ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics

The effect of language proficiency on unemployment duration. Evidence from the Netherlands.

Bachelor Thesis

Abstract
This study examines the effect of language proficiency on unemployment durations. Reduced language skills are hypothesized to reduce job search effectiveness and therefore lead to longer unemployment durations. Using data from the LISS (Longitudinal Internet Studies for the Social sciences) panel administered by CentERdata, logistic regression models are estimated. To avoid biases, competing risks- and multilevel aspects are added to the analysis. While language problems are found to significantly reduce the probability of exiting unemployment to paid employment when no controls for origin are used, this significance disappears when controls for origin are added. Stratified models are also estimated to test whether the effects of language proficiency differ for natives and first- and second generation immigrants. These stratified models show no significant effects. The results of this study thus suggest that language proficiency does not affect unemployment durations. Limitations of this study, which are mostly data-related are discussed in the last section.

Name: Sjors Hoogeveen
Student number: 412903
Supervisor: Prof. Dr. H.D. Webbink
Second reviewer: Prof. Dr. E.M. Bosker
Date: 23 August 2018

Table of contents

1 Introduction

Today, immigration is a very important topic. In the recent decennia worldwide immigration has rapidly increased (United Nations, 2017). With the increase of immigration, public concern about the consequences of the arrival of great numbers of immigrants has also been growing. In many countries, immigration policy takes an important place in politics. Examples of the prominent role which immigration has taken in public debate are anti-immigrant rhetoric among politicians such as Trump (USA), Le Pen (France) or Wilders (The Netherlands), the Brexit,[1] and the recent German cabinet crisis (see: C., 2018).

Many objections people have against increased immigration are based on economic reasons (see: Sides & Citrin, 2007). Immigrants are believed to lower wages for natives, 'steal' native's jobs or they are believed to rely more heavily on welfare benefits, thus imposing a net fiscal cost. Another commonly heard objection is based on the belief that immigration increases crime. All of these objections have some relationship with the labour market. Naturally, individuals only rely on welfare benefits if they are unemployed and there is a consensus that unemployment tends to increase the probability of criminal behaviour (Baumann & Engelhardt, 2016). Since labour market perspectives and outcomes are important determinants of the fiscal costs of immigration, immigration related crime and the effect of immigration to GDP (OECD, 2014), it is important to study immigrant labour market outcomes.

Numerous studies have been conducted on the labour market outcomes of immigrants. In many studies, special attention is given to the effect of language proficiency on labour market outcomes. Ever since the work of Gary S. Becker (1964), human capital is assumed to play an important role in labour market outcomes. Productivity enhancing traits and characteristics, such as education and labour market experience, constitute an individual's human capital stock. Language skills can be seen as a special type of human capital. Like all human capital, language proficiency has a direct positive impact on productivity because it enables efficient communication with both customers and colleagues (Chiswick & Miller, 2003). Not only does language proficiency enable efficient communication, but it might also lead to increased job performance because it deepens the level of cultural knowledge and it enables social integration (Bird & Dunbar, 1991). Language skills are however a special type of human capital because they are complementary to other aspects of human capital. This means that the productivity enhancing effect of 'regular' human capital partly depends on language proficiency (Berman, Lang, & Siniver, 2003; Di Paolo & Raymond, 2012; Esser, 2006; Friedberg, 2000). An individual who lacks language skills is less able to efficiently make use of his

[1] Several polls show that immigration was one of the main motivations to vote for an exit from the European Union (Ipsos MORI, 2016; The Economist, 2016).

human capital to enhance his productivity than an individual who has well developed language skills (Chiswick & Miller, 2003). The last important aspect of language skills is that they are country-specific. The ability to speak Dutch is less valuable in the US than it is in the Netherlands. For that reason, language skills are given special attention when the labour market position of immigrants is researched. Most immigrants will have low Dutch language skills at the time of their arrival in the Netherlands. Especially relative to natives, immigrants are likely to have more trouble with the Dutch language. Many authors have researched whether the difference in language proficiency explains differences between labour market positions and outcomes of immigrants and natives.

The impact of language proficiency on two major labour market aspects has been widely researched. In much of the economic literature a positive effect of language skill on wage earnings has been observed (Blackaby, Leslie, & Murphy, 1998; Budría & Swedberg, 2015; Grand & Szulkin, 2002; Kee, 1995; Leslie & Lindley, 2001; Mcmanus, Gould, & Welch, 1983; Trejo, 1997). Aside from this positive effect on wage earnings, language proficiency is found to increase the probabilities of being employed for immigrants (Aldashev, Gernandt, & Thomsen, 2008; Blackaby et al., 1998; Dustmann, Fabbri, Preston, & Wadsworth, 2003; Leslie & Lindley, 2001). It must be noted that some researchers however, have found no effect of language proficiency on labour market outcomes, most notably in the Netherlands (Yao & van Ours, 2015).[2]

An individual's labour market position is not only determined by his earnings and his probability of having a paid job. Another principal aspect of the labour market is unemployment duration. Unemployment duration seems to be especially important since in modern Western economies, the unemployment rate is to a great extent determined by unemployment duration (Layard, Nickell, & Jackman, 2005a). Layard *et al.* (2005a) argue that the rise in unemployment rates in Europe is not primarily a consequence of a growing labour supply relative to labour demand, but that it is mainly a consequence of longer average unemployment spells.[3]

Unemployment duration influence to what extent individuals rely on welfare benefits, since longer unemployment durations lead to a longer time of dependence on benefits. Welfare expenditure therefore goes up as unemployment durations increase, especially since the

[2] Yao and Van Ours (2015) give the fact that English is a widely used- and known language in the Netherlands as a possible explanation for the lack of an effect of Dutch language skill on labour market outcomes. In their view, being able to speak and read Dutch is not an absolute necessity because English has become a de facto *lingua franca* in the Netherlands.

[3] Layard et al. (2005a) demonstrate that the number of vacancies in many countries has been relatively stable while the unemployment rate has risen. This suggests that unemployment has risen without a large change in labour demand. Layard *et al.* (2005a) see this phenomenon as evidence for the large role of unemployment durations in the determination of the unemployment rate.

unemployment rate is largely determined by average unemployment duration. Asides from its effects on the unemployment rate and welfare expenditure, unemployment duration also affects labour market perspectives (Blanchard & Diamond, 1994; Corak, 1996) and it plays a role in personal well-being (Mckee-Ryan, Wanberg, & Kinicki, 2005). For immigrants more specifically, longer unemployment spells could hamper social- and economic integration (Aycan & Berry, 1996).

Unemployment durations are partly determined by wage factors – most notably reservation wages (Jones, 1988) and a matching labour demand and supply (Layard, Nickell, & Jackman, 2005b),[4] but there are other important factors which influence unemployment durations. Especially job search effectiveness is thought to be an important determinant of unemployment duration (Layard et al., 2005a). Job search effectiveness constitutes everything that determines the speed of finding a job. Among other things, the speed of finding vacancies, intensity and costs of job search and hiring practices of employers influence job search effectiveness. In this thesis, language proficiency is hypothesized to play a role in job search effectiveness and the probability of finding employment, especially for immigrants (Chiswick & Miller, 2003).

Research has shown that immigrants do tend to face longer spells of unemployment compared to natives (Frijters, Shields, & Price, 2005; Kogan, 2004; Uhlendorff & Zimmermann, 2006), but very little research has explicitly been done on the effect of host country language skills on unemployment duration.[5] Delander *et al.* (2005) found that Swedish language training did reduce immigrant unemployment duration and Clausen *et al.* (2009) found similar results for immigrants in Denmark. In a more general study, McQuaid (2006) found that not only for immigrants, but also for natives in the United Kingdom, higher self-perceived verbal skills significantly increased job search success. Not much research on the effect of language proficiency on unemployment duration is available, and for the Netherlands no such study on the effect of Dutch been conducted.

This thesis will focus on the effect of language proficiency on unemployment duration. Higher language skills are hypothesized to increase job search effectiveness and thus lead to a lower unemployment duration. This positive effect of language proficiency is hypothesized to affect unemployment durations for both immigrants

[4] Both regional matching and skill matching play a role in the determination of unemployment duration. For example, Ahn, De la Rica & Ugidos (1999) found that willingness to move for a job significantly lowered unemployment duration, and McQuaid (2006) found that both skill- and special mismatch were important factors in job search success.

[5] There is some research available on the effects of bilingualism on the unemployment duration. Eam Lin and Bakar (2004) found that high test scores in English classes reduced the unemployment duration of Malay students. In Finland, the Swedish speaking minority face shorter unemployment spells, which Saarela and Finnäs (2003) among other things impute to the fact that the Swedish speaking minority tends to be bilingual – they speak both Swedish and Finnish. In these studies however, the second language has a more complementary character. The ability to speak English in Malaysia or Swedish in Finland is valuable, but these languages are not the primarily used languages.

and natives, since job search effectiveness is likely affected by language proficiency for both natives and immigrants. For example, there seems to be no particular reason to assume that a native with problems reading and writing the Dutch language will be more successful finding vacancies in papers or properly writing application letters than an immigrant with similar characteristics and comparable circumstances. The main hypothesis of this thesis is:

> *Individuals without any Dutch language related problems have a higher hazard rate of exiting unemployment to paid employment.*

This main hypothesis is thus concerned about individuals, and not only about immigrants. The data that will be used contains observations on unemployment spells of both natives and foreign born immigrants, which allows this general hypothesis to be researched. Both of these groups reported to have problems with the Dutch language, although a much higher percentage of immigrants reported to have problems with the Dutch language. This more general hypothesis is especially relevant for questions about immigrant labour market outcomes, since immigrants likely to have low language skills – but it is also relevant for native related language problems such as illiteracy.

For various reasons – which will be discussed later on – the assumption that having problems with the Dutch language has (equal) negative effects for both natives and immigrants might not hold. Only immigrants might face longer unemployment durations as a consequence of having language problems or the effects of language proficiency might be different for immigrants and natives. Therefore, a secondary hypothesis is formulated:

> *Foreign born immigrants without any Dutch language related problems have a higher hazard rate of exiting unemployment to paid employment.*

To test these hypotheses, analyses will be conducted for the full sample and for different strata based on immigrant status separately.

The outline of this thesis is as follows. In the next paragraph, a theoretical framework will be given and the ways in which language proficiency is hypothesized to affect unemployment duration will be discussed in more detail. The used methods will then be given and the data will be examined. In paragraph 5, the results of the used regressions methods are shown and briefly explored. These results will be interpreted in paragraph 6, were possible issues of this study with the data and methods will be discussed as well.

2 Theoretical framework

The duration T of a spell of unemployment is determined by the conditional probability of exiting unemployment to paid employment at each month t. Higher probabilities of exiting unemployment lead to lower expected unemployment durations. The most common term for this conditional probability is the hazard rate. The hazard rate of exiting unemployment to paid employment $h_{i,t}$ for individual i at month t on the condition that the transition to paid employment has not taken place yet is defined as:

$$h_{i,t} = \Pr(T_i = t | T_i \geq t; X_{i,t}) \qquad (2.1)$$

where T_i represents the month at which individual i obtains paid employment after a spell of unemployment and $X_{i,t}$ is a vector of covariates that affect the probability of the transition from unemployment to employment. These covariates can either be constant or time-varying.

Layard *et al.* (2005a) proposed that two distinct categories of variables affect the probability of finding a job, where they give special attention to the variables which influence job search effectiveness.[6] Firstly, individual characteristics influence the probability of finding a paid job and secondly, labour market-wide variables such as competition among the unemployed for vacancies influence that probability. The latter will be discussed first.

Certain aspects of the labour market can affect unemployment duration. When there is a high demand for labour relative to labour supply, an unemployed individual will face a higher probability of exiting unemployment since there are fewer job-seekers per vacancy. Monthly unemployment rates are used as a proxy for competition among job-seekers to control for the time-varying labour market factors which influence unemployment duration. A high unemployment rate indicates lower demand relative to supply and is expected to have a negative effect on the hazard rate of transitioning to paid employment, because it increases competition among job-seekers (Bover, Arellano, & Bentolila, 2002; Layard et al., 2005a).[7] Unfortunately, more specific controls for time-varying labour market factors, such as regional- or sectoral unemployment rates cannot be used. Data on regional- or sectoral unemployment rates is available on macro-level, but no data on the individual level is available to link individuals to regions and, while there is some data available linking individuals to a sector, a lot of these observations are missing. Using sectoral unemployment rates would thus result in a large loss of data. Therefore, the nation-wide monthly unemployment rate is used as a proxy for competition among job-seekers.

As already briefly discussed, several individual characteristics are likely to affect the hazard rate of leaving unemployment. Among others, gender, age and education are used as control variables in this study. Many studies have found that these variables significantly affect

[6] Layard et al. (2005a) give some examples of factors that influence job search effectiveness, such as benefit duration regimes, employment protection legislation and duration structure. Many more possible influencing factors are possible.

[7] It would be better to use the number of unemployed individuals divided by the number of vacancies as a proxy for competition for vacancies. Unfortunately, no monthly data on the number of vacancies is available for the Netherlands.

unemployment duration (e.g. Bover et al., 2002; Kettunen, 1997; Stewart, 2001). Most notably, employment duration is assumed to affect the probability of exiting unemployment to paid employment itself and is therefore also used as a control variable. Those who are unemployed for a longer period are also less likely to find paid employment because employers seem to prefer applicants who experienced unemployment spells of small durations (Blanchard & Diamond, 1994). A full overview and the description of the individual control variables can be found in appendix A.

The main variable of interest for this study is language proficiency. Frijters *et al.* (2005) found that immigrants do face lower job search effectiveness when compared to natives. They suggest that limited language skills are the cause of the reduced job search success. In this study, language proficiency is hypothesized to affect the unemployment duration in multiple ways. Already mentioned, job search effectiveness is thought to play a large role in the determination of the probability of leaving unemployment. Increased language proficiency is hypothesized to increase job search effectiveness. Language proficiency is also hypothesized to affect skill-matching and job opportunities. These channels through which language proficiency is hypothesized to affect unemployment duration will now be discussed in more detail.

Firstly, language skills are likely to influence job search effectiveness, because individuals with lower language skills have higher search costs. They are likely to have trouble finding, reading and applying for vacancies for example in newspapers or on the internet. The speed of retrieving information about available jobs is likely to be lower for individuals with language problems. Such an individual has to put in more effort to find and apply for vacancies. This could hinder job search success.

In addition to this direct effect of language proficiency on job search effectiveness through a rate of finding information about vacancies, the lack of language skills might limit the available job search methods and could act as a barrier to use formal methods of finding employment, because of a limited knowledge of labour market institutions (Urwin & Shackleton, 1999). Strikingly, the use of those formal methods of finding employment, such as public employment services (UWVwerkbedrijf) or private employment agencies are reported to result in relatively high job finding rates (Urwin & Shackleton, 1999). Informal methods do however not automatically have to result in a lower probability of finding employment, but for immigrants with low language skills, the use of informal methods might be problematic. Finding jobs through a network of friends or relatives is the one of the most important informal job finding method. Individuals with a highly developed social network are more likely to exit unemployment through informal methods (Hannan, 1999). Social capital thus affects labour market opportunities and unemployment duration. Lancee (2010) however found that not all forms of social capital have equal effects on the probability of leaving unemployment. Especially for immigrants, structural inter-ethnic contacts and the adoption of native attitudes tend to improve employment probabilities, while contacts within an immigrant's own culture have no such effect. Immigrants who lack the ability to communicate in Dutch are less likely to form networks outside their own group or adopt native attitudes and could therefore face longer unemployment durations. Frijters *et al.* (2005) also found that immigrants have a lower probability of finding paid employment when relying on their social networks, while immigrants do tend to rely on their social networks to find employment. Lacking language skills might thus be a barrier to use formal methods of finding employment, while it also might decrease inter-ethnic social contacts and the adoption of native attitudes. Lacking language skills then leads to higher use of informal methods of finding employment, while those informal methods are likely to be less effective for those individuals who lack language skills. This effect could however be only relevant for immigrants and could therefore lead to differences in the effects of language proficiency between natives and immigrants.

Secondly, language skills are likely to influence job matching and employment opportunities. Someone with language problems is capable to perform in jobs were language is less needed, but he will be unable to have a job where language proficiency plays an important role. As a result, lacking language skills decreases the number of available jobs (Beggs & Chapman, 1990; Kossoudji, 1988; Peri & Sparber, 2009). McQuaid (2006) found that individuals – so not only immigrants – with lower self-perceived verbal skills do face a lower probability of exiting unemployment to paid employment. He interpreted this partly as a consequence of an increased demand for verbal- and communication skills. Lacking language skills can thus lead to skill mismatch and there will be fewer vacancies suitable for individuals with low language skills. Therefore competition for those vacancies will be higher. This is hypothesized to results in longer unemployment durations (Layard et al., 2005a).

Not only is higher language proficiency hypothesized to increase job search effectiveness and employment opportunities, the hiring rate is also likely to increase with language proficiency. If an employer has a choice between two nearly identical candidates, but only one of them has no language problems, it is not unlikely he will hire the proficient candidate.[8] Communication skills – and therefore language proficiency

[8] Blanchard and Diamond (1994) argued in their paper that choices made by employers affect the unemployment duration. They found employers tended to hire individuals with shorter unemployment spells, because they ranked individuals based on the length of their unemployment. In a similar way, employers could favour individuals with higher language skills. This can impact the hiring rate of those with low language skills, especially since high Dutch language proficiency is common. It is not unlikely that for every vacancy a job-seeker who speaks and reads Dutch is available. Being unable to speak Dutch is the exception rather than the rule, which means that speaking Dutch is probably less viewed as a productive skill but more as a basic common skill. Being unable to speak Dutch is likely to be viewed as a severe drawback. Some research does suggest that lacking basic skills, such as illiteracy reduces employment probability

– are found to play an important role in the hiring decision of new employees (Lynch & Zemsky, 1995; Sims Peterson, 2009). Improper grammar and spelling in resumes could also result in a lower probability of being invited for a job interview (Thoms, Mcmasters, Roberts, & Dombkowski, 1999).

Having language problems is thus hypothesized to result in longer unemployment durations because of (i) lower job search effectiveness, (ii) increased competition for suitable jobs for individuals who lack language skills and (iii) a lower hiring rate.

3 Methodology

In this thesis, the effect of having language problems on the duration of an unemployment spell will be analyzed. Analysis of duration data using ordinary least squares (OLS) regression techniques or probit or logit models faces some problems, primarily because of censoring and time-varying covariates (Jenkins, 2005, pp. 8–10). Censoring can occur if an individual was unemployed when he entered the study, which is called left censoring. Right censoring occurs when the event of interest has not happened during the study. Censoring will be more thoroughly discussed in paragraph 3.4.

Because of censoring and time-varying covariates, survival analysis methods will be used to analysis unemployment duration. These methods can handle right censoring, (Guo, 2010) and time-varying covariates can be easily incorporated into the analysis. Survival analysis methods were first developed by biomedical researchers who were interested in studying mortality (Guo, 2010). Expressions such as ‘survival’ or ‘risk’ can cause some confusion in an economic context. For clarity, some expression will be shortly defined. An individual’s survival refers to an individual remaining unemployed. While unemployed, that individual is at risk of transitioning to paid employment, which is the event of interest for this study.

