and decreases according as the immigrants gets older, so the higher the age, the more difficult it will be for the immigrant to obtain new language skills. Also the degree to which the immigrant has received education seems to have a significant impact on the language proficiency. The negative effect of the number of immigrants from the country in a certain area comes quite close to the social and geographical clustering variables as we have introduced before, which again stresses the fact to be careful of endogeneity between e.g. language variables and social clustering variables. The last determinant in the functional form above is that of having children, which has an ambiguous effect. First of all, children will learn the alien language faster and are therefore able to teach the parents the language. On the other hand, children may act as interpreters, which reduced the incentives for the parents to learn the language. As a last remark on having children, we note that children could have a depressing effect on the labor supply of their mothers. If language proficiency is an investment in the labor market, then having children could largely remove the motives to make these investments.

In our data set, several variables dealing with language proficiency are available. Variables dealing with duration of residence, age, education and ethnic density will not be taken into account when constructing the factors for two reasons. First, no good variables are present in the data set that deal with the aforementioned determinant in combination with language proficiency. Secondly, those determinants will be used in other variables in the final analysis, so that leaving those variables out will remove problems of multicollinearity. Variables dealing with the marital status and children of the immigrant in combination with language ability are taken up in the analysis.

Table 1 presents the variables used and their descriptions in this factor analysis.



Table 1: Language Variables and their definitions

Variables	Definitions
Langpart	Immigrant speaks Dutch with partner
Childnl	Immigrant has children in the Netherlands
Langchil	Immigrant speaks Dutch with his/her children
Otherlan	Immigrant speaks an other language
Langconv	Immigrant has difficulty with Dutch in the interview
Langratv	Immigrant has difficulty with Dutch on radio and TV.
Langnews	Immigrant has difficulty with Dutch in newspapers and such
Letters	Immigrant writes letters in Dutch

Figure 2 presents the factor plot of these 8 variables. We used here principal components analysis in combination with the Varimax rotation scheme on the individuals belonging to the Surinamese, Turkish, Antillean and Moroccan ethnic communities. The first two factors account for 58 percent of the variance in the eigenvalues and the 6 other eigenvalues do not seem to be identical.<sup>9</sup>

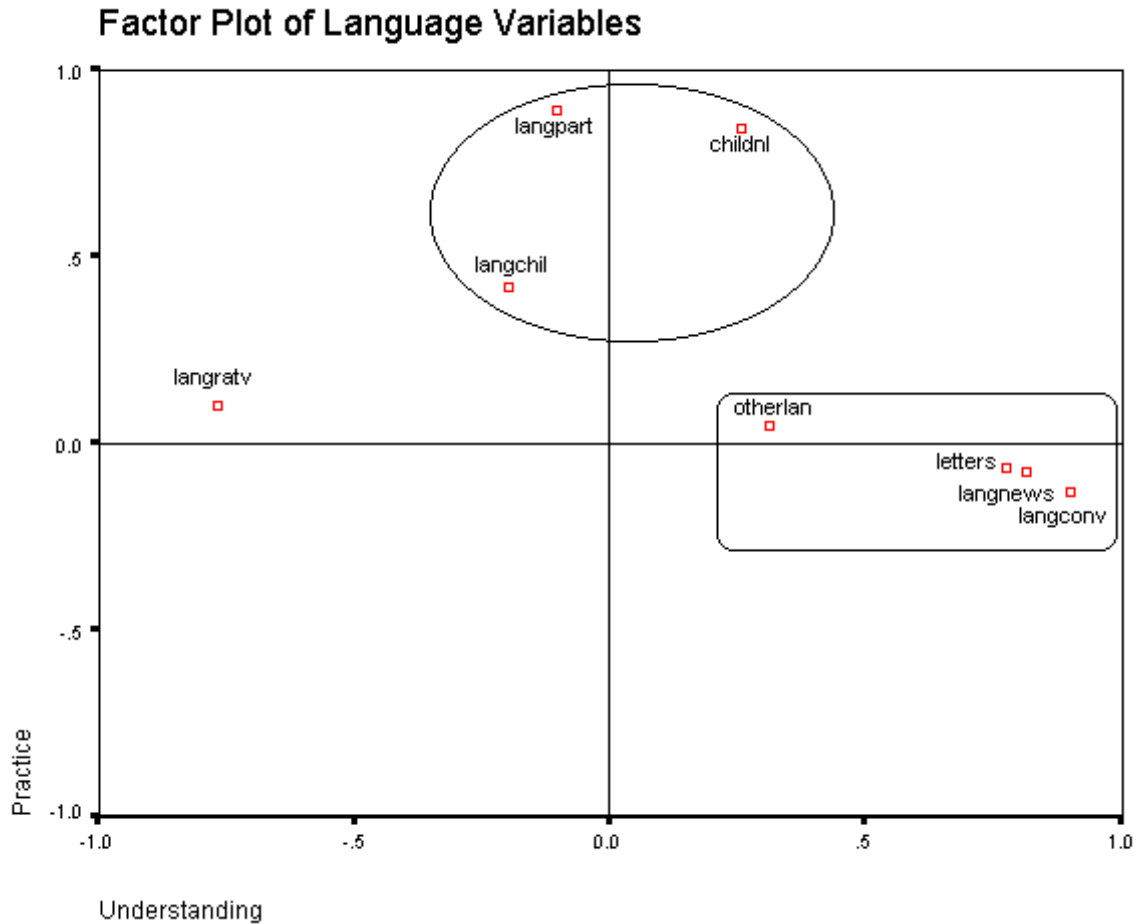


Figure 2: A factor plot of language variables in two dimensions.

