HOW IMPORTANT IS STATED HOMELAND EDUCATION FOR REFUGEES' ECONOMIC POSITION ?

joop hartog
aslan zorlu

AIAS
University of Amsterdam

Abstract

We use data on refugees admitted to the Netherlands that include registration of education in their homeland, as registered by immigration officers. Such data are seldom available. We investigate the quality and reliability of the observations and then use them to assess effects on economic position data. The most remarkable finding is the absence of returns to higher education.

Very preliminary

First draft, 080405
Keynote presentation LoWER Annual Conference, Mannheim, April 16 2005

File: refugees and education.310305

1. Introduction

Education is commonly held to be a key variable for economic success of immigrants in their destination country. Several studies have indicated that it matters very much whether this education has been acquired in the origin country prior to migration, or in the destination country. Existing studies that make this distinction never have direct observations on the decomposition: it is always inferred, usually from highest level of education attained and age at immigration. An exception is Kee (1993) who uses direct observations for immigrants to the Netherlands. In this paper we use registration by immigration officers of education attained when immigrants apply for admission. We investigate the quality of the data and then use the observations to assess the importance for economic success during the first five to six years after admission.

2. Data selection

We use our CRV/GBA/RIO dataset, and the subsample of those individuals for whom ITS has recorded education. The CRV/GBA/RIO dataset has been created by linking registration of immigrants by the immigration service IND with observations on socio-economic variables by the national statistical bureau in the RIO files, where the linking has taken place through the registration of population GBA. The dataset is described in the Appendix. Researchers at ITS checked the files of all asylum applicants from 1995 to 2000 and coded information on homeland education. They investigated some 200 000 files (all asylum applicants entering between 1995 and 2000). After linking with our CRV/GBA/RIO data, this left some 35 000 observed individuals: a reduction to 1/3 as RIO is a 1:3 sample and reductions due to requiring presence in 2000 and a decision on the status application.

As noted, in this paper we analyse data for applicants who are still present in the Netherlands in 2000. Following an entry cohort and hence, using information on returns as well, is not an attractive alternative, as it would only be feasible for cohorts entering in 1998 or later (when the electronic IND registration started): it would restrict the analysis to fairly recently arrived immigrants only. An alternative would be to use all observations in the database up to their last moment of observation, i.e endpoint 2000, or year of departure if earlier. Our choice implies that we do not observe individuals that have left before 2000. This is potentially disturbing, if it generates selectivity on variables or processes we are interested in. We are fairly confident that this is not the case, however. Our sample is restricted to those who have a valid permit to stay. We can observe departures for arrivals in 1998 or later. Among those with a permanent residential status in that sample, we only observe 5 people who have left (out of perhaps some 10 000 admissions). Those with a temporary permission to stay may be expelled when their homeland is declared safe (eg former Yugoslavia). In that case, return migration is an exogenous event and need not worry us. Divergence between the two selection rules will also occur because of sample attrition. The initial recording covers all members of a household; if someone later leaves the household this means leaving the sample. We will use the observations with selection up to last year of observation for robustness tests on our findings.

Our sample contains 13 436 observations. By year of arrival, the sample spans the decade of the 1990's, but most observations date from 1996 or later: 3.6 % arrived in 1990-1995, the remainder in 1996-2000.

To create a reasonably homogenous sample, we require individuals to have a valid permit to stay. This excludes individuals whose application is still being processed, and who may later be denied access[1]. This would disturb our observations. The records contain many statements on the applicant's formal status, but there is no track record of progress in the decision making. Timing of granting some status is not registered. So, we decided to stay on the safe side and distinguish only three categories: *A status* (permanent permission to stay; includes also immigrants granted Dutch citizenship in 2001), *AMA* (entered as independent minor, i.e. not older than 18), and *preliminary* status (all other; some of them may alter be upgraded to A status). Presumably, AMA refers to status upon entry, A status and preliminary status refer to an unknown date: time of GBA registration or IND update. Table 1 gives the distribution by status and country of origin.

Table 1. Admissions by title of residence and country of origin, Percents

	A-status	AMA	Preliminary	Total (N=100)
Iran	42.14	0.74	57.12	674
Irak	35.51	0.32	64.17	3,123
Somalia	15.9	15.33	68.76	874
China	1.17	53.38	45.45	429
Afghanistan	37.19	1.05	61.76	2,947
Sudan	28.45	4.38	67.17	594
Former Yugoslavia	27.52	0.16	72.33	1,272
Soviet Union	28.08	2.91	69.01	755
Other countries	16.11	13.51	70.38	2,768
Total	28.36	6.2	65.44	13,436

In our sample of refugees, Irak, Afghanistan and Other countries each contribute about one fifth, 11% is from former Yugoslavia; Iran, Somalia, Sudan and the Soviet Union each contribute some 5-6 and 3%is from China. About two thirds of the refugees have a preliminary status, just over a quarter has A status, 6% is AMA. AMA's are mostly from China and Somalia. Among refugees with A status Iran, Irak and Afghanistan are over-represented.

2. On measuring education

We are specifically interested in the relevance of homeland education for socio-economic position after immigration, but we have reason to be suspicious about the quality of registration of education. The original documents may register the applicant's education, but if so, registration is not according to a standardized classification system. ITS analysts read the original entries and coded them to

[1] The sample also contains individuals who are still in the application procedure and are registered at GBA. Their number is unknown but very small.

standard classification. From the analysis by ITS we know that education is missing in many cases. We also know that education is not an important variable in the decision process and that immigration officers have no special interest in the variable. In fact, they consider it irrelevant and often ignore it. Hence, prior to attempting any analysis we should assess the quality of measurement.

Table 2 Education level by country of origin, percentages

	None	1-3 basic	4-5 basic	Basic	Ext. basic	Sec. gen.	Sec. voc	Some tert.	Tert	Missi ng	TOTA L
Iran	2.82	0.89	1.34	2.23	10.09	23.74	2.23	3.12	10.98	42.58	674
Irak	4.67	1.44	2.34	4.13	9.54	8.36	3.3	3.39	12.33	50.5	3,123
Somalia	20.82	8.92	5.38	7.32	11.9	10.64	0.8	0.57	2.52	31.12	874
China	4.2	15.38	20.05	14.22	13.99	4.66	0.93	0	0.23	26.34	429
Afghanistan	5.09	1.26	2.21	4.21	4.48	9.54	0.71	3.02	9.74	59.76	2,947
Sudan	5.72	1.52	2.86	5.05	8.92	11.11	2.02	5.56	18.52	38.72	594
Former Yugoslavia	6.53	0.94	2.99	4.64	19.03	12.03	10.46	2.04	2.67	38.68	1,272
SovietUnion	4.9	2.25	1.85	4.77	16.16	16.16	5.56	1.85	13.51	32.98	755
Other countries	9.68	4.34	6.11	10.91	12.17	11.92	1.88	2.13	4.34	36.52	2,768
Total	6.97	2.9	3.86	6.1	10.54	11.06	2.9	2.63	8.45	44.6	
Total	937	390	518	820	1,416	1,486	389	353	1,135	5,992	13,436

Table 2 presents the distribution by education levels, distinguished by country of origin. The first thing to note is that in 45 % of the cases education is missing. 7 % has no education at all, 23 % has basic (including extended basic), 14 % has secondary and 11 % has tertiary education. Refugees from China, of whom many are AMA, have remarkably low levels of education and so have refugees from (former) Yugoslavia. Among refugees from Iran there is a remarkably high share with secondary education, Sudan has relatively many highly educated refugees, the distribution from the Soviet Union is rather bimodal, high share with extended lower and with high education. Refugees from Irak tend to the high end of the distribution, refugees from Somalia tend towards the low end. Table 3 gives a more compact overview, with education aggregated into three levels only. By title of residence (not shown here), refugees with A status have higher average education, AMA's have lower average level of education. Among all refugees, 28% has A status, while among refugees with tertiary education, 43 % has A status.

Table 3. Education by country of origin, three categories

	Primary	Secondary	Tertiary	Missing	Total (N=100)
Iran	7.27	39.17	10.98	42.58	674
Iraq	12.58	24.59	12.33	50.50	3,123
Somalia	42.45	23.91	2.52	31.12	874
China	53.85	19.58	0.23	26.34	429
Afghanistan	12.76	17.75	9.74	59.76	2,947
Sudan	15.15	27.61	18.52	38.72	594
Yugoslavia	15.09	43.55	2.67	38.68	1,272
Soviet Union	13.77	39.74	13.51	32.98	755
Other countries	31.03	28.11	4.34	36.52	2,768
Total (%)	19.83	27.12	8.45	44.6	
Total (N)	2,665	3,644	1,135	5,992	13,436

One might perhaps be tempted to calculate the distribution of education levels over the number of individuals excluding the missing observations. But his would assume that the share of missing observations is equal at all levels of observations and this is an assumption of which we cannot establish the reliability. Below, we will take a closer look at the missing observations.

