Migrant Wages, Human Capital Accumulation and Return Migration

Jérôme Adda * Christian Dustmann† Joseph-Simon Görlach‡

February 14, 2014
PRELIMINARY and VERY INCOMPLETE

Abstract

This paper analyses the wage dynamics of migrants focussing on their human capital accumulation and how it is affected by potential return migration. We develop a life-cycle model describing labor market participation, wages, return decisions as well as two forms of human capital, work experience and cultural integration. The model is estimated using panel data and exploits elicited return intentions as well as realised ones. We show that return intentions are key to understand the decision to invest in various forms of human capital and to explain differential wage paths. We show that conventional estimation methods overstate returns to work experience, as they fail to take into account selective return migration but also selective investment in human capital.

1 Introduction

2 Dataset and Descriptive Evidence

Rather than knowledge of actual migration durations, identification of the accumulation process of latent host country specific human capital and its effect on labor

*European University Institute.
†University College London.
‡University College London.

market outcomes requires in our context information on the expected length of stay of immigrants in their host country. Such information is rarely available, in particular not in longitudinal format. As such, the German Socio-Economic Panel (SOEP) is a unique survey, which has repeatedly been asking immigrants in Germany since the beginning of the panel in 1984, both whether they intend to stay in Germany forever and if not, how many more years they intend to stay.[1] The SOEP is a representative survey of private households, which oversamples a number of major immigrant groups in Germany. In addition to the planned length of stay of immigrants, it contains a large array of information on personal and household characteristics, including employment histories, income, and in some waves on household assets and annual savings and remittances. For identification of the integration process, we further use indices of subjective knowledge of German language and other variables indicating the degree of integration.

We use 18 waves (1984-2011) of the panel for our estimation. In order to focus on a more homogeneous group of immigrants, we restrict the analysis to men without tertiary education born in Turkey and residing in West-Germany, who were aged 16 or older at immigration, and who arrived in Germany between 1970 and 1984. The restriction to an immigration horizon of only 15 years is important insofar as calendar time is not a dimension in our model. This allows us to approximately account for the rapid macroeconomic development in Turkey during the past few decades that have changed the outside option for Turkish migrants, as we describe below. This leaves us with an unbalanced panel of 231 individuals and 2,513 observation points.

In the wave after each spell observed in the panel, the SOEP provides information on why the spell ended. Follow-ups of this kind are obiously difficult to implement and the information of why individuals ceased to be interviewed often is missing. And although this information is not crucial for our core story, it does–for individuals who are known to have left Germany or those who are known not to have emigrated–give an indication of how well immigrants anticipate the time they will stay in Germany, and we use this both as motivation and as a means of identification of an error in expected future shocks. Table 1 shows the relative frequencies of intended versus actual length of stay for those individuals for whom we observe whether they actually stay until at least age 65 or leave Germany during the sample period.

In table 2 we list the means of the major variable used in our analysis when immigrants state an intention to stay permanently and when indicating the intention

[1]The exact wording of the question is: "How long do you want to live in Germany? [1] I want to return within the next 12 months _____ [2] I want to stay several more years in Germany _____ number of years _____ [3] I want to remain in Germany permanently _____" (German Socio-Economic Panel, 1984).

Table 1: **intends to stay until at least 65 by observed to stay until at least 65**

	observed to stay until at least 65		
intends to stay until at least 65	no	yes	Total
no	28.56	36.35	64.91
yes	5.96	29.13	35.09
Total	34.52	65.48	100

Source: SOEP, own calculations

to return at some future point in time. Although most differences are modest, the table reveals some interesting facts. As would be expected, the age at which individuals arrive is positively correlated with the intention to return, which may be due to a stronger attachment to their countries of origin. Interestingly, individuals who consider themselves temporary migrants are more likely to be in work. This is not clear a priori, given that temporary migrants may have less incentive to invest in host country specific capital and to socially integrate. On the other hand, if the primary purpose of a temporary migration is the accumulation of savings, while a migrant's marginal utilities of consumption and leisure are higher in his country of origin, this may imply that temporary migrants are more eager to work. The data suggest the differences in employment are mostly driven by more frequent job matches for temporary immigrants. This interpretation also is supported by the on average higher annual savings and remittances of temporary migrants. Furthermore, mean earnings, the subjective knowledge of German language, the tendency to read German country of origin newspapers and the reported feeling of being German all are higher for permanent immigrants.

We augment the SOEP data with median net income of male workers without tertiary education in Turkey in 2006 from Eurostat. We interpolate to other years using time series on nominal compensation per employee provided by the European Commission and gross national income from the World Development Indicators. All monitary variables are adusted to 2005 Euros using consumer price indices and exchange rates from the Bundesbank and the OECD.

