Return Migration - A Duration Analysis

Sanne Schroll
CEBR and SDU
PRELIMINARY VERSION

September 1, 2008

1 Introduction

Most of the economic literature on migration views the migration decision as permanent. However, several studies have shown that more and more migrants either return to their home country after a period abroad or perhaps migrate onward to a third country, see e.g. Dustmann and Weiss (2007). The economic literature on migration is therefore now becoming more differentiated, acknowledging that migration can be both a permanent and a temporary decision and that immigrants who emigrate can be both return and onward migrants. Since the duration of the migration is very important to the economic effects of migration, it is crucial to obtain knowledge of the duration of the migrants stay in the receiving country. Further, it is very important to see whether there are any differences between migrants with long durations compared to migrants with shorter durations.

This paper will present an empirical duration analysis of migrants migrating into Denmark between 1980 to 2005. Based on data from Statistics Denmark I find the duration time for immigrants in Denmark. I include all individuals aged 15 or above immigrated to Denmark in this period. Surely some of the immigrants have not yet left Denmark by the end of the period, so the data are right censored. I find the hazard rate for immigrants, showing at what rate immigrants will emigrate.

To analyze whether there are any differences between migrants with long and short durations I find the effects on the Hazard rate from years of education, wage, tenure, unemployment, age, sex, and marital status.

I expect that the duration time for immigrants with higher levels of skills will be shorter than for immigrants with lower levels of skills. There are two reasons for this. First, individuals with higher levels of education are assumed

to have a higher income and are therefore able to accumulate wealth faster than individuals with lower incomes, this means that the highly educated will be able to return home sooner. Second, since the tax level for especially highly skilled individuals (i.e. individuals with high incomes) is relatively higher in Denmark than in similar countries, such as Germany and UK, they are assumed to have higher incentives for onward migration. Dustmann (2003) shows that when the optimal duration time for the immigrant might decrease if the wage differential between the home country and the receiving country grows larger.

This brings out another question - Does the highly education emigrants from Denmark return to their home country or do they migrate onward to another country? In the data from Statistics Denmark it is possible to see whereto the migrants move. I therefore also analyze whether emigrating immigrants are return migrants, merely moving home, or onward migrants, seeking their fortune in another country. Again I expect to see that migrants with higher level of education, and therefore probably also higher levels of income, will account for most of the onward migration, while immigrants with lower levels of education will account for most of the return migration. Further, I also analyze whether return and onward migrants differ in any of the other explanatory variables.

The rest of the paper is organized as follows. Section 2 presents the Empirical Hypotheses, and section 3 presents the data from Statistics Denmark. In section 4 I present the Econometric Model, and in section 5 the results are outlined. Finally, section 6 concludes.

2 Empirical Hypotheses

This paper focuses on the duration of immigrants in Denmark in the period 1980-2005. I want to analyze whether there are any differences between migrants with long and short durations, and between return and onward migrants.

First hypothesis: *Migrants with higher skills (longer education, longer tenure, higher wages) will have a shorter duration than migrants with lower skills.*

This hypothesis is based upon the assumption that, individuals with higher skills (e.g. higher levels of education and tenure) are assumed to have a higher income and are therefore able to accumulate wealth faster than individuals with lower incomes, this means that the highly skilled will be able to return home sooner.

Second hypothesis: *Migrants with higher skills will have a higher rate of*

onward migration than migrants with lower skills.

This hypothesis is based upon the assumption that since the tax level for especially highly skilled individuals (i.e. individuals with high incomes) is relatively higher in Denmark than in similar countries, such as Germany and UK, they are assumed to have higher incentives for onward migration. Dustmann (2003) shows that the optimal duration time for immigrants might decrease if the wage differential between the home country and the receiving country increases.

Third hypothesis: *Migrants who become unemployed will sooner return home or migrate onward to a third country.*

Several papers (See e.g. Nekby (2006), Bratsberg et al. (2007)) shows that migrants tends to out migrate (either home or onward) when the wage in the country decreases, which could be due to unemployment in the host country.