Figure 2 shows that the language variables may be divided into two factors, apart from the two variables ‘Otherlan’ and ‘Langratv’. The variables ‘Otherlan’ lies close to the origin, indicating

<sup>9</sup>Bartlett’s test indicates here a chi-square value of 7071.3026 with a probability of 0.00.

that it does not have a large effect on the two factors. Indeed, its corresponding eigenvalue is quite small. The variable ‘Langratv’ is somewhat more strange. However, the variables lies close to the x-axis and has also a large effect on the factors. Observing figure 2, we are now able to make two new latent language variables:

- *Understanding.* This is a variable, that capture the ability of the immigrant to understand Dutch media (newspaper, radio, television), and to a certain extend also the ability to understand the question of the interviewer. This variable could also be seen as a measure of being able to receive information from the outside world. A high variable ‘understanding’ must be considered as a high ability to understand the Dutch media (and the interviewer).
- *Practice.* This variable measures to which extend the immigrant actually uses the Dutch language (with his or her partner or child). The fact of having children is also important in this context, which can be explained by the educational system, that forces children to use the Dutch language and in an indirect way also their parents. If this variable is low, then the impact of the Dutch language proficiency must be considered high.

The 8 variables do not seem to be perfectly explained by the two factors. However, this may be explained by sampling error, the ordinality of the data and especially the large separation of the meaning of some of the variables. On the other hand, 58 percent is a rather good score, if one takes the different realistic meanings of the variables into account. Furthermore, our two latent variables seem to capture the underlying structure rather well. The following subsection will deal with the creation of latent social network variables.

### 3.2 Social Network Variables

The second class of latent variables we would like to investigate is that of the social network variables. From sociological literature (e.g. van Amersfoort and Cortie, 1996) but also from economic literature (Montgomery, 1991; Rees, 1966) we know that social networks are important for labor market and schooling performance. However, we do not know exactly the channels via which this process occurs. Sharing information with other individuals, though, seems to be a key variable. Considering this, we can divide the social network of an immigrant in two main elements. First of all, the social network *within* his own ethnic community and, secondly, the social network *between* the ethnic community and the indigenous population. So, at the least, we want our latent variables to have these two characteristics. Table 2 presents the observed variables we used to construct our latent variables.

Table 2: Social Network Variables and their definitions

Variables	Definitions
Concethn	Immigrant wants to live in neighborhood with own ethnic composition
Mixpart	Immigrant has partner from other ethnic group
Ethorgan	Immigrant has contact with his/her ethnic community of interest
Leisnl	Immigrant meets Dutch individuals in leisure time
Friendnl	Immigrant attaches value to having Dutch friends
Visits	Immigrant gets visited by Dutch friends
Member	Immigrant is member of an organization
Placethn	Immigrant visits places frequented by individuals from the same ethnicity

The variables in table 2 are variable that deal with networks of immigrants within their ethnic group and between several ethnic groups. The variables ‘leisnl’, ‘friendnl’, ‘visits’ gives information about the intertwinement between the immigrant and the Dutch community. ‘Mixpart’ gives even information whether the immigrant is prepared for the absolute mixture of social networks by living with a partner from another ethnic group. ‘Concethn’ informs us about the racial preferences of the immigrant for his neighborhood. ‘Ethorgan’ and ‘placethn’ provides us with information about the social network of the immigrant within his own ethnic community, while ‘member’ informs us

whether the immigrant engages in social outdoor activities. So the last variable is some kind of proxy of the total social network of the immigrant. Figure 3 gives us then the factor plot of the social network variables.

The proportion of variance explained by the first two factors is here 45%, while Bartlett's test does not indicate that the last 6 eigenvalues are identical.<sup>10</sup> One could object that the amount of variation explained here is rather low. However, just as in subsection 3.1, the variables used here are quite divers in their meaning. Together with again their ordinal values and some inevitable sample error, the proportion of variance explained could be considered quite reasonable. Note that the third largest eigenvector is marginally above, indicating a three-dimensional factor structure. For the sake of simplicity, we will continue with a two-dimensional structure.