We also have observations on education recorded by CWI, the public employment service that assists individuals in finding a job. Registration as job seeker is a requirement for obtaining social benefit. Clearly, this registration is highly selective. But we might assume that labour service agents are more dedicated in registering education, as it is an important instrument for the service they have to provide: they have an interest in accurate assessment.

Table 4. Education ITS and education CWI

	Education CWI							
	Unknown	BO	Ibo/MAVO	MBO/HAVO/	HBO	WO	Missing	Total
Education ITS								
No education	9.09	9.99	5.69	1.33	1	0.66	8.68	6.97
1-3year Primary	3.74	4.17	2.97	0.96	0.14	0.26	3.41	2.9
4-5year Primary	5.88	5.78	4.17	1.33	0.14	0.4	4.35	3.86
Primary	4.81	8.47	7.08	3.91	0.29	0.26	6.8	6.1
Extended primary	13.9	11.68	14.74	9.14	1.85	1.32	11.3	10.54
Secondary, general	6.95	9.49	13.35	20.58	9.56	4.5	10.13	11.06
Secondary, vocational	0.53	2.57	3.67	5.68	3.71	1.06	2.44	2.9
Some Tertiary	2.67	1.48	1.64	5.01	5.71	5.17	2.16	2.63
Tertiary	8.56	3.2	3.23	8.85	28.96	40.26	5.63	8.45
Missing	43.85	43.17	43.45	43.22	48.64	46.09	45.1	44.6
Total (N=100)	187	2,372	1,581	1,356	701	755	6,484	13,436

From Table 4 we may note first of all that the missing observations do not match: they are not concentrated as single diagonal entry in the cell (missing ITS, missing CWI). Missing observations are due to different processes in the two agencies and are not a unique property of the respondent. The overall proportions are about equal, at 45% for ITS and 49 for CWI. This must be coincidence, as ITS missings must be due to non-registration by the immigration officer while CWI missings must be due to absence of contact with the employment service. Interestingly, the proportion of missing observations on ITS education is virtually the same for every level of CWI education. If we are justified in assuming that CWI registration is reasonably reliable, this would imply that missing observations in ITS are unrelated to level of education, and hence, that the distribution of education is representative for all refugees: we can relate the frequencies to only those individuals for whom education has been registered. If we group the levels of education in primary, secondary, tertiary (to allow for matching classifications), we can calculate that in 6.6 % of the cases, the ITS level is higher than the CWI level, while in 5 % the reverse holds. This points to some upward bias in the ITS registration relative to the CWI registration, as one might have anticipated: ITS is the individual's assessment without any check, CWI coding is based on the registration by an employment agency that has interest in accurate assessment; they translate foreign education into the guessed Dutch equivalent and might perhaps be inclined to some downward bias because of unfamiliarity with foreign schooling systems. But the bias is quite modest, which lends credibility to the ITS data. However, there is certainly no agreement between IND and CWI on individuals' level of education. Table 5, with education registered in three levels, shows this quite clearly. If we consider only cases for which both institutions record an education level (ie, exclude missing observations), the diagonal elements in Table 5 would be 0.50, 0.65 and 0.65, meaning that for given classification by CWI in no more that two thirds of the cases would ITS record the same level.

Table 5. Education according to IND and according to CWI

	CWI				
IND/ITS	Primary	Secondary	Tertiary	Missing	Total
Primary	28.06	14.20	1.58	23.24	19.83
Secondary	25.13	36.64	16.28	26.03	27.12
Tertiary	3.60	5.82	34.82	5.63	8.45
Missing	43.22	43.34	47.32	45.10	44.60
Total	100	100	100	100	100

We have analysed possible patterns of non-recording of education by IND officers by running a logistic regression. Registration of education is systematically related to some variables: education is more often registered for immigrants who are older at arrival and for men, it is better known for later arrivals, and there are significant differences between countries of origin: better known for China, Soviet Union and Somalia, less often known for Afghanistan and Yugoslavia. Statuswait and undocyears also have significant, negative, effect on the probability of registration. The variables refer to a difference in year of GBA and CRV registration, with Undocyears measuring number of years that IND registration anticipates GBA registration (set at zero if negative) and Statuswait the similar reverse measure.

Table 6. Logit on education level registered (odds ratios)

	Model_I		Model_II	
Age	1.011	***	1.002	
Woman	0.865	***	1.040	
arrival96	0.693	***	0.753	
arrival97	0.316	***	0.683	**
arrival98	0.497	***	1.200	
arrival99	1.177	**	1.060	
arrival00	3.615	***	2.623	**
Undocyears	0.846	***	0.937	
Statuswait	0.250	***	0.618	
Iran (reference)	1.000			
Irak	0.860		1.255	
Somalia	1.609	***	2.169	***

China	2.284 ***	1.786 *
Afghan	0.540 ***	1.020
Sudan	1.143	1.358
Yugosl	0.798 **	1.532 *
SovUni	1.714 ***	1.606 *
Other	1.235 **	1.490 *
OccupNone		0.456 ***
OccupUnkn		0.002 ***
OccupMiss		1.182
A-status		0.839
AMA		1.658 ***
Emigrated		0.960
Naturalised		1.259
Married		1.191 **
_cons	2.036 ***	13.324 ***
Chi2	1279.514	5256.350
Aic	17445	5715
Ll	-8704.511	-2831.545
N	13720	13720

significance levels: *p<.1; ** p<.05; *** p<.01
DEFS:
Statuswait =(Arrival year IND-Arrival year GBA)
Undocyears =(Arrival year GBA-Arrival year IND)
Emigrated= emigrated or administrative removal in 2001
Naturalised= Naturalised in 2001

We have considered using information on homeland occupation (also coded by ITS) as a variable to assess the reliability of registered education. But a cross-tabulation of education and occupation shows wide dispersion of education by occupation. Moreover, many educations are so low, and so little specific that it would be hard to use the additional information to test the reliability of education. There are auto mechanics and farm hands with tertiary education and pharmacists with just extended basic education. The matrix is simply too far removed from diagonality to yield useful additional information. Interestingly, our second logistic regression in Table 5 shows no relation between recording education and recording occupation. For those with no occupation or occupation unknown, education is also less frequently known. The occupational variables, and the other variables in the second regression take some explanatory power away from the other variables.

We have also made inquiries with IND and with immigration officers who do the intake and registration of immigrants. They could not give any explanation on the pattern of registration of education and they are absolutely unaware of any systematic effects.

We conclude that non-recording of level of education by IND (and hence, ITS) has different incidence by country of origin, years of arrival and gender of the refugees, but there is no indication of a systematic rule applied by immigration officers. From comparing ITS and CWI registration, we conclude that there is no indication of systematic upward or downward bias in the level of education recorded by ITS. But the substantial variation in the cross-classification of the two registrations indicates that measurement error in the level of education is far from negligible. Thus we should anticipate that in regression analysis estimated coefficients will be biased downwards.

We can use several approaches to get a feel for the possible magnitude of measurement errors in our education data. Let's assume that measurement errors are independent of the level of education. Suppose, for a given level of ITS education, all variation in CWI education reflects measurement error. Then, if we translate CWI education categories in years of education, the variance in years of education for the given ITS education would be the variance due to measurement error. We could do this for every row of ITS education. The weighted average of these variances across ITS educations would give an indication of overall variance in education due to measurement error. The variance in ITS education (in years) consists of the sum of true variance plus variance due to measurement error (because we assumed zero covariance). Simple subtraction of measurement variance for total variance of ITS education would give us true variance, and hence, the relation of noise to signal, ie measurement variance to true variance. Proceeding in this way, we find variance due to measurement error of 9.83 (weighted from CWI variance by ITS education, ignoring ITS missing) and ITS variance of 12.73. This would imply that ¾ of ITS education variance in years is due to measurement errors, which is not very credible.

In the calculation above, we proceed on the assumption that for ITS education, the mean of CWI education is the true value. Of course, we can reverse that, and assume that for given CWI education, the true value is the mean of ITS educations, and the dispersion is fully due to measurement error. Thus, measurement error is the weighted average of the variances within columns, rather than within rows as assumed above. We have performed both calculations (see Appendix C).

A more formal approach is the following. We have two measures of an individuals' education, S_{iT} as measured by ITS and S_{iC} as measured by CWI, both measured in years.