Economic trends over the time period we investigate are very different for Germany and Turkey. In particular earnings in Turkey have risen strongly relative to those in Germany, with median real net earnings of non tertiary-educated workers having increased from 5.9% to 28.4% of the corresponding level in Germany. Calendar time, however, is not a state variable in our model. In order to take into account differences in macroeconomic developments in Turkey relative to Germany, our simulation assumes that immigrants arrive in 1973, which is both the median and the mode year of immigration in our sample. The attractiveness of returning to Turkey as determined by expected earnings there and by the relative price level, then changes according to the development of these variables over time. Rather than assuming that individuals can perfectly predict all fluctuations, the simulation is based on earnings conversion factors and price levels as predicted by second order polynomials of years

Table 2: Descriptive statistics by intention to stay permanently

Variable	**Stay**	**Return**	**Total**
	(64.4%)	(35.6%)	(100%)
age	44.86	44.27	44.45
	(0.38)	(0.26)	(0.21)
years since immigration	20.31	17.74	18.77
	(0.28)	(0.17)	(0.15)
age at immigration	24.56	26.53	25.68
	(0.22)	(0.17)	(0.13)
intended length of stay	-	6.895	6.895
	-	(0.142)	(0.142)
work	0.7463	0.8018	0.7800
	(0.0153)	(0.0104)	(0.0085)
job finding	0.08889	0.1548	0.1302
	(0.02127)	(0.0228)	(0.0158)
job loss	0.05925	0.06320	0.06224
	(0.01001)	(0.00742)	(0.00588)
work experience in Turkey	6.551	8.223	7.504
	(0.221)	(0.176)	(0.133)
work experience in Germany	17.96	15.87	16.68
	(0.26)	(0.16)	(0.14)
log(real annual gross earnings)	10.25	10.17	10.19
	(0.02)	(0.01)	(0.01)
annual savings and remittances	1,560	2,269	1,899
	(101)	(101)	(70.56)
stock of assets	65,081	36,277	46,386
	(17,340)	(6,391)	(7,833)
German language	0.4970	0.4110	0.4388
	(0.0102)	(0.0067)	(0.0057)
German newspaper	0.3707	0.2817	0.3192
	(0.0134)	(0.0110)	(0.0086)
feeling German	0.3593	0.1706	0.2271
	(0.0141)	(0.0072)	(0.0069)

since 1973.[2]

3 Model

Our model follows foreign born individuals, from the time they arrive in Germany until retirement, either in Germany or in their home country. We focus on labor market decisions, human capital investments, as well as the decision to return to their home country. The model is set in discrete time, and each period lasts one year. Key to the model are two forms of human capital, work experience and German social capital. The former is acquired in a learning-by-doing way, while the latter is acquired by active time investment, and corresponds to language and German specific cultural assimilation. We allow both forms of human capital to affect wages in a Mincer type model, and we allow for synergies between those two forms of human capital.

The individual makes decisions based on his age, a_{it}, years since migration, ysm_{it}, work experience, X_{it}, German social capital, Γ_{it}, assets, A_{it} as well as fixed and time varying unobserved heterogeneity, capturing ability as well as preferences towards the home country. We denote by Ψ_{it} the preference towards Germany and by α_i the ability of the agent. The state variables observed by the agents are collected in the vector $\Omega_{it} = \{a_{it}, ysm_{it}, X_{it}, \Gamma_{it}, A_{it}, \Psi_{it}, \alpha_i\}$.

3.1 Primitives of the model

Budget constraint We assume a standard intertemporal budget constraint, which relates savings to past savings, net income and consumption:

$$A_{it} = RA_{it} + net(y_{it}^G) - c_{it}, \quad A_{it} \geq 0. \tag{1}$$

We denote by $net()$ a function which describes the tax schedule.[3] If the individual is unemployed, then income is equal to unemployment benefits b. If the individual works, we denote y_{it}^G the gross wage.

[2]Precisely, the price level in Turkey relative to that in Germany at which assets are converted after return is given by $x_t = 0.7946611 + 0.0312263t + 0.0007039t^2$, where t denotes the years since 1973. The conversion factor for earnings evolves according to $\rho_t = 0.040724 + 0.0086611t - 0.000067t^2$.