In survival analysis the probability that the length of unemployment T is less than time t is given by the cumulative distribution function:

$$F(t) = \Pr(T \leq t) = \int_0^t f(x)dx \tag{4.1}$$

of survival time T, which has the probability density function:

$$f(t) = \lim_{dt \to 0} \frac{\Pr(t \leq T < t + dt)}{dt} \tag{4.2}$$

The survivor function $S(t)$ gives the probability that an individual has not made the transition to employment at time t.

$$S(t) \equiv 1 - F(t) = \Pr(T > t) = \int_t^\infty f(x)dx \tag{4.3}$$

The key function of survival analysis is the hazard function, which can be expressed as follows:

(Mcintosh & Vignoles, 2001). Lacking language proficiency might in similar way reduce the probability of getting hired.

$$h(i,t) = \lim_{dt \to 0} \frac{\Pr(t \leq T_i < t + dt | T_i \geq t)}{dt} \quad (4.4)$$

This hazard function can be understood as a measure of the instantaneous rate of change from unemployment to employment at time t. The numerator of this function gives the probability that the transition to employment will occur between t and $t + dt$, on the condition that transition has not yet occurred at time t. The hazard function $h(i,t)$ can be written as a function of $f(t)$, $S(t)$ and $F(t)$:[9]

$$h(i,t) = \frac{f(t)}{S(t)} = \frac{f(t)}{1 - F(t)} \quad (4.5)$$

3.1 Measurement differences

Interestingly, both individuals with a Dutch background and a foreign background reported to have language problems. Table 3 shows the distribution of language problems by origin. In total, 14.07% of individuals reported to have Dutch language problems. The percentages of especially individuals with a first generation foreign origin are much higher. As already discussed in the introduction, the analyses will be done for both the full sample and for three strata of the full sample separately. These strata are: natives, first generation immigrants and second generation immigrants.

Table 3. Language problems by origin

	Language problems			
Origin	Yes	No	Total	% of total
Dutch background	132	1.135	1.267	10.42
First gen. foreign, Western	26	33	59	44.07
First gen. foreign, non-Western	51	45	96	53.13
Second gen. foreign, Western	7	78	85	8.24
Second gen. foreign, non-Western	3	47	50	6.00
Total	219	1.338	1.557	14.07

Treating self-reported language problems for different groups might however be inappropriate. It is not unlikely that individuals with a Dutch background use a different definition of having problems with the Dutch language than first- or second generation immigrants. For example, a native Dutchman might report to have problems with reading the Dutch language because he has trouble reading literature, while a migrant might report to have problems because he lacks knowledge on basic vocabulary and cannot read a newspaper.[10]

Because of these differences, for each method that will be discussed in the next section, the best fitting models will also be estimated separately for individuals who reported to have a Dutch background and individuals who reported to be first- or second generation immigrants.

[9] For a proof of this, see: (Guo, 2010; Jenkins, 2005)

[10] Another difference between native- and immigrant self-reported language proficiency is caused by language anxiety. There is evidence that language anxiety – the feeling of unease when using a foreign language – affects self-reported language proficiency (MacIntyre et al., 1997). As natives are not affected by language anxiety, there self-reported language proficiency likely differs from immigrants self-reported language proficiency.

3.2 Kaplan-Meier estimator

A non-parametric model of the survival function can be estimated using the Kaplan-Meier estimator (Kaplan & Meier, 1958):

$$\hat{S}(t) = \prod_{i:t_i \leq t} (1 - \frac{d_i}{n_i}) \qquad (4.6)$$

Where t_i is a time where at least one individual moved from unemployment to employment, d_i is the number of transitions to employment at time t_i and n_i is the number of individuals who remain unemployed or have been censored at time t_i. For both the unemployed with- and without Dutch language problems, survivor functions can be estimated using Kaplan-Meier estimates. To see whether these Kaplan-Meier estimates are significantly different, a log-rank test will be conducted. The main hypothesis assumes that the effects of language proficiency are equal for all groups, so that different origins do not result in different effects of language proficiency on the job search effectiveness and the probability of exiting unemployment to paid employment.

3.3 Discrete-time survival analysis

The Kaplan-Meier estimates and the log-rank tests provide a good starting point for the analysis but they are insufficient to provide a complete test for the hypotheses. The main problem of analysis based on the Kaplan-Meier estimates and the comparison of different estimates is that it assumes homogeneity and the log-rank test does not control for additional (time-varying) covariates (Guo, 2010, p. 52; Meyer, 1990). Further multivariate analysis is necessary to test the hypotheses. To estimate the effect of language proficiency on unemployment duration, discrete-time survival methods will be used. Discrete-time survival analysis was first developed by Allison (1982) and has since then been used in multiple studies on unemployment duration (see for example: Détang-Dessendre & Gaigné, 2009; Gangl, 2003; Han & Hausman, 1990; Uysal & Pohlmeier, 2011). The slightly different discrete-time hazard rate is defined as:

$$h(i,t) = \Pr\,(T_i = t | T_i \geq t, X_{it}) \qquad (4.7)$$

which is the conditional probability of a transition to employment at the tth month of unemployment on the condition that this has not happened yet, given individual's i values of the independent predictor variables X_{it} for month t.

Throughout this thesis, logistic regressions are used to estimate the discrete-time models (see also Singer & Willett, 1993). The probability of a transition from unemployment to paid employment for individual i at month t is indicated by the dependent binary variable Y_{it}, which equals 1 if this event occurs and 0 otherwise. This probability is represented by the following nonlinear equation:

$$h(i,t) = \Pr(Y_{it} | X_{it}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 Z_{it} + \beta_2 X_{it})}} \qquad (4.8)$$

where X_{it} is the vector of independent variables affecting individual i at month t, Z_{it} is a function of time to allow for duration dependence and

β_0, β_1 and β_2 the unknown regression coefficients. This nonlinear probability function can be expressed as a generalized linear model by using the logit link function:

$$\begin{aligned} \text{logit}\{h(i,t)\} = \ &\ln\left\{\frac{h(i,t)}{1-h(i,t)}\right\} \\ &= \beta_0 + \beta_1 Z_{it} + \beta_2 X_{it} \end{aligned} \tag{4.9}$$

The regression coefficients β_0, β_1 and β_2 are estimated using maximum likelihood estimators.[11]

Goodness-of-fit will be tested using the Hosmer-Lemeshow test (Guo, 2010, p. 66; Hosmer & Lemesbow, 1980). The normal Pearson chi-squared goodness-of-fit test cannot be used for this data since the data contains a relatively small number of events per combination of values of covariates, while at the same time containing a large number of covariate patterns. For example, only one exit from unemployment is observed from an individual who: is female, is aged 34-44, has children living at home, has a HAVO/WVO diploma and has language problems. The Hosmer-Lemeshow test provides a solution to this problem by grouping observations based on their expected probability of exiting unemployment. The test than assesses whether the observed number of exits from unemployment per group is significantly different from the expected number of events. The Hosmer-Lemeshow χ^2 goodness-of-fit statistic is:

$$\chi^2 = \sum_{g=1}^{g} \frac{(y_g - m_g p_g)^2}{m_g p_g (1 - p_g)} \tag{4.10}$$

where y_g denotes the number of exits from unemployment for group g, m_g denotes the number of observations in group g and p_g the probability of an exit for group g.[12] The Hosmer-Lemeshow χ^2 has $g - 2$ degrees of freedom. Goodness-of-fit will also be assessed using Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC) and the likelihood-ratio test.[13]

The choice to use discrete- rather than continuous time survival methods is based on three reasons. Firstly, the data used is discrete (grouped) time data i.e. the survival times have been grouped into

[11] For a more complete description of the estimation of β using ML techniques, see Guo (2010)

[12] To determine how many groups should be used for the Hosmer-Lemeshow test, the suggestions by Paul, Pennell & Lemeshow (2013) were followed. Their recommended number of groups is given by:

$$g = \max\left(10, min\left\{\frac{m}{2}, \frac{n-m}{2}, 2 + 8\left(\frac{n}{1000}\right)^2\right\}\right)$$

where g is the number of groups, m is the number of successes (exits to paid employment) and n denotes the number of observations. Because the test is rather sensitive to changes in g, the test is also conducted with the standard number of 10 groups.

[13] The likelihood-ratio test assesses whether the null hypothesis that all covariates have no effect on the hazard rate holds. For low p-values, this hypothesis can be rejected, which means that the model does not provide a good fit when compared to no model.

monthly intervals.[14] This means that someone who is unemployed for the duration of one day is treated the same as someone who is unemployed for exactly one month – for both individuals an unemployment duration of one month is recorded. Discrete-time survival methods are suited for analyzing unemployment durations, especially since unemployment duration may partly be a discrete phenomenon (Han & Hausman, 1990; Uysal & Pohlmeier, 2011). Using discrete time analysis as opposed to continuous time analysis does however results in a loss of precision. This loss of precision does not necessarily lead to different estimates, especially since unemployment spells usually last for several months (Allison, 1982, 2014; Jenkins, 2005).[15] Discrete-time analysis is suitable for this thesis and the loss of precision does not make it less appropriate than continuous-time analysis.

Discrete-time analysis does provide some important benefits as opposed to continuous-time analysis. The data contains individual unemployment spells which begin at different dates. This means that time-varying covariates which measure labour market demands and competition for vacancies are very important for the analysis. Although continuous survival analysis methods allow time-varying covariates, it is more convenient to incorporate time-varying covariates when using discrete time analysis (Uysal & Pohlmeier, 2011).

Lastly, discrete time survival analysis allows some important issues to be relatively easily addressed which will be discussed in the following paragraphs. These issues are: multiple events, multiple unemployment spells per individual and unobserved heterogeneity (Han & Hausman, 1990).

3.4 Competing risks

Left and right censoring were very briefly addressed at the beginning of this chapter. Lefts censoring occurs when an individual is unemployed when he first entered the study. In this case, unemployment duration is only partially known, since the beginning of the unemployment spell is unknown. Because survival analysis methods cannot handle this kind of censoring, observations, were the starting date of the unemployment spell is unknown, have been dropped (Guo, 2010, p. 28).

Right censoring on the other hand occurs when the transition to paid employment has not occurred during the study. No transition will be observed if the individual remained unemployed at his last available observation. Attrition from the study does not seem to have a relationship with employment status and therefore is assumed to be

[14] This type of data is also known as *person-time data* (e.g. Guo, 2010, p. 58)

[15] No data on average unemployment duration is available for the Netherlands but the average unemployment duration for OECD countries between 2007 and 2017 was 8.99 months. It is unlikely that the average unemployment duration in the Netherlands shows a large difference.

random (De Vos, 2009). This type of right censoring is assumed to be noninformative and incorporating these censored observations into the analysis does not lead to bias (Allison, 2014).[16]

Informative right censoring, which can lead to bias in model estimation, is also likely to occur in this study. The event of interest for this study is the transition from unemployment to paid employment. Because of this, the unemployment duration of an individual who stops looking for employment without finding paid employment is labelled as right censored at the moment he exits unemployment: the event of interest has not occurred. Since individual exiting unemployment without finding a job are likely to be individuals who have a low probability of finding a job, this type of right censoring is informative and will lead to bias (Allison, 2014). In their study on unemployment duration Narendranathan and Steward (1993) also pointed out this problem. An unemployed individual faces two distinct risks: the risk of finding paid employment and the risk of abandoning their job search for other reasons. To estimate the effect of language proficiency on the duration of an unemployment spell which ends with the transition to paid employment, a competing risk model will be used. Using a multinomial logistic model multiple outcome events can be analyzed. In this thesis, three outcomes $k \in K = \{0,1,2\}$ are analyzed: $k = 0$: remaining unemployed/right censoring, $k = 1$: exit to paid employment and $k = 2$: exit without obtaining paid employment. In a multiple event setting the relative risk of exiting unemployment by outcome k is given by the following equation:

$$\Pr(Y_{it} = k|x_{it}) = \frac{e^{\beta_{k0}+\beta_{k1}Z_{kit}+\beta_{k2}X_{kit}}}{1 + \sum_{K=1}^{2} e^{\beta_{K0}+\beta_{K1}Z_{Kit}+\beta_{K2}X_{Kit}}} \quad (4.11)$$

Where X_{kit} represents a vector of independent variables, Z_{kit} is a function of time to allow for duration dependence and β_k represents the regression coefficient for the kth outcome. By taking outcome $k = 0$ as a reference group,[17] the nonlinear probability function can be expressed as a following generalized linear model by the logit link function:

$$\begin{aligned} \text{logit}\{Y_{it}\} &= \ln\left\{\frac{\Pr(Y_{it} = k|X_{kit})}{\Pr(Y_{it} = 0|X_{kit})}\right\} \\ &= \beta_{k0} + \beta_{k1}Z_{kit} + \beta_{k2}X_{kit} \end{aligned} \quad (4.12)$$

This general linear model estimates the natural logarithm relative risk of exiting unemployment by outcome k compared to remaining in unemployment (and right censoring). Using this multinomial logistic regression model, separate coefficients are measured for both outcomes.[18] To assess goodness-of-fit, a modified Hosmer-Lemeshow test for multinomial logistic regression models will be used (Fagerland

[16] Unfortunately, there is no test available to check whether censoring is noninformative (Guo, 2010, p. 29).

[17] The probability of the reference outcome $k = 0$ is then defined as:

$$\Pr(Y_{it} = 0|x_{it}) = \frac{1}{1 + \sum_{K=0}^{2} e^{\beta_{K0}+\beta_{K1}Z_{Kit}+\beta_{K2}X_{Kit}}}$$

[18] For other examples of the use of competing risks models in unemployment duration analysis see: Arntz & Wilke (2009) or Carling, Edin, Harkman & Holmlund (1996)

& Hosmer, 2012). Hausman tests will also be performed to test the important assumption of independence of irrelevant alternatives (IIA) (Hausman & McFadden, 1984). If this assumption is violated, multinomial logistic regression is not appropriate (Alba-Ramírez, Arranz, & Muñoz-Bullón, 2007).

3.5 Multiple unemployment spells

Until this point, it has been assumed that each individual only faced one spell of unemployment. Because of the nature of the data, it is possible that some individuals have faced multiple unemployment spells while being observed. Using (multinomial) logistic regression is inappropriate to analyze data with recurring events (Singer & Willett, 1993). An important assumption of the (multinomial) logistic regression model is independence of observations. The duration of multiple unemployment spells faced by one individual cannot be assumed to be independent. Unobserved individual characteristics are likely to influence hazard rates for each period of unemployment. Multiple unemployment spell durations from the same individual will therefore be correlated (Steele, 2005). In order to avoid this problem, multilevel multinomial logistic regression methods will be used. Repeated events lead to a two-level data structure. One level consists of individuals and the other level consists of the unemployment duration spells. The unemployment spells are nested within individuals. Each individual can face multiple episodes of unemployment. Figure 1 represents this multilevel model.

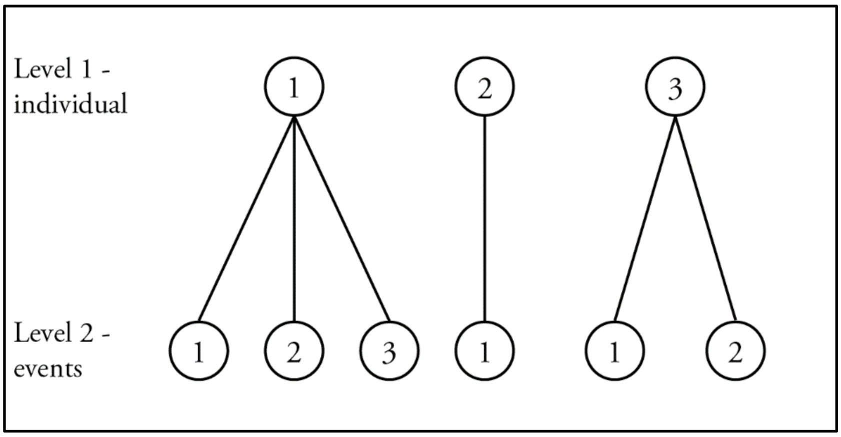

Figure 1. Graphic representation of multilevel data

In this figure, individual 1 underwent 3 spells of unemployment, individual 2 only 1 and individual 3 was observed to be unemployed 2 times. The relative probability that individual i facing unemployment spell j at month t transitioned out of unemployment by outcome k is:

$$\Pr\left(Y_{jit} = k \middle| X_{kjit}\right) = \frac{e^{\beta_{k0}+\beta_{k1}z_{kjit}+\beta_{k2}x_{kjit}}}{1+\sum_{K=1}^{2} e^{\beta_{K0}+\beta_{K1}z_{Kjit}+\beta_{K2}x_{Kjit}}} \tag{4.13}$$

which can be rewritten using the logit link function as:

$$\mathrm{logit}\{Y_{jit}\} = \ln\left\{\frac{\Pr(Y_{jit} = k)}{\Pr\,(Y_{jit} = 0)}\right\} \\ = \beta_{k0} + \beta_{k1}Z_{kjit} + \beta_{k2}X_{kjit} \tag{4.14}$$

Here $\Pr(Y_{jit} = k|X_{kjit})$ is the hazard of a k-type transition from unemployment at the tth month for individual i's jth spell of unemployment given covariates X_{kjit}. This multilevel multinomial logistic model is a version of the model used by Steele, Diamond and Wang (1996) in their paper on contraceptive use in China. In order to estimate the multilevel multinomial logistic model, Stata's Generalized Structural Equation Modelling (GSEM) will be used. Goodness-of-fit tests are unfortunately not available for GSEM.

3.6 Unobserved heterogeneity

The last problem to be addressed is the problem of unobserved heterogeneity. Unobserved characteristics of an individual are likely to affect his hazard rate of exiting unemployment to paid employment. It is highly unlikely that the independent variables that will be used in this study fully explain unemployment duration differences. For example, no control variables for reservation wages, benefit eligibility or personality characteristics are available for this study. Someone might also be exceptionally well performing at job interviews or he might have a large social network through which he can find a job quickly. Because of this unobserved heterogeneity, the observed unemployment durations and estimated logit coefficients could be different from the true risk pattern (Singer & Willett, 2003a). An individual unobserved random effect ε_i for individual i will therefore be added to the regression to control for unobserved heterogeneity:

$$\Pr(Y_{jit} = k|X_{kjit}) \\ = \frac{e^{\beta_{0k}+\beta_{1k}Z_{kjit}+\beta_{2k}X_{kjit}+\varepsilon_i}}{1 + \sum_{k=1}^{K} e^{\beta_{0k}+\beta_{1k}Z_{kjit}+\beta_{2k}X_{kjit}+\varepsilon_i}} \tag{4.15}$$

which can be rewritten using the logit link function as:

$$\mathrm{logit}\{Y_{jit}\} = \ln\left\{\frac{\Pr(Y_{jit} = k)}{\Pr(Y_{jit} = 0)}\right\} \\ = \beta_{k0} + \beta_{k1}Z_{kjit} + \beta_{k2}X_{kjit} + \varepsilon_i \tag{4.16}$$

The random intercept ε_i is assumed to follow a normal distribution $N(0, \sigma^2)$ (see also: Steele et al., 1996). This random effect can be estimated accurately as a consequence of the fact that multiple spells of unemployment per individual are observed (Singer & Willett, 2003a; Steele, 2005). Because of the limited number of individuals with recurrent unemployment spells, only a random intercept is estimated – no random slope effects are estimated – and random intercepts are assumed to be equal for different outcomes k, unemployment durations t and unemployment spells j. The random intercept model assumes the effects of the covariates to be equal for all individuals, only the intercept is allowed to vary. The regressions now consist of a fixed part, which is estimated by the coefficients $\beta_{k0}, \beta_{k1}, \beta_{k2}$ and a random part, which is estimated by ε_i and measures the variance between individuals that is not caused by the observed covariates, but by unobserved characteristics (Steele, 2005). The multilevel analysis discussed in the previous section allows random intercept to be taken into account for. To test whether the addition of a random intercept improves the competing risks model, a likelihood-ratio test will be conducted to compare the competing risks models with and without the random effect.