Assume:

(1) $S_{iT} = S_i + e_{iT}$

(2) $S_{iC} = S_i + e_{iC}$

Both measurements report individuals i's true education, but with different measurement errors. Assuming that the errors are independent of the true value, we can write for the variances:

(3) $V_T = V + V_{eT}$

(4) $V_c = V + V_{eC}$

where V measures the variance of true education across individuals, and $V_{e.}$ measures the variances in the error terms.

We can also calculate the covariance between the two measures, V_{CT} and write this as

$$(5) \quad V_{CT} = E\{(S_{iT} - E(S_{iT}))(S_{iC} - E(S_{iC}))\}$$

$$= E\{(S_{iT} - E(S_i))(S_{iC} - E(S_i))\}$$

under the assumption of zero expected measurement error and independence of true education levels. Substituting the definitions (1) and (2), we get

$$V_{CT} = E\{(S_i + e_{iT})- E(S_i))(S_i + e_{iC})- E(S_i))\}$$

$$= E\{(S_i)- E(S_i) + e_{iT})(S_i)- E(S_i) + e_{iC})\}$$

$$= V + V(e_{iT}\, e_{iC})$$

or

$$(6) \quad V_{CT} = V + \rho\sqrt{}\, V_{eT}\, V_{eC}$$

Hence, from the three equations (3), (4) and (6) we can identify the three variances if we know (or make assumptions on) the correlation between the two measurement errors. If we assume correlation zero, the identification is straightforward: the covariance identifies the true variance, and the two measurement variances follow directly from subtracting this from the two observed variances. If, at the other extreme, the correlation between measurement errors were unity, we have to solve the three unknowns from the three equations. We can use computer routines to do so.

Applying this method, we find V_T = 254.30, V_C = 401.84 and V_{CT} = 591.36, which actually rules out zero correlation between measurement errors.

3. Socio-economic status

We have estimated the effect of education, and other variables, on socio-economic status, distinguishing three options: working, not working and having some social benefit (unemployment, welfare), or not working and having no benefit (non-participating). We have estimated two separate logit models, on work versus non-work and benefit versus non-benefit. Unemployment benefit and welfare are not distinguished, as the number of observations would become too small. We have not attempted to estimate a model with three options simultaneously, as interpretation would be less straightforward. The model we estimate is a logit panel model with individual random effects, estimated on all individuals who are present in 2000.[2] We

[2] The models are estimated with STATA, xtlogit, RE

do not include refugees who have returned before 2000, or apply correction for such attrition. As noted above, among those admitted permanently (with an A-status) no one leaves and among those admitted temporarily, departures are exogenous, dictated by the political situation in their homeland.

Admitted immigrants have identical entitlements to social security benefits as native Dutch. But unemployment benefits are conditional on work history, which will disqualify immigrants in the early years after arrival. Social assistance does not depend on length of stay in the Netherlands, but is means-tested at the household level, and may disqualify marital partners or children (although the level of the benefit will depend on household composition). Refugees are provided shelter, food and a small amount of cash while their application is in process.

In Table 7, we report estimation results for the probability to work in the most extended specification. If we leave out duration slope interaction with source country, or status in 2001 (see below), this only affects the dummy coefficients on year of entry[3]. The results show a negative effect of age; women are less likely to work. We estimated a cubic function for years since migration to obtain maximum flexibility, although this cannot be extrapolated as we only estimate over a five years interval[4]. Years since migration, i.e time elapsed since registration in the population register GBA, has a monotonic positive effect on the probability to work, as one might expect. There are marked differences between source countries (using Iran as the reference source), both in intercept and slope. For easier interpretation we have plotted the profiles, both on average (Figure 1) and relative to the standard years-since-migration profile (Figure 2).

Figure 1

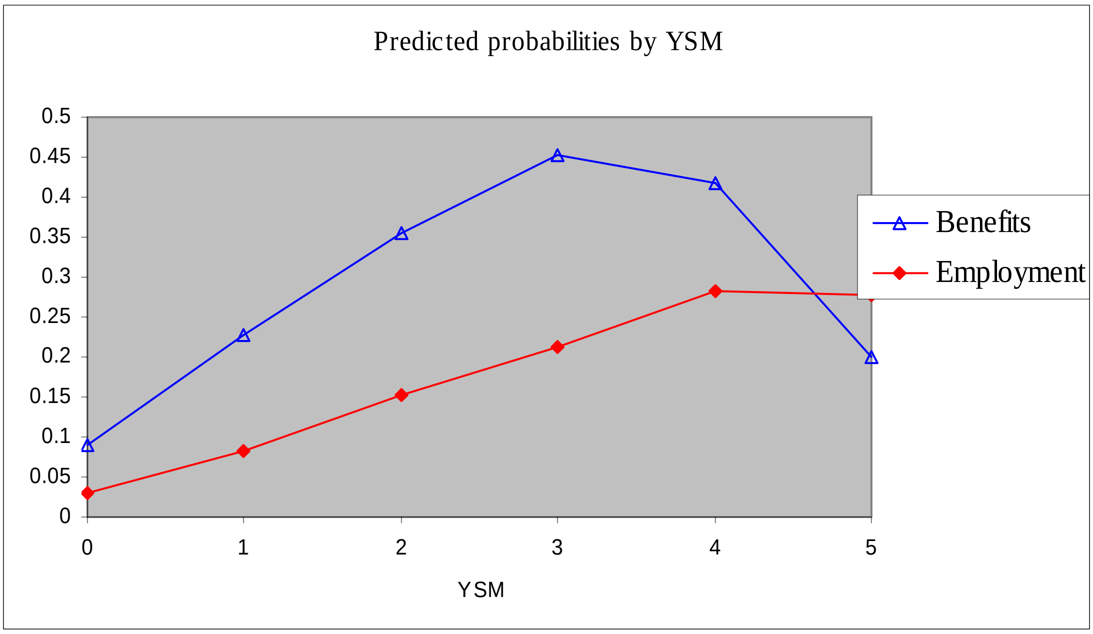


Figure 2

[3] Note that we can include year of entry dummies in addition to the duration variables because we use panel observations.
[4] Adding a quadratic for age was immaterial for our results (it was mostly insignificant).

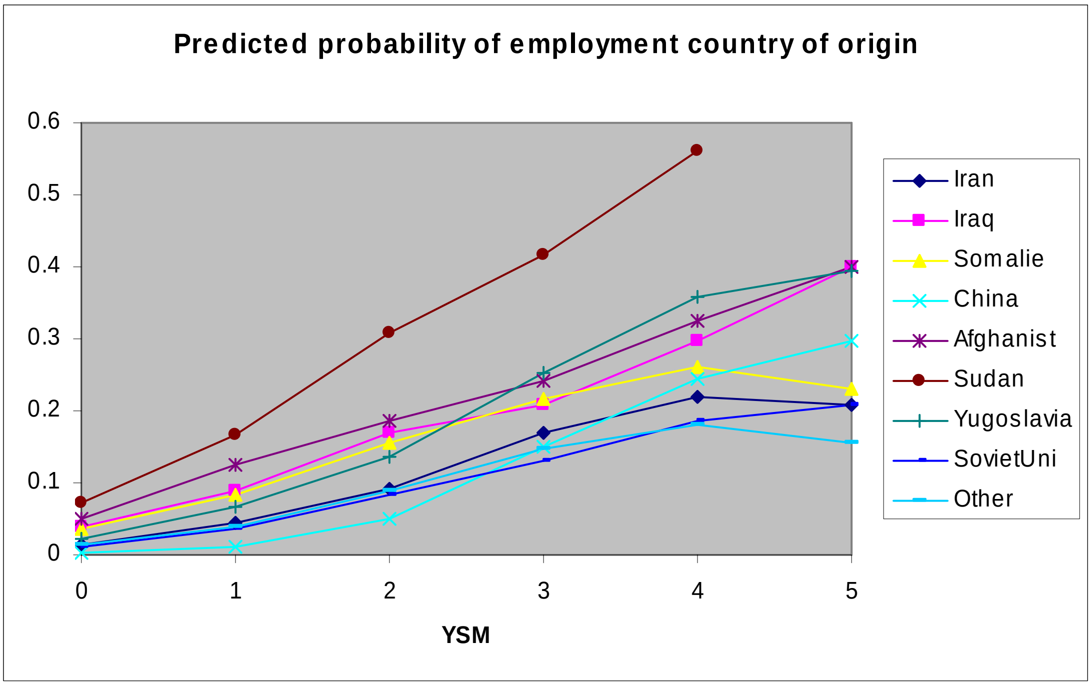
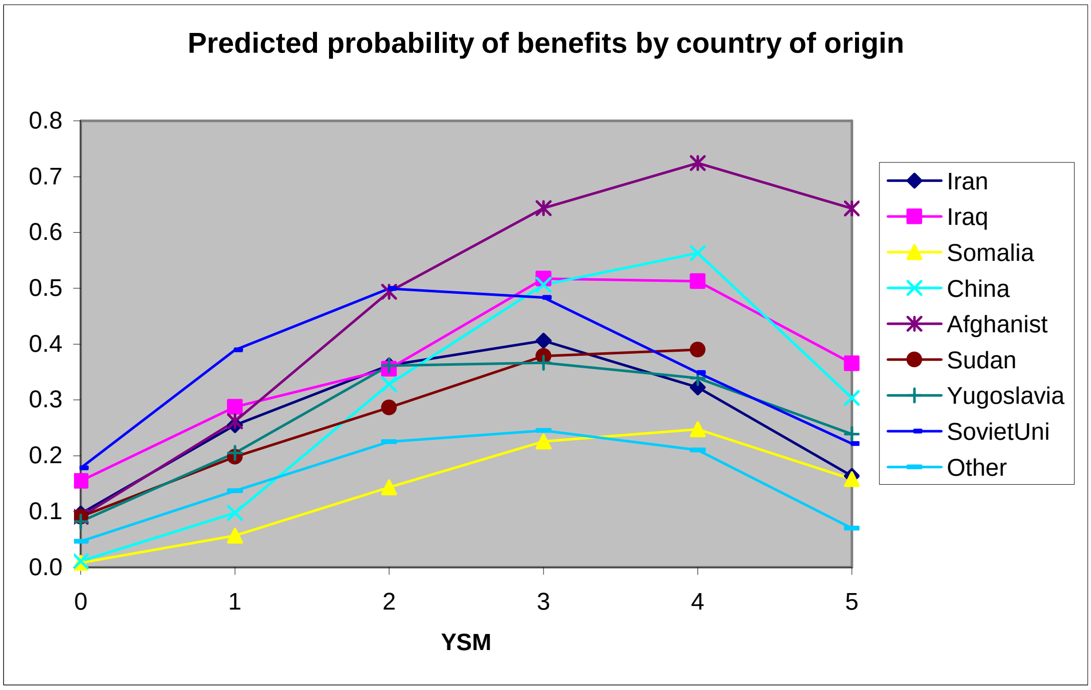