Fourth hypothesis: *The older a migrant becomes and if the migrant gets married, the duration is increasing.*

There are also several papers showing that the most mobile migrants are young and single, since they have the lowest migration costs. (See e.g. Carrington et al., 1996)

Final hypothesis: *Migrants with a Danish background tend to have the highest duration time.*

The data set includes both Danes and foreigners, and I assume that Danes migrating into Denmark will tend to stay permanently compared to other immigrants.

All these hypotheses are analyzed below and the results are reported in section 5.

3 Data

I use data from two data bases from Statistics Denmark. First, the Integrated Data base of Labour Market Research (IDA) which holds information on all Danish residents aged 15 or more in the period 1980-2005. Second, the Danish Data base of Migration which holds information on all migrations in and out of Denmark. I merge the informations from these two data bases into one data set, which then holds information on each individual migrated into Denmark once or more in the period 1980-2005.

The data set is converted so that each individual has one observation for each year of his duration in Denmark, i.e. if a person migrates into Denmark in 1985 and out again in 1990 then there will be 6 observations for this individual for this spell. In table 1 it can be seen that there are 885,720

individual immigrants in the period, and this leads to a final Migration Data set with more than 5 million observations. individuals with a duration for more than 19 years have been equalized to have a duration for 19 years.

The explanatory variables comes from the IDA data base. I have included age, age^2, and the individuals years of schooling ($education$) is included. (Please note that this is self reported for immigrants who has taken their education outside Denmark.) Also the individuals $Tenure$ as the number of years employed in Denmark is included. Further, a number of dummy variables is included, man for the sex of the individual, $married$, for the marital status of the individual, $Danish$ which indicates whether the individual is Danish or has another nationality (descendants from immigrants are included as other nationality). I also include three dummies for the individuals occupational situation, $Self\ employed$, $worker$, or $unemployed$. Also $income$, which is the net income from the Danish Tax Registration each year is included as four dummies: $income1$ if income is between 0 and 10,000 Danish Kroner, $income2$ if income is between 10,000 and 100,000 Danish Kroner, $income3$ if income is between 100,000 and 1 million Danish Kroner, and $income4$ if income exceeds 1 million. (This will be altered soon, so that income will just be included as one continuous variable). Table 2 and 3 presents summary statics and distributions for the explanatory variables.

The analysis is executed in Gauss, where I begin by drawing a 5 per cent sample of the data set. The number of observations is then 292,020. I follow Kiefer (1990) and drop the last spell of each right censored spells, and the final number of observations is then 271,683.

4 Duration Analysis - The Econometric Model

Duration analysis has its origin in Survival analysis, where the duration of interest is the survival time of a subject. However, in Economics duration analysis has recently been more and more applied to analyze the duration time of different variables. Such studies include the duration of unemployment, see e.g. Lalive et al. (2006), the duration of new firms life time, see e.g. Mata et al. (1994), the impact of immigrants on the job duration for natives, see e.g. Malchow-Møller et al. (2007) and several others.

Duration analysis has also been used to analyze the duration of migration and return migration, see. e.g. Dustmann (2003). Migration duration data are characterized by a number of observations being right censored, since when we observe the migrants coming into the country during some period, not all will have left when we end our observation period. So we have both data for individuals with a well defined duration time in the country, and

data for individuals who have migrated in to the country and who are still there. For these individuals we do not know whether they will stay for ever, or leave the country at some future date. The duration of these individuals are right censored. It is therefore natural to use duration analysis which allow for right censored data.

First, I estimate a single risk duration model, where the only possibility for a spell to end is the migrant leaving the country. Later, I estimate a competing risk duration model, where the spell can end for two reasons, either the migrant migrates back to his home country, or he migrates onward to a third country, i.e. return and onward migration.