4 Data

The data is used from the LISS (Longitudinal Internet Studies for the Social sciences) panel administered by CentERdata (Tilburg University, The Netherlands). From November 2007 to February 2018, panel members are asked to complete online questionnaires every month. In total 25.098 individuals have participated in the (unbalanced) panel.[19] The Background Variables[20] are updated each month, while the more specific Core Studies on for example Religion and Ethnicity or Work and Schooling are conducted once every year.[21]

4.1 Unemployment duration

Unemployment spell duration is derived from the Background Variables. Monthly data on each individual's primary occupation is available. Individuals who either reported their primary occupation to be *Job seeker following job loss* or *First-time job seeker* are recorded as unemployed. From the first moment of unemployment, the number of months of unemployment is counted until the individual reports his primary occupation to be *Paid employment*, which is regarded as the transition from unemployment to paid employment ($k = 1$). If an individual reports his primary occupation to be anything else after a spell of unemployment, this is regarded as exiting unemployment without gaining paid employment ($k = 2$). If the individual remains unemployed this recorded as $k = 0$, which either indicates that an individual remains unemployed while he remains in the study or when he exits the study.

As noted earlier, survival analysis is unable to handle left censored unemployment spells. Unemployment spells which started before an individual entered the study therefore are dropped. In some cases where an individual first entered the study unemployed, data from the Work & Schooling study on the time since an individual was looking for a job could be used to determine the starting point of an unemployment spell. These observations were kept.

[19] There is attrition from the panel and periodically, refreshment samples have been recruited. On average 8000 individuals participated in each wave.

[20] These include variables such as sex, age, primary occupation and income among others.

[21] Because these more specific studies were only done once a year, the use of these variables was sometimes problematic, since changes were only recorded once a year. Therefore the use of these variables has been done so very carefully.

4.2 Language proficiency

Following Yao and Van Ours (2015), a binary indicator for Dutch language proficiency is created based on self-reported language problems. In the Religion and Ethnicity questionnaire each panel member is asked: *When having conversations in Dutch, do you ever have trouble speaking the Dutch language?* and *When reading newspapers, letters or brochures, do you ever have trouble understanding the Dutch language?* The possible answers are: (i) *no, never*, (ii) *yes, sometimes* and (iii) *yes, often have trouble/do not speak Dutch*. If an individual answered at least one of this questions with *yes, sometimes* or *yes, often have trouble/do not speak Dutch*, then the language proficiency indicator equals 1. It equals 0 otherwise.

Individuals with missing data on language proficiency are dropped from the analysis. Because questionnaires on language proficiency were only filled in once a year, some assumptions had to be made about these variables. The main assumption is that, once an individual reports to have no language problems, he is assumed to have no language problems in the following months were no data is available on his language proficiency. For individuals who reported to have language problems, it is assumed that he had problems in the previous months. So, if an individual reported to have language problems in 2015, he is also assumed to have had problems in 2014 or 2007 and all the previous years. If an individual reported to have no language problems in 2008, he is assumed to have no problems 2009 or 2012.[22] Only a limited number of individuals reported different language proficiency variables during their spell of unemployment. A change from having problems to having no problems was reported by 55 individuals, of which 46 had a Dutch origin, 4 were first generation foreigners and 5 were second generation foreigners. A change from having no problems to having problems has been reported by 48 individuals, of which 41 had a Dutch background, 4 were first generation foreigners and 3 were second generation foreigners. For these individuals, their changed language status has been used since the moment they reported it to be different than their previously reported proficiency.

4.3 Descriptive statistics

In total 1.577 spells of unemployment are used in the analysis. Of these 1.577 spells, 936 ended with a transition to paid employment, 371 ended with an exit from unemployment without finding paid employment and 250 spells are right censored. 1.143 individuals contributed to the 1.577 unemployment spells. The distribution of the unemployment spells over individuals is shown in table 1. Almost a quarter of these individuals experienced more than 1 spell of unemployment while participating in the panel.

Table 1. Distribution of unemployment spells

[22] This assumption is held, unless the individual for example reported to have no language problems in 2011, but that he did have language problems in 2012.

Number of spells	Individuals	% individuals	Total spells	% spells
1	869	76.03	869	55.81
2	191	16.71	382	24.53
3	56	4.90	168	10.79
4	14	1.22	56	3.60
5	5	0.44	25	1.61
6	3	0.26	18	1.16
7	2	0.17	14	0.90
8	2	0.17	16	1.03
9	1	0.09	9	0.58
Total	1.143	100.00	1.577	100.00

In table 2, some descriptive statistics of the data are given. The number of unemployment spells and median survival times is given for the full sample and for the sample split up based on language problems, origin, education, age and gender. Median survival times are defined as month t after which at least 50% of the individuals who will experience outcome k have experienced that outcome. For the full sample, the median survival time for an exit to paid employment is 5 months and median survival time for other exits from unemployment is 9 months. Individuals with language problems do have a higher median survival time than individuals without language problems, but the difference is only 1 month. More notably, first generation foreigners have more than double the median survival times relative to natives. Only individuals with a primary school education have longer median survival times relative to all other education categories. Unemployment durations also seem to rise with age, while gender seems to matter not so much. Median survival times for the other exits from unemployment are almost always higher than median survival times for exits to paid employment. This does suggest that individuals who exit unemployment without obtaining paid employment do indeed face a lower probability to get a paid job.

Table 2. Number of spells and median survival times grouped by and outcome individual characteristics

Characteristics	Exit to paid employment			Other exit		
	Amount	%	Median	Amount	%	Median
Language problems	118	12.60	6	76	18.06	7
No language problems	818	87.40	5	304	81.94	9
Native Dutch	790	84.40	5	304	81.94	8
First gen. foreign	72	7.69	12	37	9.97	12
Second gen. foreign	74	7.91	5	30	8.09	14
Primary school	26	2.78	8	20	5.39	5
VMBO	161	17.20	5	88	23.72	11
HAVO/WVO	96	10.26	4	43	11.59	8
MBO	272	29.06	5	92	24.78	10
HBO	225	24.04	5	89	23.99	8
WO	156	16.67	4	39	10.51	6
<24 years old	169	18.06	3	46	12.40	4
25-34 years old	231	24.68	5	45	12.13	4
35-44 years old	218	23.29	5	71	19.14	7
45-54 years old	198	21.15	6	76	20.49	13
55-64 years old	120	12.82	7	133	35.85	17
Male	423	45.19	5	189	50.94	7
Female	513	54.81	5	182	49.56	9
Full sample	936	100.00	5	371	100.00	9

In appendix B, life tables are reported for the full sample and for both the group of individuals with- and without language problems status.

The survival data is also graphically described. The earlier mentioned Kaplan-Meier estimator can be used to graphically show survivor or failure functions. In figure 2, Kaplan-Meier survival estimates are plotted for both the individual with- and without Dutch language problems. It can be seen that for all months, the survival function is higher for individuals with problems than individuals without problems. This graphical representation of the data suggests that individuals with language problems face a different, higher survival function and thus face longer unemployment durations. When tested using a log-rank test, the null hypothesis that both groups face the same survival function cannot be rejected at the 5% significance level ($p > \chi^2 = 0.1218$).

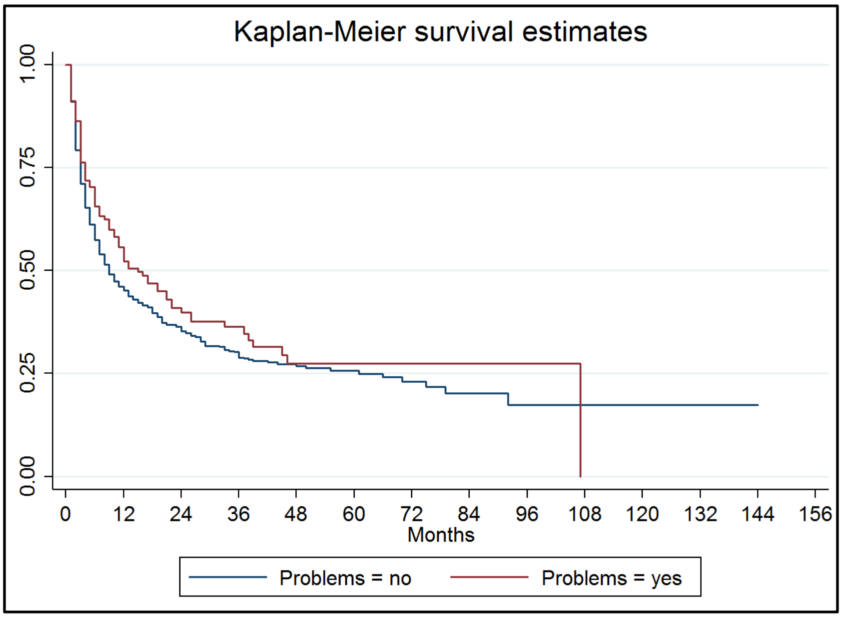

Figure 2. Kaplan-Meier survival estimates, by language problems

The use of the Kaplan-Meier estimator for this study is however problematic. The Kaplan-Meier estimator is inappropriate when applied to competing risks, because exits from the labour market without obtaining paid employment are treated as censored observations, which leads to bias (Coviello & Boggess, 2004). In a competing risks setting the cumulative incidence is a more correct method to graphically represent data. Cumulative incidence is estimated as:

$$\widehat{CI}_k(t) = \sum_{i|t_i<t} \hat{S}(t_{i-1}) \frac{d_{ki}}{n_i} \tag{5.2}$$

where $\hat{S}(t_{i-1})$ is the overall Kaplan-Meier estimator and different outcomes are represented by k (Coviello & Boggess, 2004). For both outcomes and language problem status, cumulative incidence functions are graphed in figure 3. These cumulative incidence functions suggest that having problems with the Dutch language decreases the incidence of an exit from unemployment to paid employment, while language problems at the same time increase incidence of other exits from unemployment. These graphs can however not be interpreted as evidence that language problems cause longer unemployment durations. They merely serve as graphical descriptions of the data, since no controls for other covariates are applied.

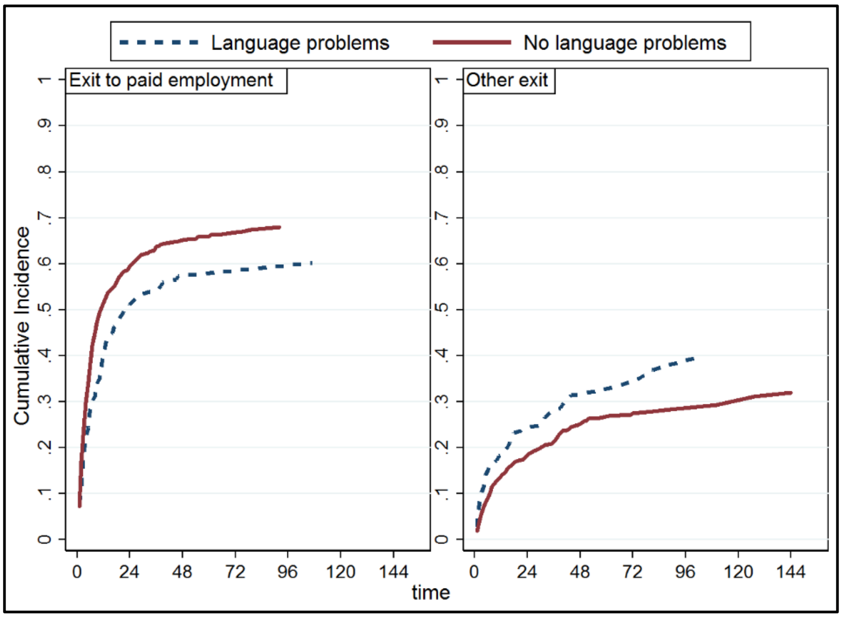

Figure 3. Cumulative incidence, by outcome and language problems

5 Results

In this section, all the discussed regression method will be conducted and their results will be interpreted.

5.1 Logistic regression

The results from the regressions are shown as odds-ratios. In the methodology section, the logistic regression function was formulated as follows:

$$\operatorname{logit}\{h(i,t)\} = \ln\left\{\frac{h(i,t)}{1-h(i,t)}\right\} = \beta_0 + \beta_1 Z_{it} + \beta_2 X_{it} \tag{6.1}$$

where the regression coefficients β_0, β_1 and β_2 are estimated using maximum likelihood estimators. To ease interpretation of these coefficients, they are transformed to odds by anti-logging:[23]

$$\left\{\frac{h(i,t)}{1-h(i,t)}\right\} = e^{\beta_0+\beta_1 Z_{it}+\beta_2 X_{it}} \tag{6.2}$$

The odds-ratio for a binary dummy variable which equals 1 as opposed to a situation where the dummy variable equals 0 is defined as:

$$\left\{\frac{\text{odds}(X_{it}=1)}{\text{odds}(X_{it}=0)}\right\} = \frac{e^{\beta_0+\beta_1 Z_{it}+\beta_2}}{e^{\beta_0+\beta_1 Z_{it}}} = e^{\beta_2} \tag{6.3}$$

Using the odds-ratio, β_2 can be more easily interpreted. In model (1), the odds-ratio of language problems equals 0.806. This means that someone who has problems with speaking or reading the Dutch language (language problems = 1), is 0.806 times as likely to exit

[23] The only non-transformed coefficient is the constant.

unemployment to paid employment compared with someone who has no problems (language problems = 0).

This results of the logistic regressions are shown in table 2. The first simple model (1) contains only two predictor variables: the language problems variable and dummy variables measuring unemployment duration to control for duration dependence. Unemployment duration is specified using tri-monthly dummies which are constructed for the first 60 months of unemployment and a last dummy is constructed if an individual is unemployed for more than 60 months. Using these dummy variables rather than a linear, cubic or logarithmic specification of time allows for a highly flexible baseline function (see also: Singer & Willett, 1993, 2003b).[24] However, alternative specifications of unemployment duration are also explored. In appendix C, the results of some models which use alternative specifications are shown. None of the alternative specifications result in a better model, because they do reduce flexibility, while they do not provide a much better AIC or BIC. Especially since many observations are available, the tri-monthly dummy variable can be used, because the ratio of events per predictor variable is large (see also: Peduzzi, Concato, Kemper, Holford, & Feinstein, 1996). For now, unemployment duration is therefore specified using the tri-monthly dummy variables.

Model (1) does suggest that having problems with the Dutch language does significantly lower the probability of exiting unemployment to paid employment. This model also suggests that the probability of exiting unemployment is significantly affected by unemployment duration. This duration dependence must however be interpreted with caution, because this model does not control for other possible predictors of unemployment duration. Heterogeneity could also cause this observed effect of unemployment duration on the probability of exiting unemployment. The individuals with a low probability remain unemployed longer than those with a higher probability of exiting unemployment to paid employment. This results in lower predicted probabilities for individuals who stay unemployed longer. The fact that a longer unemployment duration leads to a lower probability of leaving unemployment can thus be a result of both duration dependence and unobserved heterogeneity (Heckman & Borjas, 1980)**.**

Using model (1) as a starting point, additional covariates are added to the regression. By gradually adding more predictor variables, the model improves. Both the AIC and BIC decrease, while the Hosmer-Lemeshow test suggests that all models fit the data sufficiently. In model (2), control variables are added for age, the seasonally adjusted unemployment rate and education, which all have significant effects.

[24] Tri-monthly dummies are used instead of monthly dummies, because the data consists of a large number of time periods. Some unemployment spells last over 100 months. When monthly dummies were to be used, this would decrease statistical power of the analysis (Singer & Willett, 2003a). The last dummy is constructed for a larger number of months than three months, because there are relatively few exits to paid employment after 10 years of unemployment.

Model (3) adds controls for gender, and whether the individual has children living at home and whether he has a partner. Once controls for origin are added in model (4), the effect of having language problems on the probability of exiting unemployment loses significance. In all the following models, this effect remains insignificant. Based on the AIC and BIC, model (6) is the best performing model. The differences in estimated coefficients between models (4)-(8) are however very small. All these models show no significant effect of having language problems on the probability of exiting unemployment to paid employment. Unemployment rates, age and unemployment duration do significantly lower the odds-ratios of exiting unemployment to paid employment, while having a university education significantly increases this probability. Model (7) and (8) also suggest that performing unskilled manual work lowers the odds of finding paid employment.