Figure 3



Duleep and Regets (1999) have formulated the hypothesis of a negative relationship between intercept and slope of the profile of integration in the host country. Immigrants with a greater gap upon entry have greater incentive to invest in host country human capital. Their poorer socio-economic position implies lower opportunity cost of investment, hence they will invest more and grow faster. We have plotted intercepts and slopes across countries in Figure 4 and find clear confirmation of this hypothesis

Figure 4

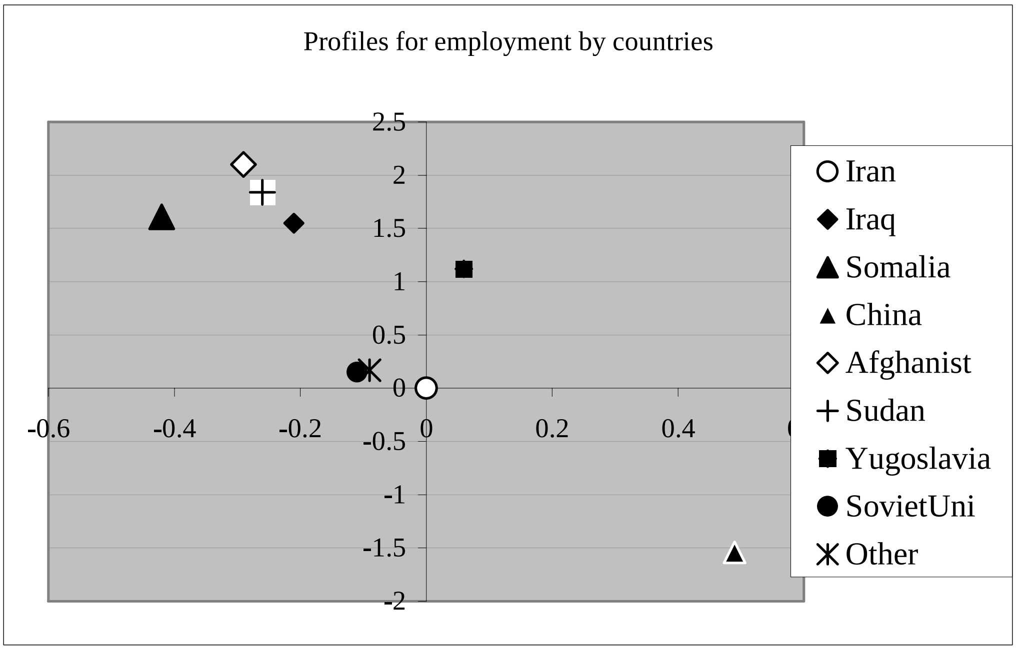
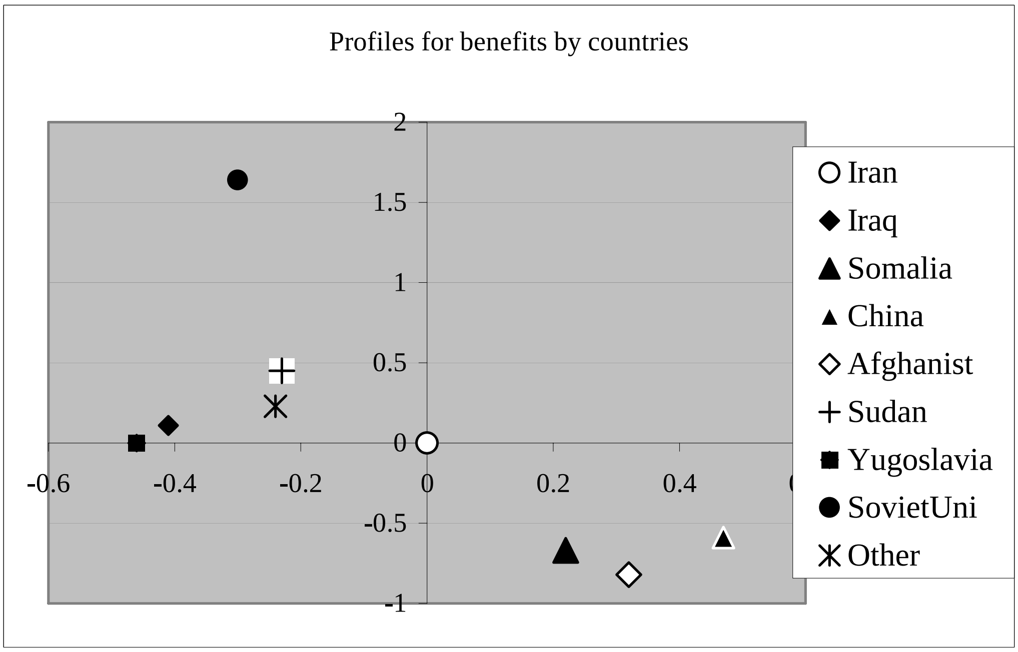


Figure 5



Our special interest is in the effect of home country schooling. Generally, there is a positive effect of schooling on the probability to work. Considering magnitudes and significance levels, it seems one may distinguish three steps: less than basic education, basic up to secondary general, secondary vocational and higher. The probability to work increases markedly between steps and is quite similar within the steps. Interestingly, and perhaps not surprisingly, the effect of secondary level schooling is split: secondary general has no marginal effect on the probability to work, secondary vocational has a marked positive effect. Within the highest step, the effect of education diminishes, consistently but not significantly; this may be because individuals are engaged in obtaining additional schooling in the Netherlands.

The effect of A status and of AMA status is negative, which is not surprising for the latter, but it is for the former. Two other duration variables have a very interesting effect. Undoc years have a positive effect: refugees who have been in the Netherlands as undocumented workers before reporting to IND have higher probability to work. This is as anticipated: as undocumented workers they will mostly have worked, and effectively this adds experience to their " years-since-migration". Statuswait has a negative effect. Spending more time in the application procedure reduces the probability to work, even after controlling for the other duration variables.

Finally we considered the effect of the situation in the year after our observation interval. Those who will have been naturalized in 2001 are more likely to work, those who will have returned (or administratively removed) are less likely to work.