However, let me begin by introducing the theory behind the single risk duration model. I will present the extended model here, where I control for heterogeneity, however, in the next section where I present the results, I will include results for both a basic model with out any controls and then for the model presented here.

Even though my data are quite extensive I believe that there will always be some heterogeneity left, since not all characteristics can be measured, e.g. the urge to travel and see new places and experience new countries and cultures is very difficult to measure, and will probably influence the duration in a set of migration data. I capture these unobserved migrant characteristics by specifying a mixed proportion hazard model for return migration:

$$\theta = (t|x_t, v) = \lambda(t) \cdot \exp(\beta x_t + v) \tag{1}$$

Where the first term $\lambda(t)$ is the baseline hazard capturing the time dependence in return migration. That is for each t, $\lambda(t)$ is the instantaneous rate of leaving per unit of time. The second term, $\exp(\beta x_t + v)$ is the systematic part given the observed time varying characteristics, x_t, and the unobserved characteristics, v.

Given that my data set is annual, my duration variable, T, will be grouped in $K+1$ intervals $\{[0, t_1), [t_1, t_2), ..., [t_k, \infty)\}$. The econometric specification surely has to take this into account. I therefore follow Kiefer (1990) and define the interval specific survival rate as follows:

$$\begin{aligned} \alpha_k &= P(T \geq t_k | T \geq t_{k-1}, x_k, v) \\ &= \exp\left[-\int_{k-1}^{k} \theta(t|x_k, v)dt\right] \\ &= \exp\left[-\exp(\beta x_k + v)\Lambda_k\right] \end{aligned} \tag{2}$$

where $\Lambda_k = \int_{k-1}^{k} \lambda(t)dt$ and $\alpha_t = \exp\left[-\exp\left(\beta x_k + v\right)\Lambda_k\right]$

The probability that a migration spell ends in the $k'th$ interval is given by the conditional probability of a migration spell ending at that interval times the probability that the spell has survived until this interval, so a spell ending in interval k, will contribute to the likelihood function with $(1-\alpha_k)\Pi_{j=1}^{k-1}\alpha_j$. A right censored spell will contribute with $\Pi_{j=1}^{k}\alpha_j$. The contribution to the likelihood function from a migration spell ending in interval k, can be written as:

$$L = (t|x_1, ..., x_k, v) = (1-\alpha_k)^{dR}\alpha_k^{(1-dR)} \cdot \Pi_{j=1}^{k-1}\alpha_j \tag{3}$$

where dR is a return dummy. If the spell is right censored $dR = 0$

In many duration analyses a functional form is imposed on the baseline hazard function, but here I allow the baseline hazard to be flexible by simply estimating the interval specific baseline parameters, Λ_k.

Unobserved heterogeneity is as mentioned specified by the stochastic variable, v, with the CDF denoted as $F(v)$. It is assumed that the v can only take two values, where one of the support points in the hazard function is normalized to zero. This specification is flexible and widely applied (see e.g. van der Berg, 2001). The complete contribution to the likelihood function from a migration spell is therefore given by

$$\begin{aligned} L &= \int_{\underline{v}}^{\overline{v}} L(t|x_1, ..., x_k, \beta, v)dF(v) \\ &= L(t|x_1, ..., x_k, \beta, \overline{v})P(\overline{v}) \\ &\quad + L(t|x_1, ..., x_k, \beta, \underline{v})(1-P)(\underline{v}) \end{aligned} \tag{4}$$

Van den Berg (2001) includes more details on this class of mixture distributions in duration models.

Now, let's move on to the empirical results.

5 Results

The empirical hazard for the immigrants in Denmark from 1980 - 2005 is presented in table 4 and figure 1. The probability of leaving Denmark is decreasing in the number of years the migrant has already stayed in Denmark.

Table 5 includes the results for the two estimations. Migrationbasis refers to the basic model, where there is not controlled for heterogeneity, and migration refers to the extended model. It is clear that there is no significant

difference between the results for the two model specifications, so I will only refer to the coefficients form the latter model.