As mentioned in the methodology section, treating individuals with a Dutch background and first- and second generation immigrants equal might be inappropriate, because language problems might be defined differently by the different groups. The results from the logit regressions suggest that there is another reason to estimate models for each origin-group separately. Mentioned earlier, the effect of having language problems was significant until controls for origin were added to the regression. An explanation for the change of significance of the effect of language problems is that origin is likely to be a confounding variable. Not only does origin affect the probability of exiting unemployment to paid employment,[25] but origin is also highly related to language problems. Especially first generation foreigners who learned Dutch as a second language will have a high probability of having problems with the Dutch language. Second generation foreigners on the other hand might face language related problems because their parents have trouble speaking Dutch, or because Dutch was less spoken at home. As seen earlier, the percentage of individuals who report to have language problems differs per origin group, with first generation foreigners having the highest percentages of individuals who report to have problems. Having problems with the Dutch language for them is likely caused by the fact that the first generation foreigners are not born in the Netherlands. Being an immigrant also likely causes a lower hazard rate of exiting unemployment to paid employment, as the results suggest. Origin thus influences both the

[25] Carlsson (2010) for example found that first- and second generation migrants have a significantly lower probability of being invited for a job interview than natives. The results of this study remained significant when control variables for language proficiency were added. Also noteworthy is the fact that the medium length of an unemployment spells for both natives as second generation foreigners is equal to 5 months, while the median length for first generation foreigners is equal to 12 months. While this is no evidence for a causal relationship, it does suggest that the probability of exiting unemployment for first generation foreigners is lower when no controls for heterogeneity are used.

dependent variable and the independent variable measuring language proficiency.
In order to control for this confounding variable, separate regression models were estimated for the different groups, also known as ex post stratification for categorical variables (Wunsch, 2007). The separate model estimations are based on model (6), which has the lowest AIC and BIC.
A major drawback of this stratification is the fact that it reduces sample size and statistical power. Especially the first- and second generation immigrant groups contain a much smaller number of events than the overall sample. The reduced sample sizes also mean that fewer observations are available per tri-monthly dummy, and the number of predictor variables per event is higher. For some dummies, no events are observed. For these dummies, Stata could not estimate logit coefficients and these dummies are dropped from the regression. Since a limited number of variables per dummy make the estimation of its coefficient difficult or impossible and too many predictor variables can lead to bias, a reduction of covariates is desirable (Peduzzi et al., 1996; Singer & Willett, 1993). Therefore the stratified models are also estimated using an alternative specification of unemployment duration to increase the number of predictor variables per event. Based on the models estimated in appendix C, the third degree polynomial of the number of months of unemployment is chosen as the alternative specification of unemployment duration. The results of the stratified regression models containing the tri-monthly dummy variables and the third degree polynomials are shown in table 4.
In none of the models the probability of finding paid employment is significantly affected by having language problems. Interestingly, the coefficient estimates for the stratified groups show large differences. For example, for first generation foreigners, age seems to play no role, and unemployment duration seems to play only a limited role in the determination of their probability of exiting unemployment to paid employment. For second generation foreigners, education seems to play no role at all. Compared to natives, both first- and second generation foreigners seem to be less affected by the covariates which have been used in this study. A Chow (Chow, 1960) test to assess whether the coefficients for model (6) are equal to the coefficients of the stratified models rejects this hypothesis when the third degree polynomial is used as a specification of unemployment duration ($p > \chi^2 = 0.0121$) and when the tri-monthly dummy variables are used ($p > \chi^2 = 0.0031$). The full model cannot assumed to be equal for each stratified model. This suggests that immigrants are affected differently by the used covariates. Also, different covariates than the ones used in this study could affect immigrants probability of exiting unemployment to paid employment.
The differences between the estimated models' groups might however also be a consequence of sample size differences. Both the first- and second generation groups only contain 72 and 74 exits from

unemployment to paid employment in 3.115 and 1.981 person-time observations. The native Dutch group contains over 790 exits to paid employment divided over 14.377 person-time observations. Comparison of these models should therefore be done with caution.

Something can be said about the hypotheses of this thesis, without having to compare the models. In none of the estimated models with controls for origin a significant effect of having trouble with the Dutch language on the probability of exiting unemployment to paid employment has been found. This result holds for both the overall sample and for each origin based group. In the next section, a competing risks element will be added to the regressions.

Table 3. Logit regression models. Coefficients are reported as odds ratios.

Model	1	2	3	4	5	6	7	8
Language problems (problems=1)	0.806*	0.764**	0.766*	0.940	0.937	0.932	0.943	0.939
Unemployment duration (reference: 1-3 months)								
4-6	0.781**	0.829*	0.828*	0.835	0.834*	0.834*	0.837	0.837
7-9	0.630***	0.710**	0.709**	0.730**	0.729**	0.728**	0.732**	0.732**
10-12	0.412***	0.490***	0.490***	0.511***	0.510***	0.510***	0.512***	0.513***
13-15	0.324***	0.398***	0.398***	0.418***	0.417***	0.417***	0.418***	0.418***
16-18	0.220***	0.274***	0.275***	0.291***	0.291***	0.290***	0.290***	0.290***
19-21	0.313***	0.404***	0.404***	0.430***	0.430***	0.428***	0.430***	0.430***
22-24	0.188***	0.250***	0.250***	0.266***	0.266***	0.265***	0.267***	0.267***
25-27	0.201***	0.275***	0.276***	0.295***	0.295***	0.294***	0.294***	0.294***
28-30	0.201***	0.285***	0.286***	0.308***	0.308***	0.306***	0.307***	0.306***
31-33	0.104***	0.151***	0.152***	0.164***	0.164***	0.163***	0.164***	0.163***
34-36	0.177***	0.261***	0.262***	0.287***	0.287***	0.286***	0.288***	0.287***
37-39	0.194***	0.285***	0.287***	0.313**	0.314**	0.312**	0.315**	0.314**
40-42	0.0601***	0.0854***	0.0862***	0.0937***	0.0937***	0.0931***	0.0939***	0.0934***
43-45	0.0685***	0.0946***	0.0954***	0.105**	0.105**	0.104**	0.105**	0.104**
46-48	0.163***	0.231**	0.234**	0.263**	0.264**	0.262**	0.264**	0.262**
49-51	0.0473**	0.0660**	0.0669**	0.0764*	0.0765*	0.0761*	0.0767*	0.0763*
52-54	0.0530**	0.0716**	0.0724**	0.0834*	0.0836*	0.0833*	0.0836*	0.0834*
55-57	0.123**	0.166*	0.168*	0.195*	0.196*	0.196*	0.196*	0.196*
58-60	0.0694**	0.0933*	0.0939*	0.114*	0.114*	0.114*	0.114*	0.114*
60+	0.0728***	0.104***	0.105***	0.125***	0.125***	0.125***	0.124***	0.123***
Age category (reference= <24)								
25 - 34 years		0.819	0.835	0.903	0.904	0.889	0.863	0.847
35 - 44 years		0.683***	0.690***	0.742**	0.738**	0.728**	0.699**	0.688**
45 - 54 years		0.538***	0.544***	0.566***	0.561***	0.553***	0.535***	0.527***
55 - 64 years		0.301***	0.305***	0.305***	0.302***	0.296***	0.289***	0.282***
Seasonal adjusted unemployment rate		0.893***	0.892***	0.887***	0.887***	0.887***	0.887***	0.888***
Education (reference: primary school)								
VMBO		1.364	1.359	1.277	1.274	1.264	1.245	1.243
HAVO/VWO		1.276	1.251	1.229	1.229	1.231	1.225	1.237
MBO		1.584*	1.565*	1.45	1.445	1.443	1.401	1.411
HBO		1.553*	1.546*	1.463	1.458	1.445	1.431	1.433
WO		2.119***	2.130***	2.030**	2.031**	2.003**	1.931**	1.918**
Female			0.942	0.942	0.942		0.964	
Children living at home			1.062	1.078	1.066		1.075	
Partner			1.083	1.047	1.02		1.011	
Origin (reference: Dutch background)								
First gen. foreign, Western				0.628*	0.637*	0.633*	0.641*	0.635*
First gen. foreign, non-Western				0.448***	0.452***	0.455***	0.475***	0.479***
Second gen. foreign, Western				0.74	0.741	0.749	0.749	0.757
Second gen. foreign, non-Western				0.646*	0.647*	0.644*	0.652*	0.653*
Natural logarithm of net household income					1.021	1.029	1.017	1.026
Profession (reference= higher academic or independent professional)								
Higher supervisory profession							0.783	0.79
Intermediate academic or independent professional							0.784	0.772
Intermediate supervisory or commercial							0.849	0.841
Other mental work							0.805	0.797
Skilled and supervisory manual work							0.817	0.819
Semi-skilled manual work							0.885	0.884
Unskilled manual work							0.667*	0.663*
Agrarian profession							0.848	0.855
Intercept	-2.226***	-1.575***	-1.626***	-1.513***	-1.634***	-1.667***	-1.349***	-1.368***
Observations	19473	19473	19473	19473	19473	19473	19473	19473
Likelihood-ratio χ^2	457.3	621.0	624.1	659.8	660.5	658.9	666.2	665.0
Likelihood-ratio test ($p > \chi^2$)	0.0000	0.0000	0.0000	0.0000	0.000	0.000	0.000	0.000
Hosmer-Lemeshow χ^2 (groups)	15.09(41)	468.13(468)	464.76(468)	426.92(468)	422.93(468)	492.32(468)	452.04(468)	522.14(468)
Goodness-of-fit test ($p > \chi^2$)	0.9998	0.4635	0.5076	0.9024	0.9243	0.1926	0.6700	0.0366
Hosmer-Lemeshow χ^2 (groups)	0.40(8)	11.50(10)	13.45(10)	1.54(10)	1.63(10)	0.75(10)	2.55(10)	6.12(10)
Goodness-of-fit test ($p > \chi^2$)	0.9989	0.1748	0.0973	0.9921	0.9904	0.9994	0.9593	0.6339
Akaike information criteria	7094.8	6951.1	6954.0	6926.3	6927.6	6923.2	6937.9	6933.1
Bayesian information criterion	7268.1	7203.2	7229.7	7233.5	7242.7	7214.6	7316.0	7287.5

* p<0.05, ** p<0.01, *** p<0.001

Table 4. Logit regression by origin, model (6)

	Dutch		Foreign			
			First generation		Second generation	
Language problems	0.954	0.95	0.88	0.819	0.929	0.937
Age category (reference= <24)						
25 - 34 years	0.944	0.941	1.13	1.1	0.615	0.621
35 - 44 years	0.762*	0.760*	0.96	0.917	0.386*	0.388*
45 - 54 years	0.554***	0.550***	0.86	0.869	0.573	0.576
55 - 64 years	0.303***	0.302***	0.588	0.569	0.249**	0.251**
Seasonal adjusted unemployment rate	0.878***	0.878***	0.841	0.835	0.877	0.877
Education (reference: primary school)						
VMBO	1.541	1.529	0.678	0.677	1.854	1.798
HAVO/VWO	1.62	1.603	0.243*	0.245*	1.965	1.928
MBO	1.760*	1.741*	0.936	0.933	1.471	1.424
HBO	1.683	1.665	0.899	0.833	2.463	2.386
WO	2.269**	2.249**	0.975	0.969	6.644	6.513
Natural logarithm of net household income	1.044	1.045	0.936	0.938	1.041	1.043
Unemployment duration (reference: 1-3 months)						
4-6	0.842		1.558		0.547	
7-9	0.767*		0.638		0.632	
10-12	0.494***		1.321		0.341	
13-15	0.422***		0.501		0.499	
16-18	0.278***		0.769		0.225*	
19-21	0.402***		1.444		0.259	
22-24	0.254***		0.87		0.144	
25-27	0.302***		1.031		1.000	
28-30	0.304***		1.177		1.000	
31-33	0.179***		1.000		0.195	
34-36	0.257**		0.467		0.424	
37-39	0.156**		1.895		0.556	
40-42	0.126**		1.000		1.000	
43-45	0.0698**		1.000		0.385	
46-48	0.275*		0.757		1.000	
49-51	0.108*		1.000		1.000	
52-54	1.000		0.769		1.000	
55-57	0.288		1.000		1.000	
58-60	1.000		1.000		0.831	
60+	0.0267***		0.453		0.49	
Months		0.923***		1.034		0.889*
Months^2		1.001		0.998		1.002
Months^3		1.000		1.000017*		1.000
Intercept	-1.916***	-1.752***	-1.711	-1.649	-2.135	-1.985
Observations[26]	14196	14377	2756	3115	1658	1981
Likelihood-ratio χ^2	518.0	520.6	25.93	28.77	55.91	66.00
Likelihood-ratio test $(p > \chi^2)$	0.0000	0.000	0.467	0.0172	0.000373	0.000
Hosmer-Lemeshow χ^2 (groups)	366.96(395)	378.29(395)	37.36(34)	26.43(34)	41.77(37)	31.78(37)
Goodness-of-fit test $(p > \chi^2)$	0.8228	0.6942	0.2364	0.7443	0.2002	0.6245
Akaike information criteria	5643.3	5631.3	695	688.0	600.9	597.7
Bayesian information criterion	5877.7	5752.5	854.9	784.7	741.7	687.2

* p<0.05, ** p<0.01, *** p<0.001

[26] Differences in the number of observations between the regressions using the tri-monthly dummies and the natural logarithm of time occur because some of the tri-month dummies predict failure perfectly and are therefore dropped.

5.2 Competing risks

As discussed in the methodology section, unemployed individuals are not only at risk of exiting unemployment by obtaining paid employment, but they can also leave unemployment trough other exit outcomes. Individuals might become inactive, retired or they could start working as a freelancer. If these exit from unemployment are recorded as censored, it is likely to cause bias in the estimates of the effect of language proficiency on unemployment duration. To control for this bias, exits from unemployment without obtaining paid employment are recorded as the competing risk. The results of the competing risks regression are shown in table 5. The coefficients are reported as odds ratio,[27] with a non-event as the baseline event ($Y_{it} = 0$). Interpretation of these odds ratios is similar is in the previous section. A unit change of a predictor variable changes the odds ratio of outcome k relative to no event by a factor of the odds ratio parameter of that predictor variable.

Analogue to the logit regression models estimated in the previous section, model (1) contains two covariates: the language problems dummy variable and the tri-monthly unemployment duration dummy variables. Additional variables are gradually added to the regressions. Similar to the estimated models from the previous the models without control for origin show a significant negative effect of having language problems on the probability of exiting unemployment by obtaining paid employment. When controls are added for origin, this effect becomes insignificant. Interestingly, the effect of having language problems on the probability of exiting unemployment without obtaining paid employment is significantly positive in models (3)-(8). Interpretation of the parameter coefficients for the exit outcome without obtaining paid employment has to be done with utmost caution. Stated earlier, the exit from unemployment without obtaining paid employment covers a wide variety of outcomes, such as entrepreneurship, assistance in a family business, receiving additional education or retirement. Furthermore, exits from unemployment without obtaining paid employment might depend on a different set of predictor variables.

As discussed earlier, treating natives equal as first- and second generation foreigners might result in biases because of the confounding aspects of the origin variables or because of measurement errors which differ by group. As in the previous section the model with the lowest AIC and BIC will be estimated for each group separately. Model (6) has both the lowest AIC and BIC. Not unimportant, a Hausman test on the independence of irrelevant alternatives (IIA) assumption confirms the hypothesis for this model that the different outcomes are independent of their alternatives.[28] The multinomial logistic regression

[27] These odds ratios are obtained be exponentiation the multinomial logit coefficients.
[28] For the exit to paid employment: $p > \chi^2 = 0.999$. For the exit without paid employment: $p > \chi^2 = 1.000$. The IIA assumption is confirmed to hold by the Hausman test.

models are therefore appropriate to use. The modified goodness-of-fit tests suggest that all the estimates models, except model (3), do provide a good fit to the data.
The estimates of the regression models separated by origin are shown in table 6. For all groups, both the models using tri-monthly dummy variables and the third degree polynomial as a specification for unemployment duration are estimated. The first models use the whole model without controls for origin. Similar to the models estimated in the previous section, these models suggest that hardly any of the covariates affect the hazard rate significantly for non-natives. The probability of exiting unemployment to paid employment for first generation non-natives is only significantly lower for individuals with a HAVO/VWO education.
Unexpected results are found in the models for second generation foreigners. The estimated odds ratios of all education levels except WO are extremely high, with extremely huge standard deviations as well. These estimates are thus very imprecise.[29]
As mentioned in the previous section, stratification of the sample leads to the case where for some tri-monthly dummy variables no events are observed. The estimated logit coefficients for these dummies are very low, but they have a huge standard error. These coefficients could not be reliably estimated and the odds-ratios for these dummies are therefore unreliable as well.[30]
It must be stressed that the fact that very few covariates have a significant effect on the hazard rate for first- and second generation foreigners might be a consequence of sample size differences. As in the previous section, a Chow test is used to test whether the stratified models are nested in the model using the full sample and dummies to control for origin. When the models are using the third degree polynomial to specify unemployment duration, equality of the coefficients of the different regression models has to be rejected ($p > \chi^2 = 0.0363$). When the tri-monthly dummies are used as a specification of unemployment duration equality of coefficients can however not be rejected ($p > \chi^2 = 0.1934$).
Briefly mentioned earlier, estimation problems might occur because the number of events per predictor variable (EPV) is rather small in the models for first- and second generation foreigners. As a rule of thumb, a minimum of 10 EPV has been proposed to avoid overfitting (Peduzzi et al., 1996).[31] The models using the third degree polynomial as a

[29] When an alternative specification of education is used (a dummy variable which equals 1 if the individual received higher education), a more reliable estimate for the effect of education on the hazard rate is produced. This different specification does however not lead to the estimation of a significant effect of language proficiency on the hazard rate. For the purpose of this study it is therefore irrelevant. No abnormalities in the data could be found.
[30] For example, the reported odds-ratio for a first generation foreigner who is unemployed for 31-33 months is 2.06E-07, which cannot be assumed to be correct.
[31] Other researchers did find that the minimum number of EPV for logistic regression with a primary binary predictor might be as low as 5 EPV (Vittinghoff & McCulloch, 2007).

specification for unemployment duration contain 15 predictor variables. For the first generation foreigners, only 72 exits to paid employment are observed, and for the second generation foreigners, 74 exits to paid employment are observed. Both of these models contain less than 10 EPV. To check whether significantly different results are be estimated when the EPV was higher, the multinomial regression models were estimated with a reduced number of predictor variables – by differently specifying unemployment duration, age and education. The results can be found in appendix D. The reduction in the number of predictor variables did not lead to very different estimated odds ratios.

Again, the results of the stratified models should still be carefully interpreted, since sample sizes are low. Based on all the models estimated in this section, no significant effect of having language problems on the probability of exiting unemployment to paid employment can be assumed for both the overall sample and more specifically for immigrants.

Table 5. Multinomial logit regression estimates with competing risks.