Table 7. Logit estimates for employment and benefit status

	EMPLOYMENT		BENEFITS	
	Coeff.	Odds ratio	Coeff.	Odds ratio
Age	-0.09	0.915	0.11	1.115
	0.004	0.004	0.004	0.005
Woman	-2.01	0.134	0.73	2.068
	0.083	0.011	0.081	0.168
YSM	1.69	5.420	2.19	8.908
	0.123	0.667	0.118	1.055
YSM2	-0.18	0.832	-0.55	0.575
	0.034	0.028	0.035	0.020
YSM3	0.01	1.013	0.04	1.044
	0.004	0.004	0.004	0.004
ysmIraq	-0.21	0.812	-0.41	0.662
	0.097	0.078	0.09	0.060
ysmSomali	-0.42	0.656	0.22	1.245
	0.111	0.073	0.115	0.143
ysmChin	0.49	1.639	0.47	1.602
	0.176	0.288	0.143	0.228
ysmAfgh	-0.29	0.750	0.32	1.383
	0.099	0.074	0.094	0.130
ysmSudan	-0.26	0.768	-0.23	0.797
	0.13	0.099	0.133	0.106
ysmJugos	0.06	1.058	-0.46	0.632
	0.102	0.107	0.096	0.060
ysmSovU	-0.11	0.896	-0.3	0.740
	0.136	0.122	0.121	0.090
ysmOther	-0.09	0.915	-0.24	0.787
	0.096	0.088	0.094	0.074
arrival96	0.22	1.242	0.73	2.080
	0.116	0.144	0.128	0.267
arrival97	0.78	2.184	0.34	1.404
	0.137	0.299	0.149	0.209
arrival98	0.77	2.166	-1.03	0.356
	0.146	0.315	0.166	0.059
arrival99	-0.24	0.785	-1.52	0.219
	0.202	0.158	0.228	0.050

arrival00	0.09	1.091	-1.44	0.237
	0.423	0.462	0.477	0.113
edu1_3y	-0.03	0.968	0.41	1.502
	0.266	0.257	0.299	0.450
edu4_5y	0.16	1.171	0.28	1.323
	0.232	0.272	0.259	0.343
eduPrim	0.45	1.576	0.13	1.142
	0.202	0.319	0.227	0.259
eduPrim_ext	0.59	1.800	0.56	1.744
	0.179	0.323	0.201	0.351
eduSec_gen	0.55	1.731	0.57	1.760
	0.178	0.308	0.195	0.344
eduSec_voc	1.06	2.888	0.28	1.329
	0.237	0.684	0.261	0.348
eduHigh_some	1.03	2.787	0.28	1.330
	0.249	0.694	0.277	0.369
eduHigh	0.87	2.397	0.39	1.471
	0.189	0.452	0.203	0.299
eduMiss	0.6	1.822	-0.08	0.923
	0.157	0.286	0.172	0.159
Iraq	1.55	4.689	0.11	1.114
	0.275	1.290	0.245	0.273
Somalia	1.61	4.980	-0.67	0.510
	0.312	1.552	0.32	0.163
China	-1.54	0.214	-0.59	0.557
	0.491	0.105	0.398	0.221
Afghan	2.1	8.179	-0.82	0.440
	0.275	2.252	0.249	0.110
Sudan	1.84	6.285	0.45	1.567
	0.321	2.017	0.33	0.517
Jugoslavia	1.12	3.074	0	1.000
	0.297	0.912	0.27	0.270
SovietUni	0.15	1.159	1.64	5.169
	0.349	0.405	0.287	1.483
OtherC	0.17	1.180	0.23	1.261
	0.277	0.327	0.25	0.315
A_Status	-0.67	0.511	3.93	50.767
	0.091	0.046	0.111	5.620
AMA	-0.91	0.401	3.53	34.121
	0.167	0.067	0.188	6.421
Undocyears	0.73	2.083	0.06	1.060
	0.052	0.107	0.056	0.059
Statuswait	-0.79	0.452	-0.28	0.752
	0.088	0.040	0.091	0.068
Naturalised	0.62	1.850	0.96	2.605
	0.106	0.196	0.117	0.305
Returned	-1.35	0.260	-1.93	0.145
	0.279	0.073	0.424	0.062
Married	-0.64	0.526	1.27	3.548
	0.082	0.043	0.09	0.318
Amsterdam	-0.11	0.897	1	2.726
	0.134	0.121	0.161	0.438
Rotterdam	-0.82	0.442	0.31	1.369
	0.157	0.069	0.162	0.222

DenHaag	-0.59	0.556	0.6	1.814
	0.15	0.083	0.157	0.284
Utrecht	0.17	1.188	0.3	1.356
	0.257	0.305	0.271	0.367
_cons	-3.79	0.023	-9.49	0.000
	0.353	0.008	0.375	0.000
lnsig2u				
_cons	1.67		1.97	
	0.052		0.049	
Statistics				
N	31323		31323	
Ll	-11364.7		-11347.9	
chi2	2188.96		2564.69	
Aic	22825.38		22791.75	

Standard errors below the coefficient and odds ratio

Table 7 also reports the results on the panel logit for the probability to receive social benefit. We use the same variables, we also report only the extended version. The sensitivity to the specification is similar to the results for work, with the addition that the coefficients on schooling also change somewhat if one allows duration slopes to differ between countries. Later arrivals are less likely to receive a benefit, in line with the need to build up rights over time, but after controlling for years since migration. Women and older refugees are more likely to receive benefits. Years since migration now has a parabolic effect, peaking at three years: refugees first build up entitlements, and then are more likely to find work, and end benefit reception status. Again, slopes and intercepts for the country specific effects of years since migration are negatively related (see plot in figure 5). The effect of homeland education is not as clear as on probability to work. The effect increases (from insignificance) for individuals with basic education or less, peaks for extended lower and secondary general and the drops off again. It is not quite clear what explains this pattern.

Refugees with A status and AMA status are strongly and significantly more likely to receive social benefits. More time spent waiting for a status decision increases the likelihood of benefit recipiency, mirroring the lower probability of working found earlier. Interestingly, refugees who have spent time as an undocumented worker are not more likely to receive a benefit, whereas they are more likely to work. Refugees who turn out to be naturalised in 2001 are more likely to receive benefits, refugees who return are less likely to receive benefits.

4.Earnings

4.1 Selecting the basic specification

In Table 8, we present estimates for earnings, for employees, ie individuals for whom labour earnings is the most important source of income during the year. It is annual labour (?)income divided by weeks worked and deflated by cost-of-living (base year 1995). Thus, there is some unknown measurement error as hours per week may vary over individuals, and because individuals may have some benefit income on top of labour earnings. We have estimated a panel GLS model with random effects.

Table 8:. Panel GLS random effect estimations

Weekly wages	Model I	Model II	Model III	Model IV	Model V	Model VI
Age	0.04***	0.06***	0.03***	0.04***	0.05***	0.04***
Woman	-0.46***	-0.43***	-0.30***			-0.43***
YSM	0.26***	0.37***	0.24**	0.26***	0.22**	0.21***
arrival96	0.16***			0.16***	0.09	0.13**
arrival97	0.18***			0.17***	0.13	0.11
arrival98	0.18***			0.16**	0.19	0.03
arrival99	-0.09			-0.11	-0.14	-0.20
arrival00	0.65**			-0.10	1.17**	0.63*
edu1_3y	0.00	0.62***	0.15	-0.07	0.27	0.00
edu4_5y	0.21*	0.45**	0.16	0.16	0.26	0.23**
EduPrim	0.28***	0.49***	0.11	0.31***	0.00	0.28***
eduPrim_ext	0.36***	0.74***	0.38*	0.32***	0.41**	0.39***
eduSec_gen	0.31***	0.51***	0.17	0.32***	0.10	0.34***
eduSec_voc	0.21*	0.45**	-0.33	0.18	0.31	0.21*
eduHigh_some	0.18	0.20	0.22	0.18	0.06	0.23*
EduHigh	0.20**	0.18	0.21	0.21**	-0.03	0.24***
EduMiss	0.13*	0.20	0.28	0.14	-0.01	
Iraq	0.35**	0.13	0.27	0.41**	0.08	0.08
Somalia	0.65***	0.61**	0.75**	0.66***	0.50	0.51**
China	-0.23	-0.56	0.53	-0.10	-0.79	-0.51
Afghan	0.29**	0.18	0.31	0.35**	0.03	0.14
Sudan	0.61***	1.40***	0.17	0.65***	0.75	0.53**
Jugoslavia	0.61***	0.89***	0.27	0.61***	0.55*	0.46**
SovietUni	0.38*	0.30	-0.19	0.26	0.44	0.35
OtherC	0.51***	0.88***	0.30	0.48***	0.63**	0.33
A_Status	0.02	-0.03	0.02	0.06	-0.10	0.00
YsmIraq	-0.05	-0.02	-0.05	-0.04	-0.04	0.06
YsmSomali	-0.08	-0.12	-0.13	-0.05	-0.15	-0.04
YsmChin	0.17*	0.20	-0.37	0.10	0.44**	0.27**
YsmAfgh	-0.03	-0.01	-0.09	-0.04	0.02	0.03
YsmSudan	-0.02	-0.28***	0.11	-0.01	-0.18	0.04
YsmJugos	-0.08*	-0.21***	-0.12	-0.07	-0.12	-0.02
YsmSovU	-0.03	0.02	0.15	0.03	-0.10	0.00
YsmOther	-0.05	-0.20**	-0.13	-0.03	-0.10	0.02
AMA	0.10	0.32**	-0.67**	0.12	-0.11	0.16
Undocyears	0.13***	0.15***	0.05	0.14***	0.08	0.12***
Statuswait	-0.14***	-0.16	-0.09	-0.19***	-0.04	-0.14**
Married	0.19***	0.19**	0.11	0.18***	0.19**	0.23***
Amsterdam	0.08	0.28**	-0.12	0.03	0.46***	0.02
Rotterdam	0.14*	0.39***	0.31	0.13	0.16	0.22**
DenHaag	0.16**	0.26	0.26	0.15*	0.29	0.08
Utrecht	0.03	0.03	-0.10	0.02	0.17	-0.02
_cons	2.36***	1.56***	3.03***	2.37***	2.10***	2.60***
N	5933	1424	1045	4762	1171	3088
Average N	1.6	1.9	1.2	1.6	1.5	1.6
Within	0.1012	0.23	0.0085	0.1024	0.1019	0.1215
Between	0.2436	0.401	0.125	0.1951	0.315	0.2541