[3]We use the tax schedule prevalent in 1999 and we do not attempt to model changes in the tax schedule over the period of analysis. We assume that the individual is married. The tax schedule also depends on the number of children, although the differences in taxation with respect to the number of children are tiny. Without much loss, we assume that individuals do not have children. The tax schedule is approximated by a third order polynomial in $\log(y^G)$ as $net(y^G) = 26.48847 - 7.956472\log(y^G) + .7837363(\log(y^G))^2 - .0252077(\log(y^G))^3$ ($R^2 = 0.9967$).

Determinants of wages and unemployment benefits Log gross wages are expressed as:

$$\log y_{it}^G = \alpha_i + \Gamma_{it}^{\alpha_\Gamma} f_y(X_{it}) + \varepsilon_{it}^y, \tag{2}$$

The function $f_y(.)$ is a piecewise linear function of work experience, with nodes at 10 and 20 years of experience, to allow for non-linear returns to experience. We allow for synergies between labor market experience and German social capital. Workers who are better at speaking German may get higher wages, conditional on labor market experience, as they may be working in better paying jobs. We allow the intercept of the wage equation to be individual specific, capturing heterogeneity in ability. We assume the error term ε_{it}^y to be normally distributed with mean zero and variance σ_y^2, and iid distributed across time and individuals.

Unemployment benefits are specified as $b = br \cdot y^G$, where br is predicted from the SOEP data as a third order polynomial of $\log(y^G)$.[4]

Preferences: The individual derives utility from consumption, c_{it}, and from leisure. We normalise total time to one unit. When working, agents use a fraction h_{it} of total time. The individual can also decide to spend a fraction of his leisure investing in German social capital. We denote by e this fraction of time and by I_{it} an indicator equal to one if investment takes place. The utility function, while in the host country, takes the following form:

$$u_{it}^G = \Psi_{it} \Gamma_{it}^{\phi_0} c_{it}^{\phi_1} \left(1 - h_{it} - e \cdot I_{it}\right)^{\phi_2}. \tag{3}$$

If the individual decides to go home, utility is derived from consumption and leisure, without the option of investing in German (or home specific) social capital.

$$u_{it}^T = c_{it}^{\phi_1} \left(1 - h_{it}\right)^{\phi_2}. \tag{4}$$

The terms $\Psi_{it}\Gamma_{it}^{\phi_0}$ capture relative preference for consumption in Germany. This relative preference consists of an individual specific random part evolving as an mean reverting AR(1) process, where the mean is individual specific,

$$\Psi_{it} = \rho \Psi_{it-1} + (1-\rho)\mu_i + \epsilon_{\Psi,it}, \tag{5}$$

with $\epsilon_{\Psi,it} \sim N(0, \sigma_\Psi)$.

[4] $br = 65.63975 - 19.84936 \log(y^G) + 2.013945(\log(y^G))^2 - .0680613(\log(y^G))^3$ ($R^2 = 0.9672$).

Human Capital Work experience is increased by one unit in each period the individual is working. Experience in Turkey, prior to immigration is discounted by a factor ξ. German social capital accumulates through active investment. and of the accumulated stock of the latent variable Γ, which can be interpreted as integration in Germany, and evolves as

$$\Gamma' = \Gamma + d_\Gamma \cdot I,$$

with I indicating whether an individual invests in integration. If he does, there is a utility loss (subracted from leisure) from this effort equal to e. The initial stock Γ_0 is age at immigration specific, with Γ_0 normalized to 1 for immigrants who arrived at age 16 (the youngest age at immigration category).

Labor market transitions In each period, employed workers face a risk of being laid off. We denote by $\delta(\Omega_{a_{it}})$ this probability, which depends on factors such as labor market experience, social capital, or age, which are all part of the state vector. Individuals who are unemployed receive a job offer with probability $\lambda(\Omega_{a_{it}})$ and decides, based on observation of the shock to income whether to accept the job or remain in unemployment.

Timing of choices Within a given period, we assume that the agent makes sequential decisions. First, the agent observes the shock to income and decides how much to consume and how much to save. The agent then observes the shocks to preferences towards location and decides whether to stay in the host country or return home. This choice is made conditional on income and consumption levels in either the host or the home country. Finally, at the end of the period, the agent observes shocks to labor supply (whether he loses his job if employed, or whether he is offered a job if unemployed), which determines the labor market status for the next period. Hence, consumption or migration decisions are made given the labor market status at the start of the period, but conditional on the expected labor market status in the next period.