Exit outcome (paid employment or other)	Paid	Other	Paid	Other	Paid	Other	Paid	Other
Model	1	1	2	2	3	3	4	4
Language problems	0.809*	1.177	0.769*	1.305	0.771*	1.335*	0.952	1.623**
Unemployment duration (reference: 1-3 months)								
4-6	0.779**	0.895	0.827*	0.907	0.826*	0.914	0.833*	0.923
7-9	0.627***	0.794	0.706**	0.809	0.706**	0.817	0.727**	0.843
10-12	0.407***	0.586*	0.485***	0.600*	0.485***	0.607*	0.506***	0.634
13-15	0.321***	0.673	0.395***	0.688	0.395***	0.693	0.415***	0.726
16-18	0.218***	0.67	0.272***	0.68	0.273***	0.684	0.289***	0.723
19-21	0.307***	0.274***	0.396***	0.275**	0.397***	0.279**	0.423***	0.296**
22-24	0.186***	0.534*	0.247***	0.539*	0.247***	0.546	0.263***	0.581
25-27	0.198***	0.452*	0.271***	0.450*	0.272***	0.456*	0.291***	0.484*
28-30	0.197***	0.347*	0.280***	0.342*	0.281***	0.343*	0.303***	0.366*
31-33	0.103***	0.653	0.150***	0.642	0.150***	0.64	0.163***	0.686
34-36	0.175***	0.596	0.258***	0.582	0.259***	0.583	0.284***	0.63
37-39	0.196***	1.503	0.288***	1.464	0.291***	1.448	0.318**	1.569
40-42	0.0600***	0.918	0.0852***	0.907	0.0859***	0.883	0.0935***	0.955
43-45	0.0686***	1.05	0.0946***	1.041	0.0954***	1.012	0.105**	1.105
46-48	0.161***	0.462	0.228**	0.453	0.231**	0.433	0.260**	0.487
49-51	0.0472**	0.908	0.0658**	0.881	0.0666**	0.844	0.0763*	0.962
52-54	0.0522**	0.398	0.0705**	0.391	0.0711**	0.372	0.0822*	0.429
55-57	0.121**	0.464	0.164*	0.456	0.165*	0.439	0.193*	0.504
58-60	0.0676**	9.12E-07	0.0908*	2.29E-07	0.0913*	2.24E-07	0.111*	5.84E-07
60+	0.0719***	0.549*	0.103***	0.520*	0.103***	0.488*	0.124***	0.551*
Age category (reference= <24)								
25 - 34 years			0.805*	0.509**	0.821	0.480***	0.889	0.534**
35 - 44 years			0.675***	0.684	0.682***	0.649*	0.735**	0.727
45 - 54 years			0.530***	0.580**	0.536***	0.551**	0.558***	0.581**
55 - 64 years			0.299***	0.869	0.302***	0.799	0.303***	0.802
Seasonal adjusted unemployment rate			0.892***	0.972	0.892***	0.979	0.886***	0.979
Education (reference: primary school)								
VMBO			1.366	1.071	1.363	1.149	1.28	1.098
HAVO/VWO			1.277	1.038	1.257	1.192	1.235	1.22
MBO			1.581*	0.911	1.567*	1.018	1.45	0.956
HBO			1.557*	1.107	1.555*	1.24	1.471	1.214
WO			2.123***	1.072	2.138***	1.138	2.037**	1.129
Female					0.939	0.881	0.939	0.875
Children living at home					1.057	0.808	1.074	0.849
Partner					1.073	0.691***	1.036	0.658***
Origin (reference= Dutch background)								
First generation foreign, Western							0.623*	0.696
First generation foreign, non-Western							0.441***	0.379***
Second generation foreign, Western							0.737	0.798
Second generation foreign, non-Western							0.641*	0.646
Natural logarithm of net household income								
Profession (reference= higher academic or independent professional)								
Higher supervisory profession								
Intermediate academic or independent professional								
Intermediate supervisory or commercial								
Other mental work								
Skilled and supervisory manual work								
Semi-skilled manual work								
Unskilled manual work								
Agrarian profession								
Intercept	-2.199***	-3.605***	-1.535***	-3.138***	-1.579***	-2.845***	-1.465***	-2.806***
Observations	19473		19473		19473		19473	
Likelihood-ratio χ^2 (df)	505.3		690.4		714.2		769.6	
Likelihood-ratio test $(p > \chi^2)$	0.0000		0.0000		0.0000		0.0000	
Hosmer-Lemeshow χ^2 (groups)	N/A		939.342(468)		1004.368(468)		917.172(468)	
Goodness-of-fit test $(p > \chi^2)$	N/A		0.427		0.049		0.629	
Akaike information criteria	10727.6		10582.5		10570.7		10531.3	
Bayesian information criterion	11074.1		11086.6		11122.0		11145.7	

* p<0.05, ** p<0.01, *** p<0.001

Table 5. Multinomial logit regression estimates with competing risks, continued.

Exit outcome (paid employment or other)	Paid	Other	Paid	Other	Paid	Other	Paid	Other
Model	5	5	6	6	7	7	8	8
Language problems	0.949	1.658***	0.942	1.645***	0.954	1.628**	0.95	1.609**
Unemployment duration (reference: 1-3 months)								
4-6	0.832*	0.935	0.832*	0.934	0.835	0.952	0.836	0.952
7-9	0.726**	0.855	0.725**	0.853	0.730**	0.877	0.730**	0.876
10-12	0.505***	0.643	0.505***	0.639	0.508***	0.658	0.508***	0.656
13-15	0.415***	0.736	0.414***	0.733	0.415***	0.753	0.415***	0.751
16-18	0.289***	0.734	0.288***	0.733	0.288***	0.749	0.288***	0.748
19-21	0.422***	0.303**	0.421***	0.301**	0.423***	0.308**	0.423***	0.307**
22-24	0.263***	0.593	0.262***	0.587	0.264***	0.603	0.264***	0.599
25-27	0.291***	0.494*	0.290***	0.489*	0.290***	0.503	0.290***	0.5
28-30	0.303***	0.373*	0.301***	0.372*	0.303***	0.379*	0.302***	0.380*
31-33	0.163***	0.699	0.162***	0.701	0.163***	0.716	0.162***	0.724
34-36	0.284***	0.635	0.283***	0.633	0.285***	0.655	0.284***	0.659
37-39	0.319**	1.562	0.317**	1.56	0.320**	1.604	0.319**	1.62
40-42	0.0936***	0.968	0.0930***	0.978	0.0938***	0.973	0.0933***	0.997
43-45	0.105**	1.122	0.104**	1.137	0.105**	1.124	0.105**	1.152
46-48	0.260**	0.488	0.258**	0.494	0.260**	0.483	0.259**	0.495
49-51	0.0764*	0.966	0.0759*	0.978	0.0766*	0.963	0.0761*	0.988
52-54	0.0823*	0.434	0.0820*	0.442	0.0823*	0.434	0.0821*	0.45
55-57	0.193*	0.506	0.193*	0.511	0.194*	0.507	0.193*	0.52
58-60	0.111*	5.99E-07	0.111*	6.09E-07	0.111*	7.92E-07	0.111*	8.07E-07
60+	0.124***	0.567	0.124***	0.585	0.122***	0.519*	0.122***	0.545*
Age category (reference= <24)								
25 - 34 years	0.89	0.523**	0.874	0.525**	0.849	0.514**	0.832	0.514**
35 - 44 years	0.732**	0.778	0.723**	0.806	0.694**	0.798	0.684**	0.825
45 - 54 years	0.554***	0.637*	0.547***	0.665*	0.529***	0.655*	0.521***	0.687
55 - 64 years	0.300***	0.903	0.295***	0.969	0.288***	0.926	0.281***	0.992
Seasonal adjusted unemployment rate	0.886***	0.977	0.887***	0.973	0.887***	0.976	0.887***	0.971
Education (reference: primary school)								
VMBO	1.278	1.124	1.267	1.075	1.25	1.152	1.247	1.102
HAVO/VWO	1.236	1.237	1.236	1.13	1.233	1.295	1.244	1.2
MBO	1.446	0.995	1.442	0.937	1.406	1.065	1.414	1.013
HBO	1.467	1.271	1.452	1.196	1.44	1.24	1.441	1.194
WO	2.039**	1.146	2.010**	1.103	1.933**	1.014	1.920**	0.992
Female	0.939	0.87			0.962	0.886		
Children living at home	1.063	0.917			1.072	0.917		
Partner	1.015	0.761*			1.006	0.753*		
Origin (reference= Dutch background)								
First generation foreign, Western	0.630*	0.63	0.626*	0.643	0.634*	0.635	0.628*	0.642
First generation foreign, non-Western	0.444***	0.352***	0.447***	0.348***	0.466***	0.338***	0.470***	0.337***
Second generation foreign, Western	0.737	0.789	0.745	0.783	0.745	0.796	0.752	0.795
Second generation foreign, non-Western	0.642*	0.647	0.639*	0.684	0.648*	0.667	0.649*	0.71
Natural logarithm of net household income	1.016	0.888***	1.024	0.856***	1.012	0.884***	1.02	0.851***
Profession (reference= higher academic or independent professional)								
Higher supervisory profession					0.769	0.542*	0.775	0.525*
Intermediate academic or independent professional					0.767	0.442**	0.754	0.441**
Intermediate supervisory or commercial					0.828	0.367***	0.819	0.369***
Other mental work					0.789	0.470**	0.78	0.460**
Skilled and supervisory manual work					0.798	0.406**	0.8	0.423**
Semi-skilled manual work					0.868	0.499*	0.867	0.531*
Unskilled manual work					0.655*	0.512*	0.650*	0.511*
Agrarian profession					0.862	1.017	0.869	1.032
Intercept	-1.559***	-2.156***	-1.589***	-2.126***	-1.253***	-1.416**	-1.268***	-1.384**
Observations	19473		19473		19473		19473	
Likelihood-ratio χ^2 (df)	782.3		772.6		804.5		795.4	
Likelihood-ratio test $(p > \chi^2)$	0.0000		0.0000		0.0000		0.0000	
Hosmer-Lemeshow χ^2 (groups)	884.723(468)		907.141(468)		858.250(468)		953.849(468)	
Goodness-of-fit test $(p > \chi^2)$	0.864		0.714		0.959		0.302	
Akaike information criteria	10522.6		10520.2		10532.3		10529.4	
Bayesian information criterion	11152.7		11103.1		11288.5		11238.3	

* p<0.05, ** p<0.01, *** p<0.001

Table 6. Competing risks regression model odds ratio estimates by group.

Group	Full sample (no controls for origin)				Dutch			
Exit outcome	Paid	Other	Paid	Other	Paid	Other	Paid	Other
Language problems	0.768*	1.306	0.766*	1.286	0.965	1.548*	0.961	1.538*
Age category (reference= <24)								
25 - 34 years	0.815	0.475***	0.812	0.466***	0.932	0.654	0.93	0.649
35 - 44 years	0.673***	0.702	0.669***	0.683	0.757*	0.838	0.755*	0.82
45 - 54 years	0.525***	0.618*	0.523***	0.606*	0.548***	0.652	0.544***	0.641*
55 - 64 years	0.295***	0.959	0.294***	0.928	0.303***	1.008	0.302***	0.984
Seasonal adjusted unemployment rate	0.891***	0.973	0.892***	0.968	0.877***	0.953	0.877***	0.943
Education (reference: primary school)								
VMBO	1.347	1.13	1.342	1.138	1.538	0.877	1.525	0.873
HAVO/VWO	1.256	1.101	1.247	1.072	1.628	1.147	1.609	1.102
MBO	1.549*	0.995	1.538*	0.98	1.756*	0.865	1.736*	0.842
HBO	1.529*	1.208	1.519	1.2	1.69	1.102	1.671	1.087
WO	2.110***	1.098	2.103***	1.113	2.263**	0.849	2.243**	0.847
Natural logarithm of net household income	1.038	0.871***	1.039	0.867***	1.038	0.853***	1.038	0.849***
Unemployment duration (reference: 1-3 months)								
4-6	0.825*	0.921			0.84	0.931		
7-9	0.705**	0.821			0.765*	0.897		
10-12	0.484***	0.608*			0.491***	0.747		
13-15	0.394***	0.697			0.419***	0.741		
16-18	0.272***	0.689			0.275***	0.668		
19-21	0.396***	0.282**			0.395***	0.340*		
22-24	0.247***	0.548			0.251***	0.578		
25-27	0.272***	0.457*			0.297***	0.355*		
28-30	0.280***	0.344*			0.300***	0.493		
31-33	0.150***	0.647			0.178***	0.831		
34-36	0.260***	0.58			0.253**	0.524		
37-39	0.291***	1.433			0.159**	1.81		
40-42	0.0856***	0.902			0.125**	0.805		
43-45	0.0952***	1.04			0.0697**	0.936		
46-48	0.231**	0.441			0.270*	0.458		
49-51	0.0666**	0.859			0.108*	0.798		
52-54	0.0714**	0.385			8.06E-08	0.613		
55-57	0.167*	0.446			0.286	0.72		
58-60	0.0924*	2.3E-07			7.98E-08	7.71E-08		
60+	0.104***	0.511*			0.0266***	0.775		
Months			0.923***	0.977			0.923***	0.977
Months^2			1.001*	1			1.001	1
Months^3			1	1			1	1
Intercept	-1.788***	-2.257***	-1.639***	-2.200***	-1.827***	-1.923***	-1.660***	-1.793***
Observations	19473		19473		14377		14377	
Likelihood-ratio χ^2 (df)	715.4		664.6		609.0		567.7	
Likelihood-ratio test ($p > \chi^2$)	0.0000		0.0000		0.0000		0.0000	
Akaike information criteria	10561.4		10544.2		8554.3		8527.7	
Bayesian information criterion	11081.3		10796.3		9054.2		8770.0	

* p<0.05, ** p<0.01, *** p<0.001

Table 6. Competing risks regression model odds ratio estimates by group, continued.

Group	Foreign							
	First generation				Second generation			
Exit outcome	Paid	Other	Paid	Other	Paid	Other	Paid	Other
Language problems	0.886	1.49	0.824	1.469	0.949	1.739	0.959	1.711
Age category (reference= <24)								
25 - 34 years	1.114	0.434	1.086	0.441	0.586	0.139**	0.591	0.148*
35 - 44 years	0.961	1.192	0.919	1.226	0.368*	0.190*	0.371*	0.223
45 - 54 years	0.861	1.131	0.871	1.17	0.553	0.351	0.557	0.376
55 - 64 years	0.586	0.723	0.567	0.686	0.244**	0.753	0.246**	0.737
Seasonal adjusted unemployment rate	0.842	1.057	0.835	1.062	0.875	0.906	0.876	0.947
Education (reference: primary school)								
VMBO	0.688	2.008	0.686	1.879	1.945	19579780	1.886	491265.6
HAVO/VWO	0.241*	0.688	0.243*	0.686	2.063	17281105	2.021	408410
MBO	0.933	0.774	0.929	0.728	1.549	21866848	1.498	519956.6
HBO	0.904	1.679	0.838	1.615	2.575	11874501	2.491	293517.9
WO	0.985	1.907	0.977	1.873	7.194	58576658	7.025	1220017.6
Natural logarithm of net household income	0.931	0.823**	0.933	0.824**	1.04	1.03	1.042	1.029
Unemployment duration (reference: 1-3 months)								
4-6	1.551	0.89			0.552	1.441		
7-9	0.633	0.82			0.63	0.769		
10-12	1.302	0.44			0.337	1.53E-07		
13-15	0.498	0.815			0.499	1.021		
16-18	0.766	0.995			0.227*	1.563		
19-21	1.414	2.29E-07			0.257	0.506		
22-24	0.866	0.98			0.143	0.56		
25-27	1.028	1.108			7.24E-08	1.2		
28-30	1.151	3.3E-07			7.29E-08	1.09E-07		
31-33	2.06E-07	0.617			0.192	1.07E-07		
34-36	0.462	0.618			0.428	1.423		
37-39	1.913	1.791			0.547	1.17E-07		
40-42	2.35E-07	1.102			6.30E-08	2.394		
43-45	2.05E-07	2.383			0.388	1.465		
46-48	0.741	1.88E-07			7.50E-08	1.592		
49-51	2.21E-07	1.81E-07			2.96E-08	4.907		
52-54	0.752	1.85E-07			6.49E-08	1.14E-07		
55-57	1.93E-07	1.87E-07			5.80E-08	1.2E-07		
58-60	1.87E-07	1.9E-07			0.819	1.27E-07		
60+	0.444	0.167			0.487	1.52E-07		
Months			1.033	0.993	0.949		0.888*	0.904
Months^2			0.998	1			1.002	1.004
Months^3			1.000*	1	0.586		1	1
Intercept	-1.663	-3.459*	-1.604	-3.591*	-2.121	-19.58	-1.969	-15.81
Observations	3115		3115		1981		1981	
Likelihood-ratio χ^2 (df)	88.93		57.21		123.6		85.65	
Likelihood-ratio test ($p > \chi^2$)	0.0214		0.00198		0.0000113		0.0000003	
Akaike information criteria	1127.7		1091.5		948.7		918.7	
Bayesian information criterion	1526.7		1284.9		1317.8		1097.6	

* p<0.05, ** p<0.01, *** p<0.001

5.3 Multiple unemployment spells and unobserved heterogeneity

The next step in the analysis is to add a multilevel aspect to the multinomial models to allow multiple unemployment spells per individual and unobserved heterogeneity. The results of the multilevel multinomial logistic regression models are shown in table 7. The random effect terms (variance and standard deviation) are reported as logit coefficients. When the multilevel multinomial models are compared with the models estimated in the previous paragraph, the multilevel models which include unobserved heterogeneity show almost identical odds-ratio estimates. The Akaike- and Bayesian information criteria are also very similar, with the AIC being slightly lower and the BIC slightly higher for the multilevel models.

Model (2) and (3) show a significant effect of having language problems on the probability of leaving unemployment to paid employment. This significance does however disappear after controls on origin are applied. As with the 'normal' competing risks models, having language problems does seem to significantly increase the probability of exiting unemployment without getting a paid job in models (3)-(8).

As in the previous section, model (6) provides the best fit to the data. When this multilevel multinomial model is compared with the competing risks model (6) of the previous section to assess whether the addition of a random intercept improves the goodness-of-fit. This is done using a likelihood-ratio test. The null hypothesis that model (6) without a multilevel component fits the data better than model (6) with a multilevel component can be rejected at a significance level of 5% ($p > \chi^2 = 0.0156$).

The random intercept shows that there is quite some unexplained heterogeneity between individuals. Taking model (6) as an example, the random intercept can be interpreted as follows. According to this model, 95% of individuals who are not affected by any covariate[32] have a probability of exiting unemployment to paid employment between:

$$\mathrm{p} = \frac{e^{\beta_0 - 1.96\sqrt{\varepsilon_i}}}{1 + e^{\beta_0 - 1.96\sqrt{\varepsilon_i}}} = \frac{e^{-1.573 \pm 1.96 * 0.386}}{1 + e^{-1.573 \pm 1.96 * 0.386}} = \{0.307, 0.593\} \tag{6.6}$$

The random effect thus leads to a substantial variability of the baseline hazard.

The same procedure as in the previous sections will be applied to control for the confounding nature of the origin variable and possible differences in measurement errors by origin. The results of the regression model estimations stratified by origin are shown in table 8.

Unfortunately, Stata's GSEM could not estimate the multilevel multinomial logit models based on model (6) for second generation

---

[32] That is: they have no language problems, are unemployed for 1-3 months, are younger than 24 years, the unemployment rate is 0, they have a primary school education, are of Dutch origin, have no net household income.

foreigners as log-likelihood iterations remained non-concave. When different predictor variables were removed from the model for second generation immigrants, it was found that the dummy variables on education caused Stata to be unable to fit the model.[33] The regression model for the second generation foreigners are therefore estimated without control variables for education.

The results of the multilevel competing risks models look very similar to the results of the competing risks regression models (table 6). At first, language proficiency seems to have a significant negative effect on the probability of exiting unemployment to paid employment when logit coefficients are estimated for the full sample without controlling for origin. When coefficients are estimated for each origin-strata separately, the effect of language proficiency on the exit probability to a paid job becomes insignificant. For natives, reported problems with the Dutch language do significantly increase the probability of an unemployment exit without obtaining paid employment. The Chow tests that have been performed in the previous two section are not available for Stata's GSEM and could therefore not be performed.

In appendix D reduced predictor variable models are estimated for first- and second generation foreigners. The results of these reduced predictor variable models do not differ much from the full predictor variable models.

[33] The estimations of the stratified models for second generation foreigners in the previous section already showed remarkable estimates for the education dummy coefficients.