Most of the variables have significant coefficients, only age^2 and the third *income* dummy are insignificant. The coefficient for *man* is positive, indicating that men have a higher probability for out-migration than women. For *age* the coefficient is negative, so the older you are the probability for out-migration is decreasing. If you are married or has Danish nationality you also have a smaller probability of out-migration. If you have a low income your probability of out-migration is also lower, but if your income is high your probability of out-migration is increasing. All three occupational variables have negative coefficients, so you have lower probability of migrating in either category compared to the rest (i.e. people outside the labour market, students, etc.). The longer your education the lower is your probability of out-migration, and the same is true for tenure.

In relation to the five hypotheses mentioned above I can verify the hypotheses four and five; the migrants duration is increasing as he gets older, if he gets married and if his nationality is Danish. Also the first hypothesis is verified, the higher education the migrant has the longer he stays.

I have not yet divided the analysis into onward and return migration, so I can not answer whether hypotheses two and three will be verified. However, I can see that the higher your income (when you are in the high end of the income range) you will be more likely to migrate out of the country, this could be interpreted as in favor for the second hypotheses that the higher skills the more likely to onward migrate. But since the evidence from years of *education* points in the other direction I will have to do more work in order to say more about the second hypothesis. The third hypothesis is actually rejected, since the coefficient for *unemplyed* is negative, meaning that unemployment decreases your probability to migrate. However, I want to include the occupational variable in a different way, so I will also come back to this hypothesis in a later version of the paper.

6 Conclusion

This version of the paper is very preliminary. I will therefore not conclude more that I already did in the last section where I go through all the hypotheses.

In stead let me give a brief review of the work I intend to do further on this paper. I have still some trouble with the data, so I hope to be able to solve these problems very soon. Then I want to alter some of the explanatory variables, so that for instance income is included as a continuous

variable instead of four dummies. Instead of including three labour market dummies (so that I have very few omitted observations) I will only include self employed and unemployed so that I can compare to workers. Further, I want to extend the analysis to a competing risk duration model so that I can distinguish between return and onward migration.

References

[1] Bratsberg, B., O. Raaum, and K. Sørlie, 2007, Foreign-Born Migration To and From Norway, *International Migration and Economic Development, World Bank, Özden (Editor), pp. 259-290.*

[2] Carrington, W. J., E. Detragiache, and T. Vishwanath, 1996, Migration with Endogenous Moving costs, *The American Economic Review, Vol.86, No. 4, pp. 909-930,* September 1996.

[3] Dos Santos, M. D. and F. Postel-Vinay, 2003, Migration as a source of growth: the perspective of a developing country, *Journal of Population Economics, Vol. 16, issue 1, pp. 161-175,* February 2003

[4] Dustmann, C., 2003, Return Migration, wage differentials, and the optimal migration duration, *European Economic Review, Vol. 47, issue 2, pp.353-369*

[5] Dustman, C. and Y. Weiss, 2007, Return Migration: Theory and Empirical Evidence from the UK, *British Journal of Industrial Relations, Vol.. 45, issue 2, pp. 236-256,* June 2007

[6] Kiefer, N.M., 1990, "Econometric Methods for Grouped Duration Data" in J. Hartog, G. Rider, and J. Theeuwes (eds.), *Panel Data and Labour Market Studies,* North Holland, Amsterdam

[7] Lalive, R., J. Van Ours, J. Zweimüller, 2006, How Changes in the Financial Incentives Affects the Duration of Unemployment, *Review of Economic Studies, Vol. 73, issue 4, pp. 1009-1038,* October 2006.