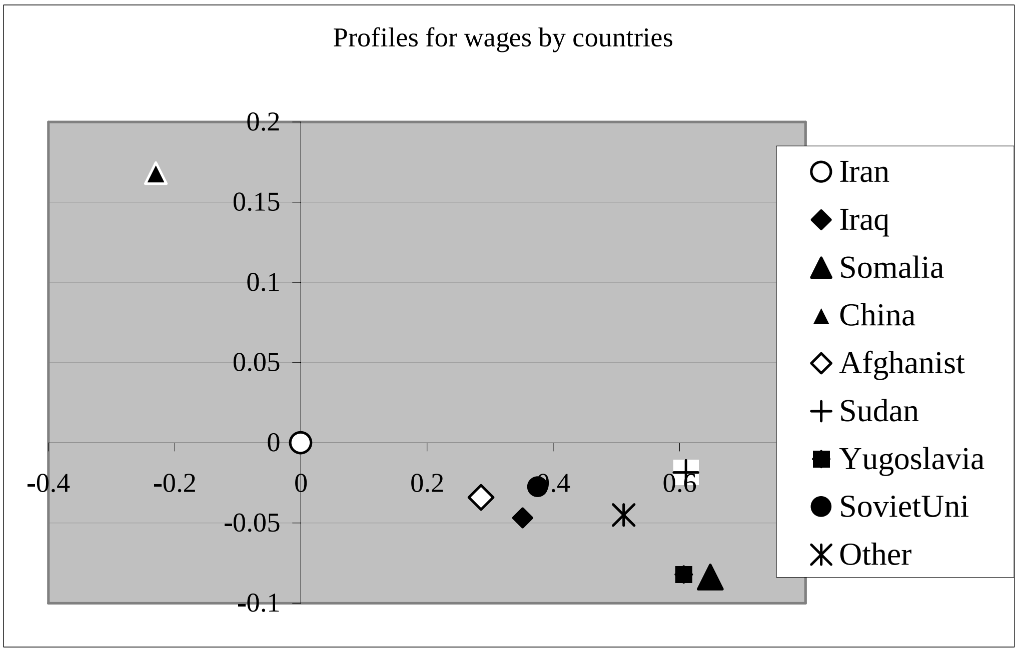
Overall	0.2503	0.402	0.1158	0.2031	0.3136	0.262
chi2	1502.8	697.3	125.97	951.61	399.6	828.26

Model I : basic specification
Model II: arrivals 1995 (IND registration)
Model III: arrivals 1998-2000 (IND registration)
Model IV: men only
Model V: women only
Model VI: observations with education missing deleted

The basic specification given in column (1) has been found after testing for several interaction effects and alternative specifications pointed out as we discuss the main findings below. Among the alternatives, we have separate estimates for men and women and separate estimates by year of arrival. The effect of age at arrival is fairly steep, with an annual growth rate of some 4%. The result is quite robust across specifications, but it drops if we estimate separately for later arrival cohort, suggesting that the disentanglement of age and years since migration is less than perfect. There is a strong and very substantial positive direct effect of years since migration. Higher orders of age and years since migration have also been tested, but they were not significant. The effect of arrival year is fairly uniform for the first three years. The strong positive effect for the latest cohort may be a selectivity effect: these are refugees who can work right in their first year of arrival, which is quite unusual. Eliminating the dummies for arrival years has no effect on the estimates for age or years since migration.

Education has an interesting parabolic effect. Most coefficients are statistically significant. The returns peak for extended primary education. One might think that this reflects selectivity, as those with higher educations might be engaged in further education in the Dutch school system. But the results of employment and benefit status in Section 3 (Table 7) do not lend much support to that interpretation. Interaction of education with years since migration is insignificant for all levels of education. One might have thought that those with the highest education may have the steepest time profiles, because of complementarity between homeland education and the intensity and returns of investment in specific Dutch human capital (Duleep and Regets, 1999). But we did not find any significant interaction between schooling and years since migration. In columns (2) and (3) we present results separately for early and late arrivals (the earliest and the latest that we can meaningfully define). (Arrival is measured by year of IND registration; individuals may have been in the country before that, so we still have variation on years since migration). The parabolic pattern of returns by education level is basically visible for the oldest and the youngest cohort, but precision is quite weak for the youngest. The oldest cohort has higher benefits from education than the youngest. It is quite remarkable that even for the oldest cohort, earnings drop for education levels beyond extended primary. Also remarkable is the strong benefits for the least educated, some years primary, after 5 years in the Netherlands. Thus, benefits from education clearly increase with time spent in The Netherlands, but the pattern by level of education is surprising. Perhaps, measurement error is not independent of education levels; this needs further investigation. We have tested for interaction between education and country of origin, but these terms were never significant: the benefits of a given education level do not differ between source countries.

Differences between source countries are marked. From top to bottom the ranking is Somalia, Sudan, Yugoslavia, Other countries, Soviet Union, Irak, Afghanistan, Iran (reference), China. Interactions between source countries and years since migration are not significant at conventional levels, but we left them in, to check on a negative relationship between intercepts and slopes across countries, as we did above. We have again plotted intercepts and slopes across countries, as above. We find a clear negative relationship, even if we were to ignore the extreme observation for Chinese immigrants.



The differences in status are not significant, except for AMA's if we split: in the youngest cohort, they are far behind, in the oldest cohort they have a premium of over 30%. This is a fantastic race through the earnings distribution. The effects of time elapsed before status obtainment are quite interesting. Years spent as undocumented worker add experience, but the pay-off is much less than the return to years since migration. Conversely, years spent waiting for a status reduces earnings, at about the same rate. These are substantial rates: a year of undocumented work adds 13% to earnings on top of the benefits from years since migration (the time scale has the same origin), another year of waiting reduces earnings by 14%. The effects are located with men, as they are not significant for women. We have also tested for selection effects, by adding a dummy for immigrants who had returned by 2001. The coefficient is not significant, further supporting our claim that in this sample, selective return migration is not an issue.

Married immigrants earn more than singles, and remarkably, on average earnings are highest in The Hague, the seat of government. But if we split between men and women, we see that men indeed earn most in The Hague, but that women earn most in Amsterdam.

As the overall regression indicates, women earn about half of what comparable men earn, which is a striking difference. We already pointed to some differences between men and women in the separate estimations, as reported in columns (4) and (5).

Several effects are essentially the same for men and women: age, years since migration, marital status. The rankings by country are very similar, suggesting that country effects relate to real underlying differences in human capital that immigrants bring. The gap between top and bottom of country effects is wider for women than for men though (1.5 versus .8). Just as for men, the coefficient on years since migration do not differ significantly between countries. In fact, significance levels are even lower, and we can only conclude that in those early years after arrival the speed of assimilation for women is identical across source countries. The only exception is Chinese women, with a strong positive effect. The parabolic effect of education that we found in the joint estimation is also visible in the results for men and women separately, but with some differences. For men, returns to education behave like a step function: zero if basic education has not been completed, some 35% for primary and extended primary, some 20% for the higher levels. For women, a single peak stands out, a significant 41% at extended primary education.