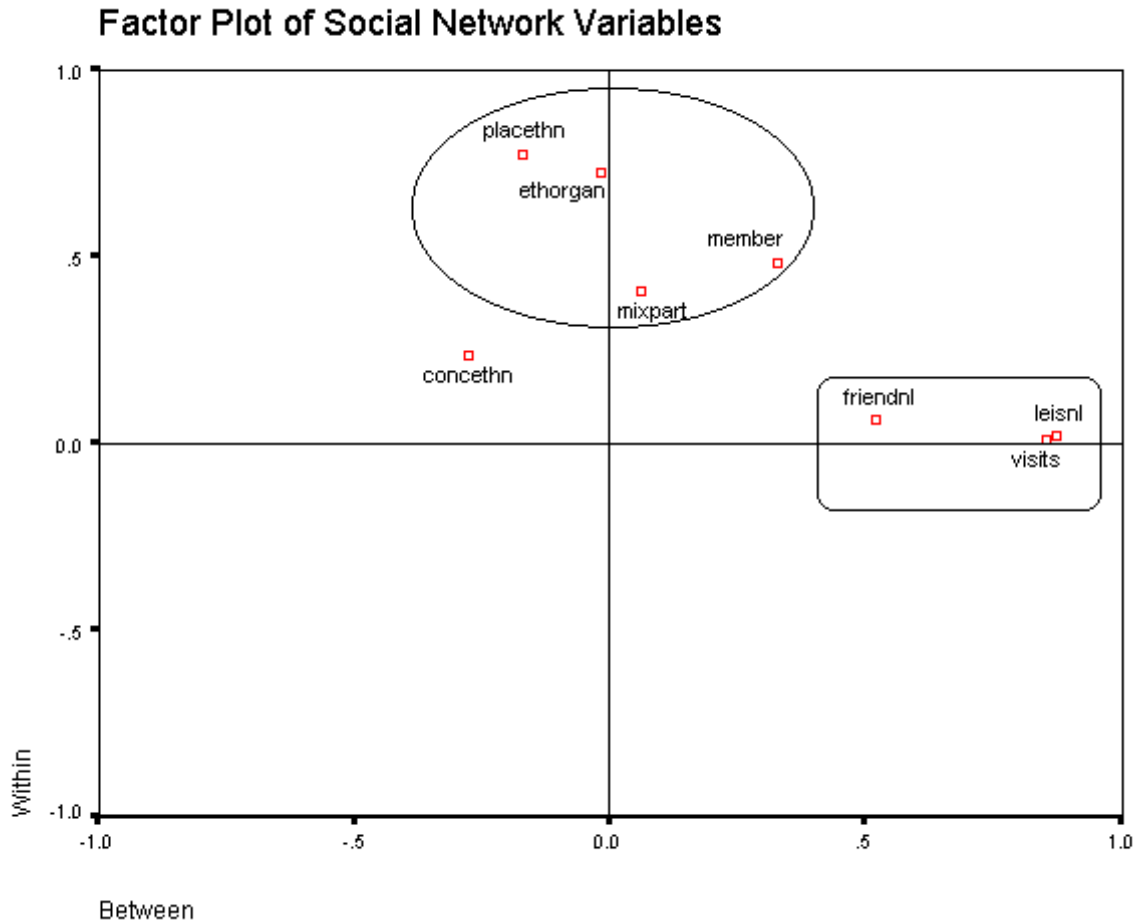


Figure 3: A factor plot of social network variables in two dimensions.

The picture drawn in figure 3 is not quite that coherent as in figure 2. However, giving interpretation to the two latent variables is rather straightforward, namely:

- *Between network*: The social network between the ethnic group and the indigenous population. The variables concerning the contact with the Dutch indigenous population are rather clustered. Even the variable regarding the preference of the immigrant of the neighborhoods racial population fits in rather well, if one takes into account that the sign of 'concethn' in the sample is reversed with respect to the other 'between network' variables. The higher the

<sup>10</sup>To be precise, Bartlett's test has as an chi-square outcome of 3290.33 with a probability of 0.00.

‘between network’ variable, the lower the impact is of the network between the two ethnic groups.

- *Within network*: The social network within the ethnic population. These variables are less clustered, mainly because ‘mixpart’ and ‘member’ are not real within network variables. The peculiar fact that ‘mixpart’ is highly determined by the within network factor, could be explained by the fact that the variable ‘mixpart’ is merely a measure about the tightness of the ethnic social network. The variable ‘member’ is definitely a combination between the between network factor and the within network, which is explained by the fact that ‘member’ is kind of proxy for the height of both these factors, namely the total social network. If the variable ‘within network’ is high, then the social network within the ethnic population of the individual has to be considered insignificant.

After constructing our network and language variables we will now turn to our third variable of labor market absorption, namely geographical clustering.

### 3.3 A Geographical Clustering Variable

It is not difficult to see that area composition effects could be an important determinant for labor market absorption. If an area exhibits a large clustering of one or more ethnic groups, then we know that the characteristics of the inhabitants of that area display a large correlation and may also differ from the mean characteristics. The correspondence analysis performed in Section 2 provides further evidence for this notion. Another cause for social interaction effects may be self selection effects of ethnic groups in their preference for a particular residential area. For instance, persons with low wages may be highly clustered, because their preference for cheap housing may be highly correlated. However, in empirical applications it is rather difficult to identify the exact reason for the impact of clustering or composition effects. For now we are only interested in the impact of geographical clustering, and we will temporarily forget the various economic reasons that could explain this impact.

Finding a proper geographical clustering variable, based on economic theory, is rather difficult. Especially, because we want our variable to have the useful property of scale independency. Because the cities in our dataset do not have the same size (see table 2A in the Appendix), scale sensitivity of an index could create a severe bias. The famous Gini’s ratio, for example, can not be used here because this variable is sensitive to extremes in the distribution and thus scale dependent. In de Graaff et al. (2000) we proposed a new geographical clustering variable, the gamma index, based on the index developed by Ellison and Glaeser (1997). This index has two major advantages. First, the index take the sizes of the other areas into account and is scale independent. Secondly, the index is based on a location choice model, where a person chooses his location based on natural advantages (observed and unobserved) and based on network externalities created by previous movers to that location.

Let  $s_{im}$  be the percentage of the ethnic group  $m$  living in area  $i$ , and let  $x_m$  be the share of the ethnic group in the total population. Furthermore, let  $m \in \{1, \dots, M\}$ , with  $M$  the total number of ethnic groups, then we may construct the following area clustering index:

$$\gamma_i = \frac{\sum_{m=1}^M (s_{im} - x_m)^2}{1 - \sum_{m=1}^M x_m^2} \quad (7)$$

The gamma coefficient,  $\gamma_i$ , is now a measure of the amount of clustering in that particular area and can be calculated for all areas, regardless of their size of population. The data we use originates from the Dutch bureau of statistics and gives us the information about the composition of 8 ethnic groups<sup>11</sup> living in each municipality in the Netherlands. This data is released the first

---

<sup>11</sup>These comprises the first and second generation immigrants. Namely, those people who are born in a foreign country of from whom at least one parent is born in a foreign country. If both parents a born in a foreign, but

of January, 1995, and can be seen as the resultant area conditions of the surveyed ethnic groups in the SPVA (these people were actually questioned in the spring of 1994). The gamma coefficient can take up a value between 0 and 1, where 0 represents complete random clustering and 1 extreme clustering. We consider a low amount of clustering when gamma is between 0 and 0.02, a medium amount of clustering when between 0.02 and 0.05 and a high amount of clustering appears when the coefficient is above 0.05.