Return migration We assume that the decision to return to the home country is final. The individual does not have the option to come back to the host country. Our model does not apply to seasonal workers, which are excluded from the sample we analyse. When back in the home country, we assume that the individual is always in work, although we assign a wage which is a weighted average of wage levels and zero income, with weights taking into account the risk of unemployment. We do so, as we do not observe individuals when they return and we only have semi-aggregate

data on wages and unemployment rates.[5] We model income in the home country as a fraction of the wage in the host country, when social capital is set to zero.

$$y_{it}^T = \rho_y y_{it}^G,$$

where the conversion factor ρ_y is the income ratio at purchasing power parity between the home and host country. Once migrants return, their assets are converted by a factor x (see the exact specification below).

Intentions We model the intended length of stay by simulating $S = 50$ future paths s of realizations of shocks to earnings, employment and preferences. Each of these paths corresponds to an optimal migration duration, and we define the median of these as the intention stated by an individual at a given age. Individuals are assumed to intend to stay forever if their intended age at return exceeds age 64.

3.2 Dynamic specification of the model

We define the inter temporal flow of utility (value function) of returning in the home country by $V^T()$ and by $V_L^G()$ the value function for staying (at least) an additional period in the home country. We index this value function by $L = U, W$, indicating whether the individual is working or not. We do so, as labor market status implies different choice sets. The value of working is defined as

$$\begin{aligned} V_W^G(\Omega_a) &= \max_{c,e} u^G(c, 1-h-eI, \Gamma, \Psi) \\ &\quad +\beta \mathrm{E}[(1-\delta(\Omega_{a+1}))\max\{V_U(\Omega_{a+1}), V_W(\Omega_{a+1})\} \\ &\quad +\delta(\Omega_{a+1})V_U(\Omega_{a+1})], \end{aligned} \tag{6}$$

where E is the expectation operator, which denotes expectations over future shocks. The individual choses consumption and investment in social capital to maximise the flow of current and future utility. We denote by β the discount factor. The value of being in unemployment is defined as:

$$\begin{aligned} V_U^G(\Omega_a) &= \max_{c,I} u^G(c, 1-eI, \Gamma, \Psi) \\ &\quad +\beta \mathrm{E}[\lambda(\Omega_{a+1})\max\{V_U(\Omega_{a+1}), V_W(\Omega_{a+1})\} \\ &\quad +(1-\lambda(\Omega_{a+1}))V_U(\Omega_{a+1})]. \end{aligned} \tag{7}$$

[5] The consequence of this assumption is that individuals do not save to buffer income risk in the home country, but only save to sustain higher consumption levels while working or in retirement.

The value of being back in the home country is defined as:

$$V^T(\Omega_a) = \max_c u^T(c, 1-h) + \beta \mathrm{E}\left[V^T(\Omega_{a+1})\right],$$

Finally, the decision to stay or return is modelled as:

$$V_L(\Omega_a) = \max\left\{V_L^G(\Omega_a) + \eta^G, V^T(\Omega_a) + \eta^T\right\}, \quad L = U, W, \tag{8}$$

where η^G and η^T are two iid extreme value distributed shocks to preferences. Such distributions ensure that the probability of returning takes a logistic form, with value functions in the home and host country as arguments.

Individuals retire at age 65, and from this point until the end of life at age 80 only make consumption decisions, with state variables X, Γ and Ψ frozen at their values at age 64. During this period, individuals receive retirement benefits $y^R = 0.7\tilde{y}$, with $\tilde{y} = ((X/(1+\xi/2))/40) \cdot y^G + (1 - ((X/(1+\xi/2))/40)) \cdot y^T$, based on the assumption that an individual's working live lasts 40 years and that by the time of return, 1/3 of total experience has been accumulated in the country of origin.

3.3 Intended length of stay

The probability that a migrant returns to his home country at any given age a conditional on labour market status is

$$P(\Omega_a) = \begin{cases} \frac{\exp(V^T(\Omega_a))}{\exp(V^T(\Omega_a)+\exp(V_W^G(\Omega_a))} & \text{if working} \\ \frac{\exp(V^T(\Omega_a))}{\exp(V^T(\Omega_a)+\exp(V_U^G(\Omega_a))} & \text{if not.} \end{cases}$$

We compute the intended length of stay by simulations. We define the *median* age at return a_r as the *intended* point of return, so that the intended length of stay ς given states (and employment status E_a) at age a is

$$\varsigma(\Omega_a, E_a) = m \quad \text{s.t.} \quad \frac{1}{S}\sum_{s=1}^{S} \mathbf{1}\left[\sum_{j=0}^{65} j\mathbf{1}_s\left[\text{return at } a+j|\Omega_a, E_a\right] \leq m\right] = \frac{1}{2},$$

where at age a, $\mathbf{1}_s\left[\text{return at } a+j|\Omega_a, E_a\right]$ indicates whether simulation s predicts the migrant to return at age $a+j$ given current states. This allows us to have a theoretical counterpart to the stated return intentions of the individuals we observe.