4.2 Robustness checks

The core result on education can be summarised as follows. The effect of education on earnings is parabolic, not monotonic. Highest earnings are consistently found for immigrants with extended primary education. Most remarkable is the consistent drop in earnings for immigrants with education beyond secondary. How robust is this result?

In column (5) of Table 8 we have reported estimation results for the case where we drop all observations where information on education is missing. This has no effect: whether we know education or not is immaterial for the estimation of the coefficients on the other variables. Covariances between education and other variables are not responsible for the result.

There are two mean reasons for concern about the reliability of this unusual result: measurement errors in education and selective labour market participation. Above, we have already indicated that with two measures of education, we may deduce some information on the contribution of measurement error to the variance. In OLS, we can use this information to correct the estimated coefficient for the bias due to measurement error (Johnston,1972; p 282; Wooldridge, 2002, p 73-76). In logit models, such corrections have not been developed.[5]

We have also made estimates with a selection on observations for reliability of the education variable: with clear mismatch in the two classification systems we discarded the observation from the sample (Table 9). The first selection rule we applied, reported under model II, is the following:

ITS primary or less : accepted if CWI classification Basic
ITS extended primary and secondary : accepted if CWI ibo/mavo or mbo/havo
ITS secondary vocational, some tertiary: accepted if CWI mbo/havo en hbo
ITS higher : accepted if CWI university

[5] The correction is not obvious and certainly not simple. High variance of an explanatory variable reduces *all* coefficients in a logit model. Private communication, J.S. Cramer

Due to the difference in the two classification systems this is not a strict criterion for a perfect match, so some noise is inevitably left. We therefore also used as an alternative selection rule that the classifications should agree on the level of primary/secondary/tertiary. We then estimated two specifications: the usual specification with all ITS categories (Model III) and a specification with three levels only (primary/secondary/tertiary; Model IV). By requiring a credible match between ITS and CWI classification, we reduce the sample to those observations for which CWI classification is available. This is quite restrictive and certainly not random. Therefore we also re-estimated the non-restricted versions on the sub-sample for which both ITS and CWI education levels are available. This is reported as Model I. For ease of comparison we also copied the basic specification from Table 8.

Table 9. Selecting on reliable measurement of education

	C1 from T8		Model I		Model II		Model III		Model IV	
Age	0.04	***	0.02	***	0.02	***	0.02	***	0.02	***
Woman	-0.46	***	-0.38	***	-0.32	***	-0.37	***	-0.38	***
YSM	0.26	***	0.28	***	0.34	***	0.26	**	0.25	**
arrival96	0.16	***	0.14	**	0.14	**	0.15	*	0.15	*
arrival97	0.18	***	0.21	***	0.18	**	0.23	**	0.21	**
arrival98	0.18	***	0.05		0.09		0.07		0.05	
arrival99	-0.09		-0.33	*	-0.46	**	-0.25		-0.23	
arrival00	0.65	**	0.87		0.92		0.78		0.76	
edu1_3y	0.00		0.07		0.03		0.16			
edu4_5y	0.21	*	0.23	*	0.17		0.10			
eduPrim	0.28	***	0.27	**	0.29	**	0.40	**		
eduPrim_ext	0.36	***	0.32	***	0.31	**	0.25	*		
eduSec_gen	0.31	***	0.27	***	0.27	**	0.20			
eduSec_voc	0.21	*	0.39	***	0.41	***	0.35	**		
eduHigh_some	0.18		0.16		0.26		0.38	**		
eduHigh	0.20	**	0.27	**	0.28	**	0.26	*		
Iraq	0.35	**	-0.12		0.01		-0.31		-0.30	
Somalia	0.65	***	0.28		0.49		0.12		0.11	
China	-0.23		-0.98	**	-0.61		-1.43	**	-1.44	**
Afghan	0.29	**	-0.01		0.19		-0.36		-0.34	
Sudan	0.61	***	0.43		0.65	**	0.23		0.25	
Jugoslavia	0.61	***	0.39		0.54	*	0.31		0.33	
SovietUni	0.38	*	0.32		0.48		0.11		0.11	
OtherC	0.51	***	0.20		0.39		-0.03		-0.03	
A_Status	0.02		0.07		0.12	*	0.05		0.04	
ysmIraq	-0.05		0.04		-0.01		0.06		0.07	
ysmSomali	-0.08		-0.07		-0.15		-0.09		-0.08	
ysmChin	0.17	*	0.26	*	0.17		0.39	*	0.40	*
ysmAfgh	-0.03		0.00		-0.08		0.09		0.11	
ysmSudan	-0.02		-0.05		-0.11		-0.03		-0.02	
ysmJugos	-0.08	*	-0.08		-0.12		-0.06		-0.06	
ysmSovU	-0.03		-0.09		-0.10		-0.05		-0.04	
ysmOther	-0.05		-0.04		-0.09		0.00		0.01	
AMA	0.10		0.00		0.10		0.15		0.13	
Undocyears	0.13	***	0.18	***	0.17	***	0.16	***	0.16	***
Statuswait	-0.14	***	-0.18		-0.19		-0.16		-0.16	
Married	0.19	***	0.12	**	0.06		0.09		0.10	
Amsterdam	0.08		0.02		0.07		0.08		0.06	
Rotterdam	0.14	*	0.18		0.08		0.05		0.08	
DenHaag	0.16	**	0.08		0.11		0.07		0.07	
Utrecht	0.03		0.08		-0.05		-0.01		0.01	
eduITSsec									0.06	
eduITSter									0.06	
Edumissing	0.13	*								
_cons	2.36	***	3.14	***	3.01	***	3.4	***	3.59	***
chi2	1502.75		568.44		420.8		356.37		345.35	
Within	0.1012		0.1607		0.1646		0.1702		0.1695	
Between	0.2436		0.2442		0.2329		0.2577		0.2467	

Overall	0.2503	0.25	0.2442	0.2609	0.2554
N	5933	2091	1603	1236	1236
Average n per person	1.6	1.7	1.7	1.7	1.7

The effect of selective observation by the Employment Service is remarkably small. The estimated coefficients differ somewhat between the full sample and the restricted sample used for Model I, but in a qualitative sense, the conclusions are not affected. The coefficients on education are very similar, except for secondary vocational education. Immigrants with that education who visit the Employment Service are much more successful than an average immigrant with that education. Of course we cannot say whether this is due to the positive influence of the Employment Service, or due to better unobserved quality of those who visit. From inspecting results for models II and III we can clearly conclude that our key conclusion on education survives: immigrants with higher education do not earn more than immigrants with lower education. In the period we observe, education acquired at home does not pay off in the Dutch labour market. Under reliability restriction II, the earnings levels for immigrants are identical for all education levels beyond some primary, with the exception of secondary vocational. Under reliability restriction III, there is equal pay for primary education, secondary vocational and some higher level education, with all other levels earning less. Model IV is even more outspoken: there is no earnings difference between immigrants with primary, secondary or tertiary education!

4.3 Selective participation

We have considered estimation of earnings functions corrected for participation using Heckman's two-step procedure. A priori we had reservations because not many variables are available and credible exclusion restrictions are hard to determine. We estimated a wage equation for wages in 2000. If the wage equation includes education and country of origin, we get unconvincing results no matter how we specify the participation equation. In particular, the effect of years since migration is negative and the dummy for women gets a positive coefficient. We decided not to pursue this approach.

4.4 Possible explanations

The key finding that higher educations acquired at home generally do not pay off during the first five years in the Dutch labour market, can be explained in several ways. One intervening variable may be language skills. It may very well be that for may of the occupations associated with higher educations understanding the Dutch language is vital, much more so than for lower levels of education. One can do cleaning work, construction work, much manufacturing work without good fluency in Dutch. One cannot be a physician without understanding Dutch properly. As we have no information on language proficiency we cannot test this. It would be quite informative to observe the jobs that immigrants hold. Unfortunately, jobs have not been recorded either. A related explanation may be certification. Several occupations that require high levels of education also require certification in the destination country. Even if one were fluent in Dutch, a qualified physician would not be allowed to take up his profession without obtaining new professional qualification in The Netherlands. Without further data, however, we cannot assess the empirical importance of these explanations.