Table A2 in the appendix gives the gamma coefficients for the cities surveyed in the SPVA. The same pattern appears as we found for the municipalities in 1998 in de Graaff et al. (2000), namely high geographical clustering only occurs in the four largest cities in the Netherlands. The other municipalities do not display a significant amount of clustering in their population composition, even for medium sized cities. In the following Section we will investigate to what extent differences in clustering between the cities may explain differences in labor market behavior.

## 4 The Labor Market Performance

In this section we want to investigate the actual labor market performance of the four surveyed ethnic groups. In order to do so, we will first developed a proper model to examine employment spells and subsequently we will give the estimation results. Our main research objective in this Section is to research to what extent the developed language, social clustering and physical clustering variables are contributing to actual labor market performances of the four sampled ethnic groups. Furthermore, we are interested in what way our surveyed ethnic groups relate to the labor market performances of the indigenous population. Although the results are still rather preliminary, results indicate a high impact of language proficiency and social network and low impact of clustering on the city level.

### 4.1 Determinants of Employment and Unemployment

In this subsection we will developed duration models – to be precise mixed proportional hazard models – specifically designed to estimated the employment spells in our dataset correctly. These spells could also be estimated by means of regression techniques, although this would only be allowed if all spells were randomly collected and if the impact of the exogenous variables do not change when the duration of the spells changes. However, it is very likely that processes as stigmatization and the loss or gain of specific skills play a significant role on the labor market, especially if the durations of the spells are rather large. Moreover, our sample does not even have a random collection of employment spells, so it is very likely that regression estimates would be biased (see Kiefer, 1988). However, the most important reason to use duration models in our case is the fact that we are dealing with right censored stock data. This means that the distribution of employment spells is not complete, but only partly observed. These types of data are known to have an overrepresentation of long spells. Therefore, we will use the more appropriate duration models. Moreover, because a large number of individuals in our sample are first generation immigrants, we encounter two main problems in modelling:

1. Our observations are right censored, i.e. the on-going spells are normally longer than can be observed at the time of sampling (point **B** in figure 4).
2. In the case of first generation immigrants, the period of working could never be larger than the period that the migrant is in the country of destination (from point **M** in figure 4). This condition should be taken into account in the distribution functions of the duration spells.

---

different country, then the child is assigned to the country of its mother. Surinam and the Dutch Antilles are considered here as foreign countries, although they are (former) colonies.

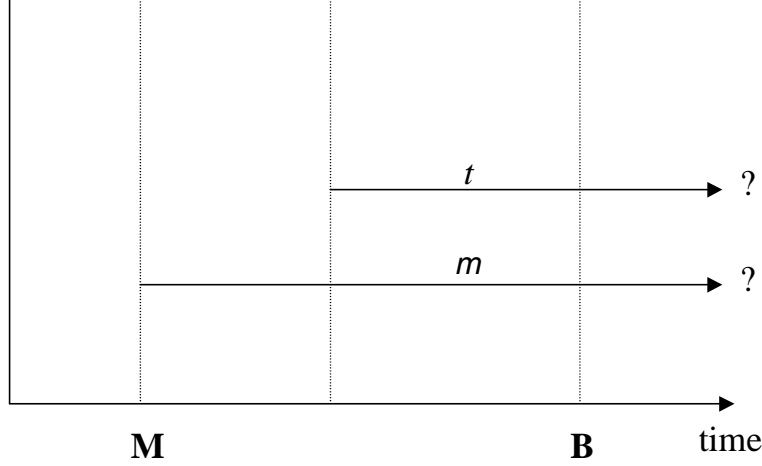


Figure 4. Duration diagram for a first generation immigrant. With **M** the moment of migration, **B** the moment of sampling,  $t$  the period in the spell and  $m$  the period in the country of destination.

The first problem is nowadays rather straightforward to solve (see e.g. Kooreman and Ridder, 1983) where the second problem is not difficult to encapsulate in a duration model either. As mentioned before we will use a duration model in the form of a Mixed Proportion Hazard (MPH) model<sup>12</sup>, mainly because of the model's flexibility and also because it is now the most popular and best known duration model. MPH models have the following form:

$$\delta_i(t) dt = \lambda_i(t) \cdot \tau(t) \cdot \underline{u}_i \cdot dt. \quad (8)$$

In this expression,  $\delta_i(t)$  is the probability of leaving a spell for each individual (the hazard rate),  $\lambda_i(t)$  is the effect of exogenous variables on the hazard rate,  $\tau(t)$  takes into account the duration dependence on the hazard rate and  $\underline{u}_i$  is an individual random error term due to omitting variables. Because our data consists only on incomplete spells and because the first generation migrants must be conditioned on their time of migration, the whole density function of our incomplete spells may be denoted as:

$$\tilde{h}(t) = \begin{cases} \frac{\Pr(\underline{t} > t)}{\int_0^M \Pr(\underline{t} \geq s) ds} = \frac{1-F(t)}{E(\underline{t} | \underline{t} \leq M)} & \text{For first generation immigrants} \\ \frac{\Pr(\underline{t} \geq t)}{\int_0^\infty \Pr(\underline{t} \geq s) ds} = \frac{1-F(t)}{E(\underline{t})} & \text{Other} \end{cases}. \quad (9)$$

Here,  $E$  denotes the expectation operation,  $\underline{t}$  is a random variable,  $F(\cdot)$  is the distribution of the complete spell and **M** is the time of immigration for a first generation immigrant.