Table 7. Multilevel multinomial regression model.

Model	1	1	2	2	3	3	4	4
Exit outcome	Paid	Other	Paid	Other	Paid	Other	Paid	Other
Language problems	0.813	1.183	0.773*	1.31	0.775*	1.334*	0.975	1.655**
Unemployment duration (reference: 1-3 months)								
4-6	0.809*	0.929	0.864	0.945	0.86	0.951	0.878	0.973
7-9	0.672***	0.852	0.765*	0.874	0.761*	0.88	0.8	0.93
10-12	0.447***	0.642	0.539***	0.663	0.535***	0.669	0.574***	0.721
13-15	0.359***	0.75	0.447***	0.775	0.444***	0.777	0.482***	0.844
16-18	0.247***	0.757	0.312***	0.777	0.310***	0.777	0.341***	0.856
19-21	0.352***	0.314**	0.463***	0.319**	0.459***	0.321**	0.508**	0.356*
22-24	0.214***	0.615	0.290***	0.628	0.287***	0.632	0.319***	0.704
25-27	0.230***	0.525	0.322***	0.53	0.319***	0.533	0.357***	0.595
28-30	0.231***	0.405*	0.335***	0.406*	0.332***	0.403*	0.376**	0.454
31-33	0.121***	0.765	0.179***	0.763	0.178***	0.754	0.203***	0.855
34-36	0.206***	0.702	0.311**	0.696	0.309**	0.69	0.357**	0.791
37-39	0.234***	1.793*	0.353**	1.775	0.352**	1.738	0.407*	2.004*
40-42	0.0721***	1.105	0.105**	1.114	0.105**	1.074	0.121**	1.243
43-45	0.0820***	1.258	0.116**	1.275	0.116**	1.227	0.135**	1.434
46-48	0.191**	0.552	0.279*	0.554	0.280*	0.524	0.335*	0.63
49-51	0.0565**	1.09	0.0814*	1.086	0.0817*	1.027	0.0994*	1.254
52-54	0.0624**	0.478	0.0870*	0.482	0.0872*	0.454	0.107*	0.562
55-57	0.145**	0.558	0.205*	0.565	0.204*	0.539	0.254	0.669
58-60	0.0808*	3.88E-08	0.114*	9.46E-09	0.113*	6.47E-09	0.146	6.60E-09
60+	0.0849***	0.649	0.128***	0.638	0.128***	0.596	0.163***	0.725
Age category (reference= 24-)								
25 - 34 years			0.793*	0.504**	0.809	0.475***	0.873	0.527**
35 - 44 years			0.650***	0.657*	0.657***	0.628*	0.705**	0.7
45 - 54 years			0.501***	0.544**	0.509***	0.520**	0.523***	0.542**
55 - 64 years			0.276***	0.797	0.279***	0.738	0.273***	0.721
Seasonal adjusted unemployment rate			0.888***	0.972	0.887***	0.979	0.882***	0.979
Education (reference: primary school)								
VMBO			1.36	1.063	1.358	1.123	1.245	1.043
HAVO/VWO			1.287	1.034	1.259	1.174	1.208	1.172
MBO			1.591*	0.905	1.573*	0.999	1.431	0.917
HBO			1.583*	1.127	1.576*	1.246	1.464	1.195
WO			2.136**	1.069	2.145**	1.125	2.009**	1.09
Female					0.941	0.879	0.939	0.87
Children living at home					1.058	0.808	1.076	0.849
Partner					1.082	0.694**	1.04	0.654***
Origin (reference= Dutch background)								
First generation foreign, Western							0.585*	0.648
First generation foreign, non-Western							0.397***	0.328***
Second generation foreign, Western							0.711*	0.754
Second generation foreign, non-Western							0.618*	0.603
Natural logarithm of net household income								
Profession (reference= higher academic or independent professional)								
Higher supervisory profession								
Intermediate academic or independent professional								
Intermediate supervisory or commercial								
Other mental work								
Skilled and supervisory manual work								
Semi-skilled manual work								
Unskilled manual work								
Agrarian profession								
Intercept	-2.257***	-3.662***	-1.542***	-3.162***	-1.586***	-2.854***	-1.446***	-2.783***
Random effect								
Variance	0.121		0.129		0.121		0.154*	
Std. deviation	0.348		0.359		0.348		0.392*	
Observations	19437		19437		19437		19437	
Akaike information criteria	10725.8		10580.1		10568.9		10527.2	
Bayesian information criterion	11080.2		11092.1		11128.2		11149.5	

* p<0.05, ** p<0.01, *** p<0.001

Table 7. Multilevel multinomial regression model. Continued

Model	5	5	6	6	7	7	8	8
Exit outcome	Paid	Other	Paid	Other	Paid	Other	Paid	Other
Language problems	0.969	1.680***	0.964	1.668**	0.975	1.653**	0.971	1.636**
Unemployment duration (reference: 1-3 months)								
4-6	0.875	0.983	0.876	0.982	0.878	1.001	0.879	1.000
7-9	0.795	0.938	0.797	0.936	0.8	0.962	0.8	0.961
10-12	0.570***	0.725	0.571***	0.722	0.573***	0.742	0.574***	0.739
13-15	0.478***	0.85	0.480***	0.848	0.479***	0.869	0.480***	0.866
16-18	0.338***	0.863	0.338***	0.862	0.338***	0.879	0.338***	0.877
19-21	0.503**	0.362*	0.504**	0.361*	0.504**	0.367*	0.505**	0.366*
22-24	0.316***	0.712	0.316***	0.708	0.317***	0.725	0.318***	0.722
25-27	0.353***	0.6	0.354***	0.597	0.353***	0.612	0.354***	0.609
28-30	0.371**	0.458	0.372**	0.458	0.371**	0.466	0.371**	0.467
31-33	0.200***	0.862	0.200***	0.868	0.201***	0.883	0.200***	0.893
34-36	0.352**	0.789	0.353**	0.79	0.353**	0.811	0.353**	0.816
37-39	0.402*	1.973*	0.403*	1.980*	0.404*	2.018*	0.403*	2.038*
40-42	0.120**	1.242	0.120**	1.261	0.120**	1.248	0.119**	1.275
43-45	0.133**	1.441	0.133**	1.467	0.133**	1.441	0.133**	1.474
46-48	0.331*	0.623	0.331*	0.635	0.330*	0.616	0.328*	0.631
49-51	0.0982*	1.244	0.0980*	1.266	0.0981*	1.242	0.0974*	1.272
52-54	0.106*	0.565	0.106*	0.578	0.106*	0.566	0.105*	0.585
55-57	0.251	0.664	0.251	0.673	0.25	0.669	0.25	0.682
58-60	0.145	3.01E-08	0.145	8.15E-09	0.144	4.08E-08	0.144	2.54E-08
60+	0.161***	0.731	0.161***	0.753	0.158***	0.684	0.157***	0.708
Age category (reference= 24-)								
25 - 34 years	0.875	0.521**	0.859	0.525**	0.833	0.512**	0.817	0.514**
35 - 44 years	0.703**	0.748	0.694**	0.772	0.665**	0.765	0.655***	0.792
45 - 54 years	0.521***	0.593*	0.513***	0.617*	0.496***	0.610*	0.488***	0.640*
55 - 64 years	0.272***	0.813	0.267***	0.868	0.261***	0.836	0.255***	0.894
Seasonal adjusted unemployment rate	0.882***	0.976	0.883***	0.972	0.882***	0.976	0.883***	0.971
Education (reference: primary school)								
VMBO	1.246	1.07	1.237	1.038	1.226	1.095	1.226	1.063
HAVO/VWO	1.21	1.194	1.215	1.105	1.214	1.253	1.231	1.177
MBO	1.429	0.954	1.428	0.909	1.4	1.021	1.413	0.984
HBO	1.461	1.246	1.449	1.184	1.443	1.216	1.448	1.184
WO	2.012**	1.103	1.989**	1.069	1.901**	0.954	1.894**	0.943
Female	0.939	0.867			0.967	0.885		
Children living at home	1.066	0.919			1.076	0.918		
Partner	1.02	0.757*			1.012	0.745*		
Origin (reference= Dutch background)								
First generation foreign, Western	0.593*	0.6	0.588*	0.621	0.595*	0.605	0.589*	0.622
First generation foreign, non-Western	0.404***	0.305***	0.406***	0.301***	0.424***	0.290***	0.428***	0.290***
Second generation foreign, Western	0.713*	0.749	0.721	0.745	0.722	0.755	0.73	0.756
Second generation foreign, non-Western	0.620*	0.601	0.617*	0.638	0.628*	0.62	0.629*	0.665
Natural logarithm of net household income	1.015	0.889***	1.024	0.856***	1.011	0.885***	1.02	0.852***
Profession (reference= higher academic or independent professional)								
Higher supervisory profession					0.748	0.512*	0.751	0.500*
Intermediate academic or independent professional					0.729	0.413**	0.716	0.416**
Intermediate supervisory or commercial					0.792	0.341***	0.782	0.345***
Other mental work					0.755	0.438**	0.745	0.432**
Skilled and supervisory manual work					0.77	0.383**	0.767	0.404**
Semi-skilled manual work					0.843	0.459*	0.837	0.493*
Unskilled manual work					0.630*	0.482*	0.625*	0.486*
Agrarian profession					0.825	0.968	0.829	1.001
Intercept	-1.531***	-2.132***	-1.573***	-2.116***	-1.194**	-1.333*	-1.217**	-1.325*
Random effect								
Variance	0.145*		0.149*		0.146*		0.146*	
Std. deviation	0.381*		0.386*		0.382*		0.382*	
Observations	19437		19437		19473		19473	
Akaike information criteria	10519.1		10516.4		10528.7		10525.7	
Bayesian information criterion	11157.1		11107.2		11292.8		11242.5	

* p<0.05, ** p<0.01, *** p<0.001

Table 8. Multilevel multinomial regression model. Logit coefficient estimates – by origin.

Group	Full sample				Dutch			
Exit outcome	Paid	Other	Paid	Other	Paid	Other	Paid	Other
Language problems	0.772*	1.304	0.770*	1.284	0.978	1.561*	0.976	1.554*
Age category (reference= 24-)								
25 - 34 years	0.803	0.476***	0.796	0.466***	0.918	0.643	0.912	0.636
35 - 44 years	0.649***	0.678	0.640***	0.655*	0.735*	0.812	0.727*	0.790
45 - 54 years	0.498***	0.581**	0.489***	0.562**	0.520***	0.616*	0.509***	0.599*
55 - 64 years	0.272***	0.883	0.266***	0.837	0.280***	0.927	0.275***	0.891
Seasonal adjusted unemployment rate	0.887***	0.972	0.887***	0.966	0.875***	0.956	0.875***	0.946
Education (reference: primary school)								
VMBO	1.344	1.113	1.338	1.116	1.524	0.861	1.506	0.853
HAVO/VWO	1.265	1.100	1.253	1.069	1.620	1.134	1.595	1.090
MBO	1.559*	0.983	1.548	0.964	1.756*	0.852	1.732*	0.828
HBO	1.554	1.218	1.548	1.210	1.696	1.100	1.675	1.087
WO	2.125**	1.086	2.120**	1.097	2.236**	0.825	2.209**	0.820
Natural logarithm of net household income	1.039	0.871***	1.039	0.868***	1.038	0.853***	1.039	0.849***
Unemployment duration (reference: 1-3 months)								
4-6	0.859	0.959			0.872	0.966		
7-9	0.760*	0.883			0.82	0.961		
10-12	0.534***	0.669			0.538***	0.817		
13-15	0.444***	0.781			0.466***	0.823		
16-18	0.310***	0.783			0.310***	0.75		
19-21	0.458***	0.325**			0.450**	0.387*		
22-24	0.287***	0.635			0.288***	0.66		
25-27	0.319***	0.533			0.345***	0.41		
28-30	0.332***	0.404*			0.351**	0.573		
31-33	0.178***	0.763			0.209***	0.969		
34-36	0.310**	0.687			0.299**	0.614		
37-39	0.352**	1.721			0.190**	2.143*		
40-42	0.104**	1.095			0.150**	0.964		
43-45	0.115**	1.262			0.0835*	1.123		
46-48	0.280*	0.533			0.324	0.547		
49-51	0.0816*	1.047			0.130*	0.956		
52-54	0.0872*	0.471			3.22E-09	0.742		
55-57	0.205*	0.548			0.346	0.873		
58-60	0.114*	6.68E-09			3.22E-09	3.14E-09		
60+	0.128***	0.620			0.0317***	0.915		
Months			0.938***	0.989			0.988	0.937**
Months^2			1.001	1			1	1.001
Months^3			1	1			1	1
Intercept	-1.795***	-2.270***	-1.664***	-2.220***	-1.832***	-1.934***	-1.684***	-1.816***
Random effect								
Variance	0.121		0.153*		0.102		0.124	
Std. deviation	0.348		0.391*		0.319		0.352	
Observations	19473		19473		14377		14377	
Akaike information criterion	10559.5		10540.6		8554.3		8526.9	
Bayesian information criterion	11087.3		10800.6		9061.8		8776.9	

* p<0.05, ** p<0.01, *** p<0.001

Table 8. Multilevel multinomial regression model. Logit coefficient estimates – by origin.

Group	First generation foreign				Second generation foreign			
Exit outcome	Paid	Other	Paid	Other	Paid	Other	Paid	Other
Language problems	0.900	1.508	0.850	1.513	0.765	2.23	0.77	2.178
Age category (reference= 24-)								
25 - 34 years	1.212	0.466	1.263	0.499	0.68	0.171*	0.672	0.175*
35 - 44 years	1.030	1.213	1.032	1.277	0.481	0.295	0.484	0.319
45 - 54 years	0.911	1.14	0.960	1.196	0.432	0.338	0.433	0.364
55 - 64 years	0.597	0.706	0.590	0.666	0.246**	0.79	0.247**	0.768
Seasonal adjusted unemployment rate	0.839	1.051	0.830	1.049	0.952	0.974	0.961	1.013
Education (reference: primary school)								
VMBO	0.664	1.832	0.650	1.639				
HAVO/VWO	0.231*	0.628	0.227*	0.598				
MBO	0.903	0.730	0.887	0.666				
HBO	0.886	1.593	0.817	1.490				
WO	0.980	1.843	0.980	1.789				
Natural logarithm of net household income	0.931	0.827**	0.931	0.831**	0.996	0.976	0.996	0.981
Unemployment duration (reference: 1-3 months)								
4-6	1.585	0.904			0.534	1.413		
7-9	0.654	0.839			0.576	0.722		
10-12	1.369	0.454			0.310*	4.00E-09		
13-15	0.533	0.855			0.466	1.006		
16-18	0.826	1.055			0.224	1.619		
19-21	1.546	6.32E-09			0.246	0.541		
22-24	0.956	1.059			0.137	0.607		
25-27	1.15	1.214			1.78E-09	1.418		
28-30	1.301	9.11E-09			1.80E-09	3.07E-09		
31-33	5.71E-09	0.682			0.189	3.16E-09		
34-36	0.524	0.684			0.429	1.724		
37-39	2.213	2.045			0.555	3.16E-09		
40-42	6.53E-09	1.272			1.79E-09	2.859		
43-45	5.65E-09	2.708			0.401	1.727		
46-48	0.849	5.23E-09			2.15E-09	1.881		
49-51	6.11E-09	5.04E-09			8.68E-10	6.413		
52-54	0.86	5.15E-09			1.99E-09	3.22E-09		
55-57	5.38E-09	5.25E-09			1.89E-09	3.32E-09		
58-60	5.22E-09	5.35E-09			1.053	3.55E-09		
60+	0.512	0.196			0.727	4.17E-09		
Months			1.001	1.047			0.877*	0.904
Months^2			1.000	0.998*			1.003	1.005
Months^3			1.000	1.000*			1.000	1.000
Intercept	-1.758	-3.491*	-1.790**	-3.669*	-1.435	-2.997*	-1.315	-2.991*
Random effect								
Variance	0.105		0.180		0.287		0.262	
Std. deviation	0.324		0.424		0.536		0.512	
Observations	3115		3115		1981		1981	
Akaike information criterion	1129.6		1093.0		949.8		919.4	
Bayesian information criterion	1534.5		1292.5		1268.6		1048.0	

* p<0.05, ** p<0.01, *** p<0.001

6 Conclusion and discussion

Many different models were estimated in this thesis. First, simple logistic models were estimated using different combinations of covariates. While the first estimated models suggested that having language problems has a significant negative effect on the probability of exiting unemployment to paid employment, adding controls for origin resulted in insignificance of this effect. Because of the confounding nature of origin, stratified models based on origin were also estimated, where the estimated effects of having language problems remained insignificant. To control for bias that could be a consequence of the fact that exits from unemployment without obtaining paid employment were recorded as right-censored observations, multinomial logistic regression models were estimated. These models treated exits from unemployment without obtaining paid employment as the competing risk of getting a paid job. Again, the models without controls for origin showed significant effects, while the models with controls for origin and the stratified models showed no significant effect of having language problems on the probability of exiting unemployment to paid employment. Lastly, a multilevel component was added to the regressions to allow recurrent unemployment spells and add unobserved heterogeneity. Similar to the simple- and competing risks regression, the effect of having language problems remained insignificant in the models with controls for origin and the stratified models.

In all of the models, the negative effect of having language problems on the probability of exiting unemployment to paid employment became insignificant when controls for origin were added or when the models were estimated for each origin-group separately. These results might be due to a phenomenon known as Simpson's Paradox (Blyth, 1972). Non-natives might face a lower baseline probability of exiting unemployment to paid employment while they have a higher risk of having language problems at the same time. The descriptive statistics do show that especially first generation foreigners have higher median unemployment durations. A higher percentage of first generation foreigners also reported to have language problems. Without controls for origin, a significant effect of language problems can then be estimated. When controls for origin are added, or the models are stratified, the estimated effects become insignificant. Not having language problems, but being a non-native predicts a low probability of leaving unemployment to paid employment, for example because of discrimination. That is, both natives and foreigners who report to have language problems do not face significantly lower probabilities of exiting unemployment to paid employment when compared to individuals from the same origin-group without language problems.

The main hypothesis of this study:

Individuals without any Dutch language related problems have a higher hazard rate of exiting unemployment to paid employment.

has to be rejected based on the results. No significant negative effect of having language problems on the probability of exiting unemployment to paid employment has been observed in this study. Language proficiency was thought to affect hiring rates and job search efficiency, but the results show that it does not affect the probability of exiting unemployment to paid employment.
The secondary hypothesis:

Foreign born immigrants without any Dutch language related problems have a higher hazard rate of exiting unemployment to paid employment.

has to be rejected as well. Foreign born immigrants without Dutch language problems do not face a higher hazard rate of exiting unemployment to paid employment. The stratified models show no significant effect of having language problems on the probability of exiting unemployment to paid employment for all origin based groups. Interestingly however, the competing risks- and multilevel multinomial models with controls for origin suggested that having problems with the Dutch language significantly increased the probability of exiting unemployment without obtaining paid employment. When models were estimated for each origin-group separately, this effect only remained significant for Dutch natives. This significant effect of having language problems should be interpreted with great caution, since exits from unemployment without obtaining paid employment cover a wide variety of exits. Not only could an individual become inactive or retired, he could also start a business or go to school. Further research could focus on the effect of language proficiency on the various other unemployment exits.