[8] Mata, J. and P. Portugal, 1994, Life Duration of New Firms, *Journal of Industrial Economics, Vol. 42, No. 3, pp.227-245,* September 1994

[9] Munch, J.R., M. Rosholm, and M. Svarer, 2006, Home ownership, Job Duration and Wages, *IZA DP No. 2110*, May 2006

[10] Malchow-Møller, N., J.R. Munch, and J.R. Skaksen, 2007, What happens to the Employment of Native Co-Workers when Immigrants are Hired, *Study no. 17, The Rockwool Foundation Research Unit,* Copenhagen 2007

[11] Nekby, Lena, 2006, The Emigration of Immigrants, Return vs. Onward Migration: Evidence from Sweden, *Journal of Population Economics, Vol. 19, Number 2, pp. 197-226,* June 2006

[12] van der Berg, 2001, Duration Models: Specifications, Identifications, and Multiple Durations, in J.J. Heckman and E. Leamer (eds.), *Handbook of Econometrics, Vol. V,* North Holland, Amsterdam, 2001

Tables:

Table 1: Migration Spell Statistics	
Number of persons	885,720
Number of spells	1,005,752
Persons with more than one spell (share)	0.1193
Mean duration of spells (years)	8.03
Proportion of spells:	
right-censored spells	83.35
emigration to birth country	8.48
emigration to last home country	3.11
emigration to other country	5.05

Table 2: Summary Statistics		
	Mean	Stdv.
Male	0.491515	0.499928
Age	35.85879	11.818736
Married	0.520559	0.4995772
Danish	0.545532	0.4979226
INC1 (< 10 tkr)	0.039108	0.1938522
INC2 (10-100 tkr)	0.185689	0.3888557
INC3 (100t - 1 mill kr)	0.429287	0.4949745
INC4 (> 1 mill kr)	0.002619	0.0511128
Self employed	0.045877	0.2092188
Worker	0.550203	0.4974733
Unemployed	0.077221	0.2669418
Years of education	10.35369	5.56608
Tenure	3.47861	4.67407

Table 3: Distribution of explanatory variables	
	Percent
Male	49.15
Married	52.06
Danish	54.55
INC1 (< 10 tkr)	3.91
INC2 (10-100 tkr)	18.57
INC3 (100t - 1 mill kr)	42.93
INC4 (> 1 mill kr)	0.26
Self employed	4.59
Worker	55.02
Unemployed	7.72
Years of education:	
0	18.17
8	7.67
9	6.81
10	5.18
12	35.34
14	4.88
16	13.77
18	7.71
20	0.48

Table 4: Empirical Hazard
0.19363054
0.09172306
0.06074299
0.04804159
0.03779811
0.03221706
0.02633660
0.02329797
0.02190427
0.01621533
0.01727247
0.01254348
0.01677616
0.01094355
0.01165644
0.01044803
0.01105088
0.00638350
0.00855205