Because our hazard rate of (8) can be denoted as  $\delta(t) = f(t)/(1-F(t))$ , the following relation between (8) and  $1-F(t)$  can be easily derived:

$$1-F(t) = \exp\left(\int_0^t -\delta_i(s) ds\right). \quad (10)$$

For the function representing the exogenous variables we now define  $\lambda_i(\cdot)$  as usual:

$$\lambda_i(X(t)) = \lambda_i(X) = e^{X\beta}, \quad (11)$$

with  $\beta$  the parameters to be estimated. We define the duration dependence  $\tau(t)$  as 1 or, alternatively, as a Weibull specification form:

<sup>12</sup>For an overview of hazard models see inter alia Gorter (1991), Kiefer (1988) and Lancaster (1990).

$$\tau(t) = \alpha.t^{\alpha-1}. \quad (12)$$

$\alpha$  is here a duration parameter to be estimated; if  $\alpha = 1$ , there is no duration dependence, if  $\alpha < 1$  there is negative duration dependence and if  $\alpha > 1$  there is positive duration dependence. And last, the error term  $\underline{v}_i$  in our model is assumed to have a gamma distribution with mean 1 and variance 1 or, alternatively, the variance is assumed to follow a gamma distribution with parameters  $(\sigma^{-2}, \sigma^{-2})$ :

$$g(\underline{v}) = \frac{(\sigma^{-2})^{\sigma^{-2}}}{\Gamma(\sigma^{-2})} .v^{\sigma^{-2}-1} .e^{-v.\sigma^{-2}} \quad (13)$$

The main reason of adding (13) is to capture unobserved heterogeneity. Taking (8), (10) - (13) together we are now able to specify our density function of our incomplete spells as in (9) (see for an elaboration on the derivations: Kooreman and Ridder, 1983):

For our first specification, we take for  $\underline{v} = 1$ , and  $\tau(t) = \alpha.t^{\alpha-1}$  as in (12):<sup>13</sup>

$$\tilde{h}_i(t) = \begin{cases} \frac{\alpha(e^{X\beta})^{1/\alpha}}{\mathbb{M}_{\mathbf{M}(\beta,\alpha)}(1/\alpha)} \exp(-e^{X\beta}t\alpha) & \text{For first generation immigrants} \\ \frac{\alpha(e^{X\beta})^{1/\alpha}}{\Gamma(1/\alpha)} \exp(-e^{X\beta}t\alpha) & \text{Other} \end{cases}, \quad (14)$$

where  $\mathbb{M}_{\mathbf{M}(\beta,\alpha)}(1/\alpha)$  is  $\Gamma(1/\alpha)$  times an incomplete gamma distribution (between zero and  $e^{X\beta}\mathbf{M}\alpha$ ), with parameters 1 and  $1/\alpha$ . Now for  $\underline{v} = g(\underline{v})$  and for the duration dependence constant ( $\tau(t) = 1$ ):

$$\tilde{h}_i(t) = \begin{cases} \frac{e^{X\beta}(1-\sigma^2)}{1 - \left(\frac{1}{1+\sigma^2 e^{X\beta}M}\right)^{\frac{1-\sigma^2}{\sigma^2}}} (1 + \sigma^2 e^{X\beta}t)^{-1/\sigma^2} & \text{For first generation immigrants} \\ e^{X\beta}(1-\sigma^2)(1 + \sigma^2 e^{X\beta}t)^{-1/\sigma^2} & \text{Other} \end{cases}. \quad (15)$$

And last for both  $\underline{v} = g(\underline{v})$  and  $\tau(t) = \alpha.t^{\alpha-1}$ :

$$\tilde{h}_i(t) = \begin{cases} \frac{\alpha(\sigma^2 e^{X\beta})^{1/\alpha}}{\mathbb{B}_{\mathbf{M}(\alpha,\beta,\sigma^2)}\left(\frac{1}{\alpha}, \frac{1}{\sigma^2} - \frac{1}{\alpha}\right)} (1 + \sigma^2 e^{X\beta}t\alpha)^{-1/\sigma^2} & \text{For first generation immigrants} \\ \frac{\alpha(\sigma^2 e^{X\beta})^{1/\alpha}}{\mathbf{B}\left(\frac{1}{\alpha}, \frac{1}{\sigma^2} - \frac{1}{\alpha}\right)} (1 + \sigma^2 e^{X\beta}t\alpha)^{-1/\sigma^2} & \text{Other} \end{cases}, \quad (16)$$

with  $\mathbb{B}_{\mathbf{M}(\alpha,\beta,\sigma^2)}\left(\frac{1}{\alpha}, \frac{1}{\sigma^2} - \frac{1}{\alpha}\right)$  defined as  $\mathbf{B}\left(\frac{1}{\alpha}, \frac{1}{\sigma^2} - \frac{1}{\alpha}\right)$  times an incomplete beta distribution function between zero and  $1/\{1 + \sigma^2 e^{X\beta}M\alpha\}$ , with parameters  $1/\alpha$  and  $(1/\sigma^2 - 1/\alpha)$ , and finally  $\mathbf{B}\left(\frac{1}{\alpha}, \frac{1}{\sigma^2} - \frac{1}{\alpha}\right)$  denoting the beta function:

$$\mathbf{B}\left(\frac{1}{\alpha}, \frac{1}{\sigma^2} - \frac{1}{\alpha}\right) = \int_0^1 x^{\frac{1}{\alpha}-1} (1-x)^{\frac{1}{\sigma^2}-\frac{1}{\alpha}-1} dx = \frac{\Gamma\left(\frac{1}{\alpha}\right)\Gamma\left(\frac{1}{\sigma^2} - \frac{1}{\alpha}\right)}{\Gamma\left(\frac{1}{\sigma^2}\right)} \quad [x \in (0,1)].$$

(14) - (16) can be estimated by means of maximum likelihood. The loglikelihood can be easily constructed, by splitting the distribution functions of the incomplete spells in two separate loglikelihoods and adding both loglikelihoods up for both first generation immigrants and for both non-first generation immigrants. In the following subsection we will show the estimation results for employment spells of the Turks, Moroccans, Antilleans, Surinamese and the Dutch.

<sup>13</sup>We do not consider the most straightforward specification here, namely the exponential case:  $1 - F(t) = \exp(-\lambda_i(\cdot)t)$ . Because the exponential distribution is a no-memory process, differentiating between first and second generation immigrants does not make a difference here, which may give biased results.

## 4.2 Results

Figure 4 shows the histogram of all employment spells in our data set. The stepsizes between the separate bars is 20 months and the total number of observations is 1841. Although we are working with duration models, empirical survival functions – like e.g. Kaplan-Meier estimates – are unfortunately not attainable, because we only have right censored spells. Therefore, only direct information about the length of the spells is available. Figure 4 displays a remarkable amount of short spells, which is rather striking because censored data should provide us with an overrepresentation of long employment spells. We can observe that more than 25 percent of the individuals in our database have started their job less that 20 months before the moment of sampling. Thus, it seems that the individuals in our database have a high probability to leave employment rather soon after the commencement of their job. In terms of duration models: the hazard of leaving the current job is high and seems to be increasing in time (certainly for the first few years). This does not necessarily mean that people who leave their job become unemployed. It may also be the case that they receive a better job offer and accept it or that they leave their job for early retirement.

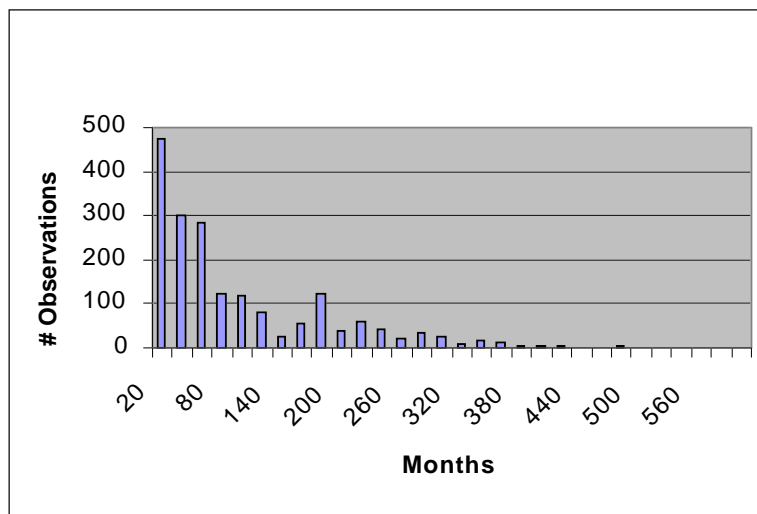


Figure 4: A histogram of employment spells.

Because we conjecture our data to possess positive duration dependencies, we will estimate first the MPH model of (8) with  $\tau(t)$  defined as in (12) and  $g(\underline{v}) = 1$ , so the model does not contain heteroskedasticity. The second column in table 3 presents the maximum likelihood estimation results from the likelihood results as defined in (14).

The estimates presented in table 3 are satisfactory and according to intuition, except for two cases. The dummies for the ethnic groups are not significant, while one would expect them to be and the duration parameter is – while positive as was to be expected – rather high. The last characteristic would imply that individuals in the sample witness their probabilities to leave employment quickly increasing in time.

The low (insignificant) parameter value for the clustering effect is not surprising. The size of the area – namely on the level of cities – seems to be too large to capture network or peer effects, although table A2 in the appendix shows that there is still some variation in clustering behavior between the separate cities. Table A2 also shows that there is a large difference in clustering behavior between the four largest cities (Amsterdam, Rotterdam, The Hague and Utrecht) and the nine smaller cities in the Netherland. An insignificant value of the clustering variable may also indicate that there is no difference in the labor market performance between smaller or larger cities among the ethnic groups.

The coefficient of the language variables, ‘understanding’ and ‘practice’, have the expected sign and are statistically significant. Recall that a high value of the variable ‘understanding’ means that the individual has a high ability to actually understand Dutch, while a low value of the



variable ‘practice’ indicate a high level of using the Dutch language. So, individuals who have difficulties with understanding Dutch and are also not inclined to exercise the language have a higher probability to leave their job than others.