There are some possible explanations for the absence of a negative significant effect of having language problems on the probability of exiting unemployment to paid employment. First of all, Yao and Van Ours (2015) did not find any effect of language proficiency on immigrants wages in the Netherlands using the same data as this study. Their explanation of the lack of an effect of language proficiency on wage earnings was that English is a *de facto* lingua franca in the Netherlands. This means that the ability to speak Dutch is not a necessity in the Netherlands, because English is spoken and understood as well. This might also explain the lack of an effect of language proficiency on unemployment duration.
As already mentioned in the theoretical framework, immigrants tend to apply for 'bad' jobs and jobs for which native labour supply is low (Fullin & Reyneri, 2010; Kossoudji, 1988; Peri & Sparber, 2009; Premji, Duguay, Messing, & Lippel, 2010). While this was hypnotized

to lead to increased competition for job vacancies, it could also lead to diminished importance of Dutch language skills for immigrants. As a consequence, language might play no or only a limited role in the unemployment duration for those with language problems.[34]

Based on this study, job search effectiveness does not seem to be affected by language proficiency, since reduced language proficiency does not result in lower predicted probabilities of finding paid employment and lower unemployment durations. As discussed earlier however, job search effectiveness is constituted by a wide range of aspects, such as the speed of finding information about vacancies or the ability of writing appropriate application letters. Further research could be conducted to see whether having language problems does affect some specific aspects of job search effectiveness. A more experimental approach could for example be used to study the effect of language proficiency on the speed of finding suitable vacancies or the success of job interviews.

The lack of a significant effect might however not be a consequence of a true lack of a significant effect. The data or the methods used in this study could have resulted in bias results. First of all, biases could have occurred because of omitted control variables which were not available. For example, multiple studies have found that eligibility for unemployment benefits and benefit duration have a significant effect on the exit rates from unemployment (Bover et al., 2002; Carling, Edin, Harkman, & Holmlund, 1996; Katz & Meyer, 1990; Lalive, 2008). Other factors that could have affected unemployment duration are reservation wages (Holzer, 1984), which were mentioned earlier, health (Stewart, 2001) or personality traits (Kanfer, Wanberg, & Kantrowitz, 2001; Uysal & Pohlmeier, 2011). All of these omitted variables could not only directly affect unemployment duration, but they could also be correlated with language proficiency. For example, low language proficiency is related with a lower access to (mental) health services (Sentell, Shumway, & Snowden, 2007), or certain personality traits could be correlated with both job search effectiveness and (second) language proficiency because personality traits can have a direct effect on language learning (Dulay, Burt, & Krashen, 1982). Individuals of certain origins might also be more likely to have certain personality traits (Allik & McCrae, 2004) while those individuals at the same time could be more likely to face problems with the Dutch language (Beenstock, Chiswick, & Repetto, 2001). Because no control variables were used for these possible covariates, omitted-variables bias could have occurred. To avoid these biases, further research could focus on the collection of data related to the omitted variables or use a more experimental approach.

---

[34] For these 'bad' jobs, language skills could play only a very limited role in the hiring decision and employers might be willing to put in more effort to find individuals with low language skills for these jobs, therefore reducing the negative effects of low language skills on information search speed and effectiveness.

Aside from the possible problems as a consequence of omitted variables, the models could also have been wrongly specified. The stratified regression model estimates for example showed that only a very limited number of predictor variables seemed to significantly affect exit probabilities for non-natives, while almost all predictor variables significantly affected native exit probabilities. This could be a result of the fact that exit probabilities of non-natives are affected by different variables than the probabilities for natives.

The major problem of this study however is the limited sample size. Especially for the stratified samples, the number of observed events is low for both first- and second generation foreigner's strata. Because of the low number of observed events the probability of a rejection of the hypotheses even though the hypotheses are true (type II error) is high. Further research could therefore focus on the gathering of larger amounts of data to see if the results of this study hold when samples are larger.

As pointed out in many studies on language proficiency and labour market outcomes, reverse causality might cause significant bias (Yao & van Ours, 2015). The probability of exiting unemployment could affect language proficiency. Firstly, longer spells might reduce an individual's Dutch language skills because Dutch might be used and practiced less outside work environments for those who have trouble with the Dutch language. Secondly, unemployment reduces income. Learning and practicing Dutch might be complicated with reduced income. Further research could focus on using methods to address these issues such as Instrumental Variable (IV) techniques.

Data problems could occur because of measurement errors of language proficiency (MacIntyre, Noels, & Clément, 1997). The language proficiency variables are self-reported and measurement errors could therefore easily occur. As discussed earlier, different groups might have different subjective opinions what having problems with the Dutch language exactly means. Also, the used variables might not capture the full concept of language proficiency. More objective measures of language proficiency, such as standardized language tests or expert evaluation might therefore be used to measure language proficiency more accurately. Lastly, some groups are underrepresented in the LISS panel. Most important for this study is the fact that the LISS panel is conducted in Dutch only. Household were no adult member is capable of understanding the Dutch language are not included in the LISS panel.

7 Appendices

A Overview of used variables

Table A.1. Overview of used variables.

Variable	Description
Age category	Categorical variable measuring age.
Gender	Dummy variable. Equal to 0 if male and 1 if female.
Household income	Continuous variable. Measures net household income.
Education category	Categorical variable based on the Statistics Netherlands education categories.
Partner	Dummy variable. Equal to 1 if individual lives with a partner and 0 if not.
Children living at home	Dummy variable. Equal to 1 if individual has children living-at-home and 0 if not.
Occupation	Categorical variable of primary occupation. Of special interest are: Paid employment (1), Job seeker following job loss (4) and First time job seeker (5)
Profession	Categorical variable on the individual's profession. For unemployed individuals, the last profession is used, and for first time job-seekers, their future profession is used, when available. Examples of professions are: higher academic or independent professional or semi-skilled manual worker.

B Life tables

Table B.1. Life tables grouped exit outcome and language proficiency.

	Full sample					No language problems					Language problems				
		Exit outcome					Exit outcome					Exit outcome			
Interval	At risk	Paid job	Other	Total	Censored	At risk	Paid job	Other	Total	Censored	At risk	Paid job	Other	Total	Censored
1 - 2	1868	142	35	177	18	1588	119	27	146	17	280	23	8	31	1
2 - 3	1673	167	38	205	19	1425	153	27	180	16	248	14	11	25	3
3 - 4	1449	138	38	176	2	1229	118	29	147	2	220	20	9	29	0
4 - 5	1271	95	29	124	18	1080	85	25	110	16	191	10	4	14	2
5 - 6	1129	69	19	88	8	954	63	14	77	7	175	6	5	11	1
6 - 7	1033	72	19	91	9	870	61	16	77	7	163	11	3	14	2
7 - 8	933	65	19	84	8	786	60	17	77	5	147	5	2	7	3
8 - 9	841	42	21	63	8	704	39	21	60	6	137	3	0	3	2
9 - 10	770	41	12	53	6	638	34	11	45	6	132	7	1	8	0
10 - 11	711	32	15	47	7	587	29	13	42	7	124	3	2	5	0
11 - 12	657	22	11	33	7	538	17	8	25	6	119	5	3	8	1
12 - 13	617	21	9	30	7	507	14	9	23	7	110	7	0	7	0
13 - 14	580	25	5	30	11	477	18	5	23	10	103	7	0	7	1
14 - 15	539	17	9	26	6	444	15	6	21	5	95	2	3	5	1
15 - 16	507	8	13	21	3	418	7	10	17	3	89	1	3	4	0
16 - 17	483	10	7	17	2	398	9	5	14	2	85	1	2	3	0
17 - 18	464	8	7	15	8	382	5	4	9	8	82	3	3	6	0
18 - 19	441	13	9	22	6	365	12	7	19	6	76	1	2	3	0
19 - 20	413	15	3	18	7	340	12	3	15	6	73	3	0	3	1
20 - 21	388	11	4	15	5	319	11	4	15	5	0	0	0	0	0
21 - 22	368	10	3	13	9	299	7	2	9	8	69	3	1	4	1
22 - 23	346	2	1	3	2	282	0	1	1	1	64	2	0	2	1
23 - 24	341	4	6	10	5	280	4	4	8	3	61	0	2	2	2
24 - 25	326	13	6	19	4	269	12	6	18	4	57	1	0	1	0
25 - 26	303	4	5	9	2	247	3	5	8	2	56	1	0	1	0
26 - 27	292	7	2	9	4	237	5	2	7	3	55	2	0	2	1
27 - 28	279	5	3	8	6	227	4	3	7	5	52	1	0	1	1
28 - 29	265	7	2	9	2	215	6	2	8	2	50	1	0	1	0
29 - 30	254	5	4	9	4	205	5	3	8	4	49	0	1	1	0
30 - 31	241	2	2	4	6	193	2	2	4	5	48	0	0	0	1
31 - 32	231	1	5	6	6	184	1	5	6	6	0	0	0	0	0
32 - 33	219	2	5	7	3	172	2	2	4	1	47	0	3	3	2
33 - 34	209	4	1	5	6	167	3	0	3	4	42	1	1	2	2
34 - 35	198	1	1	2	0	160	1	1	2	0	0	0	0	0	0
35 - 36	196	1	6	7	6	158	1	4	5	4	38	0	2	2	2
36 - 37	183	9	3	12	1	149	8	3	11	1	34	1	0	1	0
37 - 38	170	3	5	8	2	137	2	5	7	2	33	1	0	1	0
38 - 39	160	3	7	10	2	128	2	7	9	2	32	1	0	1	0
39 - 40	148	2	4	6	5	117	1	3	4	4	31	1	1	2	1
40 - 41	137	0	6	6	3	109	0	5	5	3	28	0	1	1	0
41 - 42	128	0	1	1	0	0	0	0	0	0	27	0	1	1	0
42 - 43	127	2	1	3	3	101	2	1	3	3	0	0	0	0	0
43 - 44	121	0	2	2	4	95	0	2	2	4	0	0	0	0	0
44 - 45	115	1	5	6	3	89	1	3	4	3	26	0	2	2	0
45 - 46	106	1	2	3	2	82	0	2	2	2	24	1	0	1	0
46 - 47	101	3	0	3	0	78	2	0	2	0	23	1	0	1	0
47 - 48	98	0	1	1	2	76	0	1	1	2	0	0	0	0	0
48 - 49	95	2	3	5	1	73	2	3	5	0	22	0	0	0	1
49 - 50	89	0	2	2	0	68	0	2	2	0	0	0	0	0	0
50 - 51	87	1	2	3	1	66	1	2	3	1	0	0	0	0	0
51 - 52	83	0	1	1	2	62	0	1	1	1	21	0	0	0	1
52 - 53	80	0	2	2	3	60	0	2	2	2	20	0	0	0	1
53 - 54	75	0	0	0	2	56	0	0	0	2	0	0	0	0	0
54 - 55	73	1	0	1	2	54	1	0	1	1	19	0	0	0	1
55 - 56	70	2	1	3	1	52	2	0	2	1	18	0	1	1	0
56 - 57	66	0	1	1	3	49	0	0	0	2	17	0	1	1	1
57 - 58	62	0	1	1	1	47	0	1	1	1	0	0	0	0	0
59 - 60	60	0	0	0	3	45	0	0	0	2	15	0	0	0	1
60 - 61	57	1	0	1	1	43	1	0	1	1	0	0	0	0	0
61 - 62	55	2	2	4	0	41	2	2	4	0	0	0	0	0	0
62 - 63	51	0	0	0	2	37	0	0	0	2	0	0	0	0	0
63 - 64	49	0	0	0	1	0	0	0	0	0	14	0	0	0	1
64 - 65	48	0	0	0	1	0	0	0	0	0	13	0	0	0	1
65 - 66	47	0	0	0	1	35	0	0	0	1	0	0	0	0	0
66 - 67	46	1	1	2	3	34	1	0	1	3	12	0	1	1	0
67 - 68	41	0	0	0	1	30	0	0	0	1	0	0	0	0	0
68 - 69	40	0	0	0	2	29	0	0	0	1	11	0	0	0	1
69 - 70	38	0	0	0	1	28	0	0	0	1	0	0	0	0	0
70 - 71	37	1	0	1	0	27	1	0	1	0	0	0	0	0	0

71 - 72	36	0	1	1	2	26	0	1	1	1	10	0	0	0	1
72 - 73	33	0	1	1	1	24	0	1	1	1	0	0	0	0	0
73 - 74	31	0	0	0	1	22	0	0	0	1	0	0	0	0	0
74 - 75	30	0	1	1	0	0	0	0	0	0	9	0	1	1	0
75 - 76	29	1	1	2	0	21	1	0	1	0	8	0	1	1	0
76 - 77	27	1	0	1	1	20	1	0	1	1	0	0	0	0	0
77 - 78	25	0	0	0	1	18	0	0	0	1	0	0	0	0	0
78 - 79	24	0	0	0	1	17	0	0	0	1	0	0	0	0	0
79 - 80	23	1	0	1	1	16	1	0	1	0	7	0	0	0	1
80 - 81	21	0	0	0	1	15	0	0	0	1	0	0	0	0	0
81 - 82	20	0	2	2	0	14	0	1	1	0	6	0	1	1	0
82 - 83	18	0	0	0	1	13	0	0	0	1	0	0	0	0	0
83 - 84	17	0	0	0	1	12	0	0	0	1	0	0	0	0	0
86 - 87	16	0	0	0	1	11	0	0	0	1	0	0	0	0	0
89 - 90	15	0	1	1	1	10	0	1	1	1	0	0	0	0	0
92 - 93	13	1	0	1	0	8	1	0	1	0	0	0	0	0	0
93 - 94	12	0	0	0	1	7	0	0	0	1	0	0	0	0	0
96 - 97	11	0	0	0	2	6	0	0	0	1	5	0	0	0	1
101 - 102	9	0	0	0	1	5	0	0	0	1	0	0	0	0	0
102 - 103	8	0	1	1	0	0	0	0	0	0	4	0	1	1	0
107 - 108	7	1	0	1	0	0	0	0	0	0	3	1	0	1	0
110 - 111	6	0	1	1	0	4	0	1	1	0	0	0	0	0	0
118 - 119	5	0	1	1	0	3	0	1	1	0	0	0	0	0	0
123 - 124	4	0	0	0	1	0	0	0	0	0	2	0	0	0	1
127 - 128	3	0	1	1	0	2	0	1	1	0	0	0	0	0	0
144 - 145	2	0	1	1	0	1	0	1	1	0	0	0	0	0	0
159 - 160	1	0	0	0	1	0	0	0	0	0	1	0	0	0	1

C Unemployment duration specification

To assess whether alternative specifications of the unemployment duration, multiple regressions will be estimated. Model (1) specifies unemployment duration using the tri-monthly dummies. Model (2) uses the natural logarithm of the number of months of unemployment as a specification. In model (3)-(8) unemployment duration is represented by various polynomials.

Even while the AIC and BIC are slightly lower in some of the polynomial models, the model based on the tri-monthly dummies is used because it does allow for the most flexible estimation. For a more detailed discussion on time specification, see Singer and Willett (2003a).

Table C.1. Regression models odds ratio estimates. These models differ in their specification of the unemployment duration effect.

Model	1	2	3	4	5	6	7	8
Unemployment duration (reference = 1-3 months)								
4-6	0.781**							
7-9	0.630***							
10-12	0.410***							
13-15	0.322***							
16-18	0.219***							
19-21	0.311***							
22-24	0.187***							
25-27	0.199***							
28-30	0.199***							
31-33	0.103***							
34-36	0.176***							
37-39	0.192***							
40-42	0.0597***							
43-45	0.0680***							
46-48	0.162***							
49-51	0.0472**							
52-54	0.0526**							
55-57	0.122**							
58-60	0.0690**							
60+	0.0722***							
ln(months)		0.584***						
Months			0.944***	0.921***	0.902***	0.906***	0.910***	0.98
Months^2				1.001***	1.001***	1.001	1.001	0.991
Months^3					1.000*	1.000	1.000	1.000*
Months^4						1.000	1.000	1.000*
Months^5							1.000	1.000*
Months^6								1.000
Observations	19473	19473	19473	19473	19473	19473	19473	19473
Likelihood-ratio χ^2 (df)	452.6	365.6	400.2	427.5	433.1	433.2	433.2	438.3
Likelihood-ratio test ($p > \chi^2$)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Akaike information criteria	7097.5	7146.5	7111.9	7086.6	7083.0	7084.9	7084.9	7081.8
Bayesian information criterion	7262.9	7162.3	7127.7	7110.2	7114.5	7124.3	7124.2	7129.1

* p<0.05, ** p<0.01, *** p<0.001

D Reduced predictor variable models

These models use a reduced number of predictor variables to estimate the effect of language problems on the probability of exiting unemployment to paid employment for first- and second generation foreigners. To reduce the number of predictor variables, education is measured using a binary dummy variable, which equals 1 if the individual has had higher education. Instead of categorical variable measuring age, a continuous variable is used. Unemployment duration is measured using a second degree polynomial. Results are shown in table D.1. These estimates do not differ a lot from the estimates from the competing risk models in table 6. Most important, language proficiency has no significant effect on the probability of exiting unemployment to paid employment for both first- and second generation foreigners according to both models.

Table D.1. Reduced predictor variable multinomial logistic regression models.

	First generation		Second generation	
Exit outcome (paid employment or other)	Paid	Other	Paid	Other
Language problems	0.963	1.792	0.877	1.908
Age	0.987	1.014	0.972**	1.018
Seasonal adjusted unemployment rate	0.893	1.097	0.957	1.098
Higher education	1.272	1.704	1.874*	0.966
Natural logarithm of net household income	0.957	0.813***	1.02	0.965
Months	0.957*	0.981	0.901***	0.988
Months^2	1	1	1.001***	1
Observed events	72	37	74	30
Events per variable	10.29	5.29	10.57	4.29
Observations	3115		1981	
Likelihood-ratio χ^2 (df)	35.27		59.57	
Likelihood-ratio test ($p > \chi^2$)	0.00134		0.000000139	
Akaike information criteria	1081.4		912.8	
Bayesian information criterion	1178.1		1002.3	

* p<0.05, ** p<0.01, *** p<0.001

The same reduced models have been estimated for the multilevel multinomial model using Stata's GSEM. The results are shown in table D.2. These results are shown as logit coefficients, but they differ very little from the results of table D.1.

Table D.2. Reduced predictor variable multilevel multinomial logistic regression models.