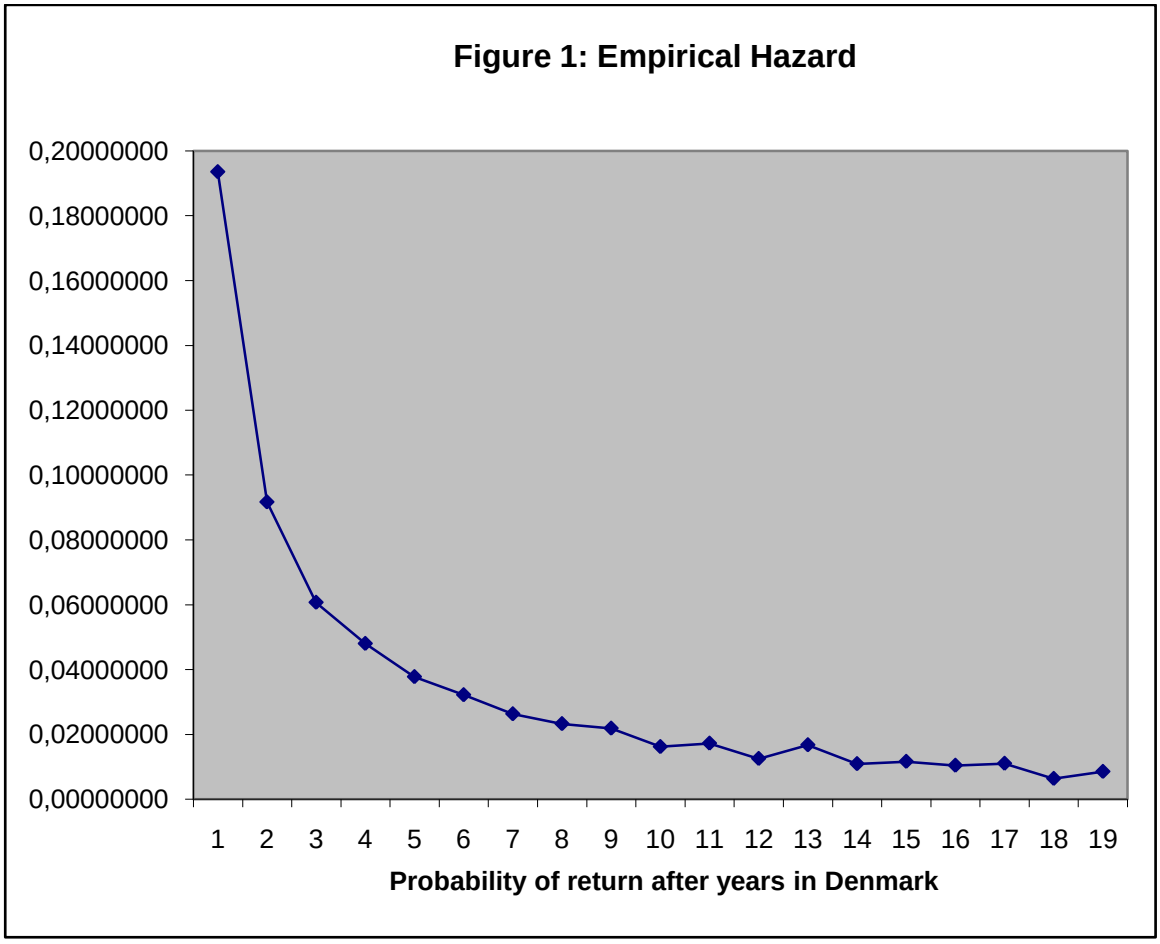
Figure 1: Empirical Hazard
0,20000000
0,18000000
0,16000000
0,14000000
0,12000000
0,10000000
0,08000000
0,06000000
0,04000000
0,02000000
0,00000000
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
Probability of return after years in Denmark

Table 5: Estimation Results		
	migrationbasis	migration
Male	0.14460	0.14000
	(9.22)***	(8.14)***
Age	-0.01656	-0.01348
	(-4.97)***	(-3.75)***
Age^2	0.00005	-0.00001
	(1.17)	(-0.30)
Married	-0.49690	-0.52880
	(-27.38)***	(-26.64)***
Danish	-0.24160	-0.28630
	(-12.74)***	(-13.37)***
INC2 (10-100 tkr)	-0.08950	-0.08310
	(-2.83)***	(-2.55)**
INC3 (100t - 1 mill kr)	0.06140	0.05440
	(1.65)*	(1.42)
INC4 (> 1 mill kr)	1.56810	1.55590
	(11.29)***	(10.57)***
Self employed	-0.19310	-0.20200
	(-3.62)***	(-3.67)***
Worker	-0.14160	-0.16710
	(-4.33)***	(-4.97)***
Unemployed	-0.20990	-0.22520
	(-6.03)***	(-6.22)***
Education	-0.09108	-0.09759
	(-63.68)***	(-58.88)***
Tenure	-0.02564	-0.02562
	(-7.31)***	(-6.66)***

Note: *** significant at 1 % level, ** 5 % level, and *10 % level