The coefficient of the network variables are also according to intuition, although the coefficient of the ‘within’ variable is only significant at a 10% percent level. From subsection 3.2 we know that low values of the ‘between’ and ‘within’ variables indicate strong social networks between the ethnic group and the Dutch and within the ethnic group itself. Table 3 shows that individuals who have a strong social network intertwined with that of the Dutch have a low probability to leave their current job. Similarly, persons who have a strong social ethnic network have a (marginally significant) higher probability to leave their job.

Table 3: Maximum Likelihood Estimation Results of the MPH Models for Employment Spells.

	With duration dependence	
<i>Individual characteristics</i>		
Constant	-9.25	(1.60)
Experience	-0.01	(0.00)
Age	-0.07	(0.02)
Generation	17.41	(20.99)
Female	0.33	(0.07)
Turks	-16.18	(21.17)
Moroccans	-16.27	(21.20)
Surinams	-16.21	(21.14)
Antillians	-15.56	(21.24)
<i>Language variables</i>		
Understanding	-0.68	(0.40)
Practice	0.31	(0.14)
<i>Clustering &amp; Network variables</i>		
Between	0.45	(0.19)
Within	0.28	(0.22)
Clustering	0.00	(0.01)
<i>Duration dependence parameter</i>		
$\alpha$	2.63	(0.37)
Mean log likelihood	-4.31	
$N$	1841	

Explanatory note: Standard errors are in parentheses.

The remaining coefficients seem according to theory, so we briefly address the outcomes of the individual characteristics. As could be expected, the higher the experience the less inclined one would be to leave a job. This phenomenon may be due to rent-seeking behavior. Also age seems to decrease the probability to leave a job. This could be caused by the fact that pensions in the Netherlands are endangered when changing jobs; something which becomes more important when getting older. On the other hand, the age effect could just as well be caused by the rent-seeking process. And last, it seems that females have higher propensities to leave their current job. This would not only be caused by a weaker labor market position, but also by the fact that women may leave jobs to give birth.

To sum up, preliminary results indicate that our latent language and network variables are indeed important to explain part of the unfavorable labor market position of ethnic minorities. Moreover, the developed clustering variable is almost zero, indicating similar labor market positions of ethnic minorities in the surveyed 13 cities in the Netherlands.

## 5 Conclusions and Further Research

As was already known, there seems to be wide differences in the characteristics across ethnic groups in the Netherlands. These differences become apparent in demographic variables, educational attainment and in labor market characteristics. It is widely recognized that part of the causality for unfavorable labor market and educational achievement characteristics may be due to the working of social networks and the correlated language proficiencies.

This paper shows that with the use of correspondence and factor analysis new latent variables, describing the same language proficiencies and social networks, can be easily constructed. This approach is especially useful when working with large surveys, containing a large number of (dummy) variables describing the same characteristics. Moreover, the created latent variables are independent of each other, which facilitates further analysis. In the empirical analysis of this paper there is shown that the newly constructed variables have a high explanatory power (except for the clustering variable), and that they are statistically significant with the expected sign. Still, the empirical results are rather preliminary, but they point in the right direction.

The trajectories of further research are quite clear. Next to looking at employment spells, most researchers are also highly interested in the duration of unemployment spells. Taken together with the fact that the unemployment rates of the four ethnic group were substantially higher than that of the indigenous population, gives us enough reason to perform the same analysis of subsection 4.1, but then on unemployed individuals in our dataset. This would also circumvent one of the intrinsic disadvantages of investigating employment spells, namely the uncertainty about the transition. If an unemployment ends, then it is in almost all cases beneficial for the individual, while the end of an employment spell may be due to a resignation or to a better job offer.

Furthermore, the estimations could be extended in the form of allowing heteroskedasticity and of more flexible duration dependence specifications. As already indicated, data on residential areas should be more detailed, preferably on the level of neighborhoods to capture spillover effects between immigrants. In the ideal case, tracts should be developed with equal size and a low number of inhabitants. However, the chances on obtaining data on residential areas are severely limited in the Netherlands.

Another interesting line of research would be to investigate the returns to education for ethnic groups. If returns to education are lower than that of the Dutch, incentives for high educational achievement could also be subdued. Moreover, if the ethnic ties are high and the binding with the Dutch low, children from ethnic communities could be less inclined to perform as well as the Dutch in a Dutch schooling system. This will allow us then to answer the research question: how easy is it for ethnic minorities in the Netherlands to get a job and what is the job worth conditional on their education?

## References

- [1] van Amersfoort, H. and C. Cortie, Social Polarization in a Welfare State? Immigrants in the Amsterdam region, *New Community*, vol. 22, no. 4, pp. 671-687, 1996.
- [2] Bénabou, R., Heterogeneity, Stratification, and Growth: Macroeconomic Implications of Community Structure and School Finance, *The American Economic Review*, vol. 86, no. 3, pp. 584-609, 1996.
- [3] Berkowitz, S.D., *An Introduction to Structural Analysis: The Network Approach to Social Research*, Butterworth, Toronto, 1982.
- [4] Bourdieu, P., *Homo Academicus*, CA:Stanford University Press, Stanford, 1988.
- [5] Carrington, W.J., E. Detragiache and T. Vishwanath, Migration with Endogenous Moving Costs, *The American Economic Review*, 1996.
- [6] Chiswick, B.R. and P.W. Miller, Language Skills and Earnings among Legalized Aliens, *Journal of Population Economics*, vol. 12, pp. 63-89, 1999.