References

Johnston, J. (1972), *Econometric Methods*, (2^{nd} edition), Tokyo: McGrawHill

Duleep, H. and M. Regets (1999), Immigrants and human capital investment, *American Economic Review*, 89 (2), Papers and Proceedings, 186-191

Kee, P (1993), The economic status of male immigrants in the Netherlands, Ph. D thesis, University of Amsterdam

Wooldridge, J. (2002), *Econometric Analysis of Cross Section and Panel Data*, Cambridge: MIT Press

Appendix A The IND/GBA/RIO data

1. *The files*

All immigration by non-Dutch citizens is registered in the Central Register Foreigners (*Centraal Register Vreemdelingen, CRV*), using information from the Immigration Police (*Vreemdelingen Politie*) and the Immigration and Naturalisation Service (*Immigratie- en Naturalisatie Dienst IND*)[6]. CBS, the Dutch Central Bureau of Statistics, has linked the data to the Municipal Register of Population (*Gemeentelijke Basisadministratie GBA*). The GBA/CRV Register includes all non-Dutch immigrants who legally entered The Netherlands during 1990-2001, except those who have returned before January 1, 1998, those who naturalised to Dutch citizenship and those who have died. As the Register takes stock every year on January 1, immigrants who left within the calendar year of arrival are also excluded. A problematic category is “administrative removal”: immigrants removed from the files of one municipality without showing up in the files of another municipality or as emigrant. Administrative removals are included among return migration. However, there is no evidence that they actually left The Netherlands. It is quite likely that many “administrative removals” remained in The Netherlands as illegal immigrant.

The GBA/CRV files have been linked to observations in the Regional Income Panel 1995-2000 (*Regionaal Inkomens Onderzoek RIO*), created by CBS. RIO is a panel of 2 million households, containing some 5 million individuals, about 30% of the population. The original GBA/CRV file covers about 600 000 individuals, from which about a third can be retrieved in the RIO panel, thus generating a GBA/CRV/RIO file of some 200 000 individuals. As immigrants in the GBA/CRV file have been linked to the RIO file in its base year 1995, the linked file covers about one third of the immigrants that have been registered in municipal population registers between 1990 and 1995 (however, the entry date in the register is not necessarily the entry date in The Netherlands, see below; registration in the municipal register is compulsory for every resident however). All immigrants registered after 1995 have been added to the data set; about one third of them could be linked to RIO. Naturalised immigrants are maintained in the RIO sample.

RIO gives panel information on disposable income and on socio-economic classification, both for individuals and for the household they belong to. The classification is based on the dominant income source during the year: employee, self-employed, on disability, social assistance or unemployment benefit, other (mostly non-participating, without an individual income). Disposable income is defined as gross income minus taxes (on income and wealth) social security premiums and other transfers (such as alimony).

Information on level of education of immigrants is available in CRV if the immigration officer has bothered to register this (immigration officers consider it mostly irrelevant for their purpose); there is also registration of education for

---

[6] Note that we only consider non-Dutch immigrants.

individuals who have contacted the government employment agency to find a (new) job, obviously a very selective group.

Observations have been weighted, with weights reflecting gender by year of birth by age (older or younger than 18) by year of arrival (since 1990) by year of exit (deceased, migrated: 1998, 1999, 2000, still present).

2. *Refugees*

Asylum migrants (refugees) enter as applicants for asylum. Registered asylum migrants are immigrants who have been admitted, ie who have obtained a title of residence (refugees with a temporary status, A status, “AMA” (independents under 18**),** admission for humanitarian reason) and immigrants waiting for a decision on their asylum application. Admitted asylum migrants in principle are always registered in GBA. Registration for asylum applicants is variable. If they are registered in GBA at all, registration takes place several months after application. Since 1998, there are two special arrangements for asylum applicants. Under *Zelfzorgarrangementen* (Independent Housing) find their own housing, with friends, relatives or otherwise. In this case they will always directly be registered in GBA. Under Central Housing, COA (*Centrale Opvang voor Asielzoekers*) takes care of housing. Asylum applicants in Central Housing are registered in GBA when they obtain asylum status or after spending one year in Central Housing (since June 2000, after spending 6 months). Most applicants were registered when they left Central Housing. This means that the group of asylum migrants contains an unknown share of asylum applicants, i.e. is an unknown mixture of admitted migrants and applicants for admission.

3. *Limitations of the dataset*

While the data set is unique in its perspective and coverage, as a follow-up on all immigrants arriving in The Netherlands, it is also imperative to point out its limitations: truncation, measurement errors, limited number of variables.

The CRV/GBA file basically includes all non-Dutch immigrants who legally entered The Netherlands during 1990 - 2001 and who have “survived” until at least January 1, 1998: they are only observed if at that date they are still living in the Netherlands as an immigrant. Thus, older cohorts of immigrants are truncated at departure (through death, emigration or naturalisation) before January 1, 1998. Moreover, information is collected on the stock of immigrants as per January 1. All immigrants leaving within the calendar year of arrival remain unobserved. This means that short durations are only observed if the interval of immigration contains January 1. In other words, precise information on short durations should be taken from durations covering January 1. This is biased information if such spells of immigration differ from spells shorter than one year that do not include January 1.

One source of measurement error is particularly disturbing. Since “administrative removal” is counted as return migration in the GBA files, while it is not at all certain that these individuals have actually left the Netherlands, return migration will include an unknown number of illegal immigrants

Information is limited to a small number of variables. GBA/CRV registers year of arrival, age, gender, country of origin, immigration motive, marital status, family composition, city of residence. RIO registers socio-economic category (employee, self-employed, disability, unemployment or welfare benefit, other), individual and household income. Category is measured from main income source during the year. Income itself is taken from fiscal records and has very high reliability. Education, a key variable, is poorly measured: in a standard classification scheme for those individuals who have visited the Employment Agency and for refugees if the Immigration Officer has bothered to fill out the entry at the application document. For the latter observations, ITS Nijmegen has coded the entries (the application document has an entry for all immigrants, but ITS only coded for refugees). We are grateful for their generous offer to add their coding to our dataset.

Appendix B Variable definitions

Observations restricted to individuals aged 15-59

Arrival.year. (instroom): year of registration IND

Settlement.year (vestiging): year of registration GBA

Age: age at arrival in The Netherlands (ie at IND or GBA registration?)

YSM: years since migration; years elapsed since registration GBA

Statuswait: arrival.year minus settlement.year, if positive, zero otherwise (year of IND registration minus year of GBA registration), hence time spent in refugee homes waiting for a decision on the application

Undocyears:: settlement.year minus arrival.year, if positive, zero otherwise (year of GBA registration minus year of IND registration); this applies when immigrants settled in the Netherlands without residential permission and without applying, undocumented immigrants could register at GBA without any sanction

Education: education as registered by IND and coded by ITS:

None
1-3 years basic
4-5 years basic
Basic
Extended basic
Secondary, general
Secondary, vocational
Some tertiary
Tertiary (higher vocational and university)
Missing

A status: permanent residential permission; date of granting status unknown

AMA: independent refugee not older than 18 at arrival

Naturalised: obtained Dutch citizenship in 2001

Returned: emigrated or administratively removed in 2001

Married: individual had marital status when arriving.

Appendix C. Assessing measurement error in education.

1. CWI variation as measurement error

	Mean	Variance
No education	7.78	8.08
1-3year Primary	8.01	8.12
4-5year Primary	8.02	8.11
Primary	8.48	9.20
Extended primary	9.31	10.97
Secondary, general	11.09	14.32
Secondary, vocational	11.23	14.37
Some Tertiary	13.29	16.03
Tertiary	15.07	13.91
Missing	11.03	19.13
average	10.33	12.22

Variance ITS: 12.73; weighted variance CWI, across rows:9.83

2. ITS variation as measurement error

CWI education									**ITS mean**	**ITS variance**
CWI opl 1995-2000		**ITS-Opl jaren**								
	0	**2**	**4**	**6**	**8**	**11**	**12**	**14**	**16**	MEAN	VARIANCE
Onbekend	0.16	0.07	0.10	0.09	0.25	0.12	0.01	0.05	0.15	7.63	27.4
BO	0.18	0.07	0.10	0.15	0.21	0.17	0.05	0.03	0.06	6.74	56.1
Ibo/MAVO	0.10	0.05	0.07	0.13	0.26	0.24	0.06	0.03	0.06	7.93	56.0
MBO/HAVO/VWO	0.02	0.02	0.02	0.07	0.16	0.36	0.10	0.09	0.16	10.74	55.8
HBO	0.02	0.00	0.00	0.01	0.04	0.19	0.07	0.11	0.56	13.83	55.5
WO	0.01	0.00	0.01	0.00	0.02	0.08	0.02	0.10	0.75	14.71	51.6
Missing	0.16	0.06	0.08	0.12	0.21	0.18	0.04	0.04	0.10	7.59	56.6