Group	First generation		Second generation	
Exit outcome (paid employment or other)	Paid	Other	Paid	Other
Language problems	1.000	1.760	0.886	2.070
Age	0.984	1.009	0.965**	1.011
Seasonal adjusted unemployment rate	0.878	1.064	0.932	1.065
Higher education	1.329	1.765	2.034*	1.020
Natural logarithm of net household income	0.947	0.815**	1.031	0.980
Months	0.968	0.993	0.919**	1.002
Months^2	1.000	1.000	1.001**	1.000
Language problems	1.000	1.760	0.886	2.070
Intercept	-1.613	-4.129***	-1.181	-4.944**
Observed events	72	37	74	30
Events per variable	9	4.63	9.25	3.75
Observations	3115		1981	
Random effects				
Variance	0.238		0.383	
Std. deviation	0.488		0.619	
Akaike information criteria	1081.7		911.6	
Bayesian information criterion	1184.4		1006.7	

* p<0.05, ** p<0.01, *** p<0.001

8 References

Ahn, N., de la Rica, S., & Ugidos, A. (1999). Willingness to Move for Work and Unemployment Duration in Spain. *Economica*, *66*(263), 335–357. https://doi.org/10.1111/1468-0335.00174

Alba-Ramírez, A., Arranz, J. M., & Muñoz-Bullón, F. (2007). Exits from unemployment: Recall or new job. *Labour Economics*, *14*(5), 788–810. https://doi.org/10.1016/J.LABECO.2006.09.004

Aldashev, A., Gernandt, J., & Thomsen, S. L. (2008). Language usage, participation, employment and earnings Evidence for foreigners in West Germany with multiple sources of selection. *Labour Economics*, *16*, 330–341. https://doi.org/10.1016/j.labeco.2008.11.004

Allik, J., & McCrae, R. R. (2004). Toward a Geography of Personality Traits. *Journal of Cross-Cultural Psychology*, *35*(1), 13–28. https://doi.org/10.1177/0022022103260382

Allison, P. D. (1982). Discrete-Time Methods for the Analysis of Event Histories. *Source: Sociological Methodology*, *13*, 61–98.

Allison, P. D. (2014). *Event History and Survival Analysis*. 2455 Teller Road, Thousand Oaks California 91320 United States: SAGE Publications, Inc. https://doi.org/10.4135/9781452270029

Arntz, M., & Wilke, R. A. (2009). Unemployment Duration in Germany: Individual and Regional Determinants of Local Job Finding, Migration and Subsidized Employment. *Regional Studies*, *43*(1), 43–61. https://doi.org/10.1080/00343400701654145

Aycan, Z., & Berry, J. W. (1996). Impact of employment-related experiences on immigrants' psychological well-being and adaptation to Canada. *Canadian Journal of Behavioural Science*, *28*(3), 240–251.

Baumann, R., & Engelhardt, B. (2016). Crime and labour market turnover. *Applied Economics Letters*, *23*(7), 536–538. https://doi.org/10.1080/13504851.2015.1085633

Becker, G. S. (1964). *Human capital: A theoretical and empirical analysis, with special reference to education* (3rd ed.). Chicago, Illinois: The University of Chicago Press Books.

Beenstock, M., Chiswick, B. R., & Repetto, G. L. (2001). The Effect of Linguistic Distance and Country of Origin on Immigrant Language Skills: Application to Israel. *International Migration*, *39*(3), 33–60. https://doi.org/10.1111/1468-2435.00155

Beggs, J. J., & Chapman, B. J. (1990). Search efficiency, skill transferability and immigrant relative unemploymen...: EBSCOhost. *Applied Economics*, *22*(2), 249–260.

Berman, E., Lang, K., & Siniver, E. (2003). Language-skill complementarity: returns to immigrant language acquisition. *Labour Economics*, *10*(3), 265–290. https://doi.org/10.1016/S0927-5371(03)00015-0

Bird, A., & Dunbar, R. (1991). Getting the job done over there: Improving expatriate productivity. *National Productivity Review*, *10*(2), 145–156. https://doi.org/10.1002/npr.4040100204

Blackaby, D., Leslie, D., & Murphy, P. (1998). The ethnic wage gap and employment differentials in the 1990s: Evidence for Britain. *Economics Letters*, *58*, 97–103.

Blanchard, O. J., & Diamond, P. (1994). Ranking, Unemployment Duration, and Wages. *Source: The Review of Economic Studies*,

61(3), 417–434.
Blyth, C. R. (1972). On Simpson's Paradox and the Sure-Thing Principle. *Journal of the American Statistical Association*, *67*(338), 364–366. https://doi.org/10.1080/01621459.1972.10482387
Bover, O., Arellano, M., & Bentolila, S. (2002). Unemployment Duration, Benefit Duration and the Business Cycle. *The Economic Journal*, *112*(479), 223–265.
Budría, S., & Swedberg, P. (2015). The impact of language proficiency on immigrant's earnings. *Revista de Economía Aplicada*, *23*(67), 63–91.
C., J. (2018). Angela Merkel has two weeks to keep Germany's centre-right together - Showdown postponed. *The Economist*, *06*.
Carling, K., Edin, P.-A., Harkman, A., & Holmlund, B. (1996). Unemployment duration, unemployment benefits, and labor market programs in Sweden. *Journal of Public Economics*, *59*(3), 313–334. https://doi.org/10.1016/0047-2727(95)01499-3
Carlsson, M. (2010). Experimental Evidence of Discrimination in the Hiring of First-and Second-generation Immigrants. *LABOUR*, *24*(3), 263–278.
Chiswick, B. R., & Miller, P. W. (2003). The complementarity of language and other human capital: immigrant earnings in Canada. *Economics of Education Review*, *22*(5), 469–480. https://doi.org/10.1016/S0272-7757(03)00037-2
Chow, G. C. (1960). Tests of Equality Between Sets of Coefficients in Two Linear Regressions, *28*(3), 591–605.
Clausen, J., Heinesen, E., Hummelgaard, H., Husted, L., & Rosholm, M. (2009). The effect of integration policies on the time until regular employment of newly arrived immigrants: Evidence from Denmark ☆. *Labour Economics*, *16*, 409–417. https://doi.org/10.1016/j.labeco.2008.12.006
Corak, M. (1996). Measuring the duration of unemployment spells. *The Canadian Journal of Economics*, *29*(Special Issue: 1), 43–49.
Coviello, V., & Boggess, M. (2004). Cumulative incidence estimation in the presence of competing risks. *The Stata Journal*, *4*(2), 103–112.
De Vos, K. (2009). Panel attrition in LISS. Tilburg: CentERdata.
Delander, L., Hammarstedt, M., Månsson, J., & Nyberg, E. (2005). Integration of immigrants. The Role of Language Proficiency and Experience. *Evaluation Review*, *29*(1), 24–41. https://doi.org/10.1177/0193841X04270230
Détang-Dessendre, C., & Gaigné, C. (2009). Unemployment duration, city size, and the tightness of the labor market. *Regional Science and Urban Economics*, *39*(3), 266–276. https://doi.org/10.1016/J.REGSCIURBECO.2009.01.003
Di Paolo, A., & Raymond, J. L. (2012). Language Knowledge and Earnings in Catalonia. *Journal of Applied Economics*, *9*(1), 89–118. https://doi.org/10.1016/S1514-0326(12)60005-1
Dulay, H. C., Burt, M. K., & Krashen, S. D. (1982). *Language two*. Oxford University Press.
Dustmann, C., Fabbri, F., Preston, I., & Wadsworth, J. (2003). Labour market performance of immigrants in the UK labour market. *Economic Journal, Royal Economic Society*, *113*, 695–717. https://doi.org/10.1111/1468-0297.t01-1-00151
Eam Lim, H., & Bakar, N. (2004). Unemployment Duration of

Graduates of Universiti Utara Malaysia: The Impact of English Language Proficiency. *Malaysian Journal of Economic Studies*, *41*(1 & 2).

Esser, H. (2006). *Migration, language and integration*. Berlin.

Fagerland, M. W., & Hosmer, D. W. (2012). A generalized Hosmer–Lemeshow goodness-of-fit test for multinomial logistic regression models. *The Stata Journal*, *12*(3), 447–453.

Friedberg, R. M. (2000). You Can't Take It with You? Immigrant Assimilation and the Portability of Human Capital. *Journal of Labor Economics*, *18*(2), 221–251.

Frijters, P., Shields, M. A., & Price, S. W. (2005). Job search methods and their success: A comparison of immigrants and natives in the UK. *Economic Journal*. https://doi.org/10.1111/j.1468-0297.2005.01040.x

Fullin, G., & Reyneri, E. (2010). Low Unemployment and Bad Jobs for New Immigrants in Italy. *International Migration*, *49*(1), 118–147. https://doi.org/10.1111/j.1468-2435.2009.00594.x

Gangl, M. (2003). *Unemployment Dynamics in the United States and West Germany*. Heidelberg: Physica-Verlag HD. https://doi.org/10.1007/978-3-642-57334-7

Grand, C. Le, & Szulkin, R. (2002). Permanent disadvantage of gradual integration: Explaining the immigrant-native earnings gap in Sweden. *Labour*. https://doi.org/10.1111/1467-9914.00186

Guo, S. (2010). *Survival Analysis* (1st ed.). Oxford: Oxford University Press.

Han, A., & Hausman, J. A. (1990). Flexible parametric estimation of duration and competing risk models. *Journal of Applied Econometrics*, *5*, 1–28.

Hannan, C. (1999). *Beyond Networks: Social Cohesion' and Unemployment Exit Rates*.

Hausman, J., & McFadden, D. (1984). Specification tests for the multinomial logit models. *Econometrica*, *52*(5), 1219–1240.

Heckman, J. J., & Borjas, G. J. (1980). Does Unemployment Cause Future Unemployment? Definitions, Questions and Answers from a Continuous Time Model of Heterogeneity and State Dependence. *Economica*, *47*(187), 247. https://doi.org/10.2307/2553150

Holzer, H. (1984). *Black Youth Nonemployment: Duration and Job Search*. Cambridge, MA. https://doi.org/10.3386/w1276

Hosmer, D. W., & Lemesbow, S. (1980). Goodness of fit tests for the multiple logistic regression model. *Communications in Statistics - Theory and Methods*, *9*(10), 1043–1069. https://doi.org/10.1080/03610928008827941

Ipsos MORI. (2016). Concern about immigration rises as EU vote approaches. Retrieved July 10, 2018, from https://www.ipsos.com/ipsos-mori/en-uk/concern-about-immigration-rises-eu-vote-approaches

Jenkins, S. P. (2005). *Survival Analysis*. Unpublished manuscript.

Jones, S. R. G. (1988). The Relationship Between Unemployment Spells and Reservation Wages as a Test of Search Theory. *The Quarterly Journal of Economics*, *103*(4), 741. https://doi.org/10.2307/1886073

Kanfer, R., Wanberg, C. R., & Kantrowitz, T. M. (2001). Job Search and Employment: A Personality-Motivational Analysis and Meta-Analytic Review. *Journal of Applied Psychology*, *86*(5), 837–855.

https://doi.org/10.1037//0021-9010.86.5.837
Kaplan, E. L., & Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. *Source Journal of the American Statistical Association*, *53*(282), 457–481.
Katz, L. F., & Meyer, B. D. (1990). The impact of the potential duration of unemployment benefits on the duration of unemployment. *Journal of Public Economics*, *41*(1), 45–72. https://doi.org/10.1016/0047-2727(92)90056-L
Kee, P. (1995). Native-immigrant wage differentials in the Netherlands: Discrimination? *Oxford Economic Papers*. https://doi.org/10.1093/oxfordjournals.oep.a042172
Kettunen, J. (1997). Education and Unemployment Duration. ~ *Pergamon Economics of Education Review*, *16*(2), 163–170.
Kogan, I. (2004). Last Hired, First Fired? The Unemployment Dynamics of Male Immigrants in Germany. *European Sociological Review*, *20*(5), 445–461. https://doi.org/10.1093/esr/jch037
Kossoudji, S. A. (1988). English Language Ability and the Labor Market Opportunities of Hispanic and East Asian Immigrant Men. *Source Journal of Labor Economics*, *6*(2), 205–228.
Lalive, R. (2008). How do extended benefits affect unemployment duration? A regression discontinuity approach. *Journal of Econometrics*, *142*(2), 785–806. https://doi.org/10.1016/J.JECONOM.2007.05.013
Lancee, B. (2010). The Economic Returns of Immigrants' Bonding and Bridging Social Capital: The Case of the Netherlands. *The International Migration Review*, *44*(1), 202–226.
Layard, R., Nickell, S., & Jackman, R. (2005a). Job Search: the Duration of Unemployment. In *Unemployment: Macroeconomic Performance and the Labour Market* (2nd ed., pp. 216–282). Oxford: Oxford University Press.
Layard, R., Nickell, S., & Jackman, R. (2005b). Mismatch: the Structure of Unemployment. In *Unemployment* (pp. 285–333). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199279166.003.0006
Leslie, D., & Lindley, J. (2001). The Impact of Language Ability on Employment and Earnings of Britain's Ethnic. *Economica*, *68*(272), 587–606.
Lynch, L. M., & Zemsky, R. (1995). *EQW National Employers Survey: First Results.* Philadelphia.
MacIntyre, P. D., Noels, K. A., & Clément, R. (1997). Biases in Self-Ratings of Second Language Proficiency: The Role of Language Anxiety. *Language Learning*, *47*(2), 265–287. https://doi.org/10.1111/0023-8333.81997008
Mcintosh, S., & Vignoles, A. (2001). Measuring and assessing the impact of basic skills on labour market outcomes. *Oxford Economic Papers*, *3*, 453–481.
Mckee-Ryan, F. M., Wanberg, C. R., & Kinicki, A. J. (2005). Psychological and Physical Well-Being During Unemployment: A Meta- Analytic Study. *Journal of Applied Psychology*, *90*(1), 53–76. https://doi.org/10.1037/0021-9010.90.1.53
Mcmanus, W., Gould, W., & Welch, F. (1983). Earnings of Hispanic Men: The Role of English Language Proficiency. *Source Journal of Labor Economics*, *1*(2), 101–130.
McQuaid, R. W. (2006). Job search success and employability in local

labor markets. *The Annals of Regional Science*, *40*(2), 407–421. https://doi.org/10.1007/s00168-006-0065-7
Meyer, B. D. (1990). Unemployment Insurance and Unemployment Spells. *Econometrica*, *58*(4), 757–782. https://doi.org/10.2307/2938349
Narendranathan, W., & Stewart, M. B. (1993). Modelling the Probability of Leaving Unemployment: Competing Risks Models with Flexible Base-Line Hazards. *Applied Statistics*, *42*(1), 63–83. https://doi.org/10.2307/2347410
OECD. (2014). *Is migration good for the economy* (Migration Policy Debates No. May).
Paul, P., Pennell, M. L., & Lemeshow, S. (2013). Standardizing the power of the Hosmer-Lemeshow goodness of fit test in large data sets. *Statistics in Medicine*. https://doi.org/10.1002/sim.5525
Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, *49*(12), 1373–1379. https://doi.org/10.1016/S0895-4356(96)00236-3
Peri, G., & Sparber, C. (2009). Task Specialization, Immigration, and Wages. *American Economic Journal: Applied Economics*, *1*(3), 135–169. https://doi.org/10.1257/app.1.3.135
Premji, S., Duguay, P., Messing, K., & Lippel, K. (2010). Are immigrants, ethnic and linguistic minorities over-represented in jobs with a high level of compensated risk? Results from a Montréal, Canada study using census and workers' compensation data. *American Journal of Industrial Medicine*, *53*(9), 875–885. https://doi.org/10.1002/ajim.20845
Saarela, J., & Finnäs, F. (2003). Unemployment and native language: the Finnish case. *Journal of Socio-Economics*, *32*, 59–80. https://doi.org/10.1016/S1053-5357(03)00007-6
Sentell, T., Shumway, M., & Snowden, L. (2007). Access to Mental Health Treatment by English Language Proficiency and Race/Ethnicity. *J Gen Intern Med*, *22*(2), 289–293. https://doi.org/10.1007/s11606-007-0345-7
Sides, J., & Citrin, J. (2007). European Opinion About Immigration: The Role of Identities, Interests and Information. *British Journal of Political Science*, *37*(03), 477. https://doi.org/10.1017/S0007123407000257
Sims Peterson, M. (2009). Personnel interviewers' perceptions of the importance and adequacy of applicants' communication skills. https://doi.org/10.1080/03634529709379102
Singer, J. D., & Willett, J. B. (1993). It's About Time: Using Discrete-Time Survival Analysis to Study Duration and the Timing of Events. *Journal of Educational Statistics Summer*, *18*(2), 155–195.
Singer, J. D., & Willett, J. B. (2003a). Extending the Discrete-Time Hazard Model. In *Applied Longitudinal Data Analysis* (pp. 407–467). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195152968.003.0012
Singer, J. D., & Willett, J. B. (2003b). Fitting Basic Discrete-Time Hazard Models. In *Applied Longitudinal Data Analysis* (pp. 357–406). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195152968.003.0011
Steele, F. (2005). *Event History Analysis* (NCRM Methods Review

Papers No. 004). Bristol.

Steele, F., Diamond, I., & Wang, D. (1996). The determinants of the duration of contraceptive use in China: a multilevel multinomial discrete hazards modeling approach. *Demography*, *33*(1), 12–23.

Stewart, J. M. (2001). The impact of health status on the duration of unemployment spells and the implications for studies of the impact of unemployment on health status. *Journal of Health Economics*. https://doi.org/10.1016/S0167-6296(01)00087-X

The Economist. (2016, June). Explaining the Brexit vote - The immigration paradox. *The Economist*.

Thoms, P., Mcmasters, R., Roberts, M. R., & Dombkowski, D. A. (1999). Resume characteristics as predictors of an invitation to interview. *JOURNAL OF BUSINESS AND PSYCHOLOGY*, *13*(3).

Trejo, S. J. (1997). Why Do Mexican Americans Earn Low Wages? *Source: Journal of Political Economy*, *105*(6), 1235–1268. https://doi.org/10.1086/516391

Uhlendorff, A. ;, & Zimmermann, K. F. (2006). *Unemployment Dynamics among Migrants and Natives* (Discussion Papers No. 2299).

United Nations. (2017). *International Migration Report 2017*. New York.

Urwin, P., & Shackleton, J. R. (1999). Search methods and transitions into employment and inactivity. *International Journal of Manpower*, *20*(3/4), 189–237. https://doi.org/10.1108/01437729910279153

Uysal, S. D., & Pohlmeier, W. (2011). Unemployment duration and personality. *Journal of Economic Psychology*, *32*(6), 980–992. https://doi.org/10.1016/J.JOEP.2011.03.008

Vittinghoff, E., & McCulloch, C. E. (2007). Relaxing the Rule of Ten Events per Variable in Logistic and Cox Regression. *American Journal of Epidemiology*, *165*(6), 710–718. https://doi.org/10.1093/aje/kwk052

Wunsch, G. (2007). Confounding and control. *Demographic Research*, *16*(4), 97–120. https://doi.org/10.4054/DemRes.2007.16.4

Yao, Y., & van Ours, J. C. (2015). Language skills and labor market performance of immigrants in the Netherlands. *Labour Economics*, *34*, 76–85. https://doi.org/10.1016/J.LABECO.2015.03.005