- [7] Claussen, S., *Applied Correspondence Analysis*, Sage University Paper series on Quantitative Applications in the Social Sciences, 07-121, Thousand Oaks, CA: Sage, 1998.
- [8] Durlauf, S.N., Spillovers, Stratification, and Inequality, *The European Economic Review*, vol. 38, pp. 836-845, 1994.
- [9] Ellison, G. and E.L. Glaeser, Geographical Concentration, *Journal of Political Economy*, vol. 105, no. 5, pp. 889-927, 1997.
- [10] Farrell, J. and G. Saloner, Installed Base and Compatibility: Innovation, Products announcement, and Predation, *The American Economic Review*, vol. 76, no. 5, pp. 940-955, 1986.
- [11] Gorter, C., *The Dynamics of Unemployment and Vacancies on Regional Labour Markets*, Ph.D. thesis, Tinbergen Institute, Amsterdam, 1991.
- [12] de Graaff, T., C. Gorter and P. Nijkamp, Spatial Clustering and Segregation of Immigrants: Statistical Evidence for the Netherlands, forthcoming, 2000.
- [13] Granovetter, M., *Getting a Job: A Study of Contacts and Careers*, Harvard University Press, Cambridge, 1974.
- [14] Greenacre, M. and J. Blasius, *Correspondence Analysis in the Social Science*, Academic Press Ltd., London, 1994.
- [15] van der Heijden, P.G.M. and J. de Leeuw, Correspondence Analysis used Complementary to Loglinear Analysis, *Psychometrika*, vol. 50, no. 4, pp. 429-447, 1985.
- [16] Jackson, J.E., *A User's Guide to Principal Components*, John Wiley & Sons Inc., New York, 1991.
- [17] Katz, M. and C. Shapiro, Network Externalities, Competition and Compatibility, *The American Economic Review*, vol. 75, no. 3, pp. 423-440, 1985.
- [18] Kiefer, N., Economic Duration Data and Hazard Functions, *Journal of Economic Literature*, vol. 26, pp. 646-679, 1988.
- [19] van der Klaauw, B., *Unemployment Duration, Determinants and Policy Evaluation*, Tinbergen Institute Research Series, Ph.D. thesis, Amsterdam, 2000.
- [20] Kooreman, P. and G. Ridder, The Effects of Age and Unemployment Percentage on the Duration of Unemployment: Evidence from Aggregate Data, *European Economic Review*, vol. 20, pp. 41-57, 1983.
- [21] Lancaster, T., *The Econometric Analysis of Transition Data*, Cambridge University Press, Cambridge, 1990.
- [22] Manski, C.F., Identification of Endogenous Social Effects: The Reflection Problem, *Review of Economic Studies*, no. 60, pp 531-542, 1993.
- [23] Mincer, J., *Schooling, Experience and Earnings*, Columbia University Press, New York, 1974.
- [24] Montgomery, J, Social Networks and Labor Market Analysis: Towards an Economic Analysis, *The American Economic Review*, vol. 81, no. 5, pp. 1407-1418, 1991.
- [25] Rees, A., Information Networks in Labor Markets, *American Economic Review*, no. 56, pp. 559-566, 1966.
- [26] Tesser, P.T.M., J.G.F. Merens and C.S. van Praag, *Rapportage Minderheden*, Sociaal Cultureel Planbureau, Den Haag, 1999.

## 6 Appendix

The following table gives several categories and the number of persons, belonging to a specific ethnic group, from our sample that fit in that category.

Table A1: A contingency table for the characteristics of the five ethnic groups.

	Turks	Moroccans	Surinamese	Antilleans	Dutch	Total
<b>Wage Variables</b>						
Wage < 1100 p.m.	42	62	87	130	108	429
Wage 1100-1500 p.m.	68	67	126	104	142	507
Wage 1500-1900 p.m.	284	307	172	103	96	962
Wage 1900-2300 p.m.	132	106	94	44	97	473
Wage 2300-2700 p.m.	73	49	49	43	99	313
Wage 2700-3100 p.m.	25	11	37	36	88	197
Wage 3100-3500 p.m.	10	4	15	11	42	82
Wage 3500-3900 p.m.	1	1	9	8	51	70
Wage > 3900 p.m.	5	0	13	9	94	121
<b>Demographic Variables</b>						
Female	101	81	406	292	359	1239
Married	629	654	286	146	452	2167
No children at home	191	279	393	345	732	1940
Partner at home	133	196	439	386	457	1611
Age > 65	9	12	51	15	191	278
<b>Education</b>						
None or low	496	584	284	131	201	1696
Low	160	92	255	189	257	953
Mediate	91	62	178	150	269	750
High	27	20	87	106	303	543
<b>Miscellaneous</b>						
Unemployed	430	489	397	303	419	2083
Never had a job	88	117	119	146	33	503
Has a religion	774	751	629	377	407	2938
<b>Total</b>	<b>3767</b>	<b>3944</b>	<b>4126</b>	<b>3074</b>	<b>4897</b>	<b>19810</b>

The next table gives the coefficients of our gamma coefficient for the 13 cities surveyed in our data.

Table A2: Gamma coefficient of the 13 cities in the SPVA in decreasing order of magnitude, measured at 1 January 1995.

Cities	City size	$\gamma_i$ coefficient	Cities	City size	$\gamma_i$ coefficient
Amsterdam	722,230	0.26	Almere	104,541	0.01
Rotterdam	598,239	0.15	Delft	92,457	0.01
The Hague	442,937	0.16	Alphen aan den Rijn	67,113	0.00
Utrecht	235,625	0.05	Bergen op Zoom	47,915	0.01
Eindhoven	196,693	0.01	Tiel	34,171	0.01
Enschede	148,034	0.02	Hoogeveen	33,676	0.00
Dordrecht	114,089	0.01			