4 Estimation

We estimate our model using a simulated method of moments estimator. We detail below its implementation.

4.1 Dimensions and other specifics

The simulation makes the following assumptions:

- individuals immigrate at ages 16, 28 or 40, where the probabilities according to which the age at immigration for each individual is drawn is taken from the data
- the working experience in Turkey prior to immigration is drawn by age at immigration from the empirical distribution
- the working experience in Germany accumulated by age 18 for individuals who immigrated at younger age is drawn from the empirical distribution
- since we do not have calendar time in the model, we restrict the empirical sample to individuals who immigrated during druing 1970-1984. We use predicted values of the income ratio from a regression of the income ratio on squared years since 1973 and assign values to the simulated individuals under the assumption that they immigrate in 1973, which for this sample is both the median and the mode year of arrival.
- the real exchange rate is calibrated in the same way as the income ratio.
- there are four heterogeneity types τ, each two for the earnings intercept α_τ and the preference component Ψ_τ
- the asset grid has four equally distributed points of support from 0 to 300,000 Euros (interpolated inbetween)
- individuals choose between consuming 80%, 90%, 100%, 110% or 120% of their income
- the grid for integration (Γ) ranges in intervals of 0.5 from 0 to 2 (interpolated inbetween)
- the grid for working experience is in intervals of 7 years (interpolated in between)

- the shock to income, ε_y has two points of support
- the integration variables ι we observe (German language knowledge (L), reading German newpapers (N), and feeling German (F)) are determined as

$$\iota = \begin{cases} 0 & \text{if} \quad \gamma_0^\iota + \gamma_1^\iota \Gamma + \epsilon^\iota \leq 0 \\ \gamma_0^\iota + \gamma_1^\iota \Gamma + \epsilon^\iota & \text{if} \quad 0 < \gamma_0^\iota + \gamma_1^\iota \Gamma + \epsilon^\iota < 1 \\ 1 & \text{if} \quad \gamma_0^\iota + \gamma_1^\iota \Gamma + \epsilon^\iota \geq 1, \end{cases}$$

  with $\epsilon^\iota \sim N(0, \sigma_\iota^2)$ for $\iota \in \{L, N, F\}$.

4.2 Empirical moments

Idenfication relies on the following data moments:

- coefficients in a regression of log annual earnings on a spline of work experience in Germany, work experience in Turkey, a spline of age and a constant, the standard deviation of the residual (denoted sd(ia) in the graphs below) and the standard deviation of the within individual mean residual (sd(i))
- coefficients in a regression of the growth in log annual earnings on last year's intention to stay permanently (denoted L.Stay in the graphs), total work experience and a constant, and the standard deviation of the residual
- coefficients in separate linear regressions of log annual earnings, on one of the three integration variables (language, newspaper and feeling German) and a constant, the standard deviation of the residual (sd(ia)) and the standard deviation of the within individual mean residual (sd(i))
- coefficients in linear regressions of switching (1) from not working to working and (2) from working to not working, each on work experience in Germany, work experience in Turkey, a spline of age and a constant, and the standard deviations of the residual
- coefficients in linear regressions of switching (1) from not working to working and (2) from working to not working, both for each one of the three integration variables (language, newspaper and feeling German) and a constant, and the standard deviations of the residual

- coefficients in a linear probability model of being in work on a spline of work experience in Germany, work experience in Turkey, a spline of age and a constant, and the standard deviation of the residual
- coefficients in regressions of the indices for (1) German language knowledge, (2) reading German newpapers, and (3) feeling German, each on intention to stay permanently, years since immigration, an interaction of the two and a constant, and the standard deviations of the residual
- coefficients in a regression of the stock of assets on age, indicators for age at immigration and a constant, and the standard deviation of the residual
- coefficients in a regression of the annual savings on intention to stay permanently, an interaction of a working indicator and annual income, the working indicator and a constant, and the standard deviation of the residual
- coefficients in a regression of intended length of stay on a working indicator, an interaction of the working indicator and annual income, an interaction with squared annual income, years since immigration and a constant, the standard deviation of the residual (denoted sd(ia) in the graphs below) and the standard deviation of the within individual mean residual (sd(i))
- coefficients in a regression of intended length of stay at age 35 on the intended length of stay at age 25 and a constant, and the standard deviation of the residual (sd(ia))
- coefficients in a regression of actually returning next period on age, a working indicator and a constant, and the standard deviation of the residual
- coefficients in a regression of the actual time until return on the intended length of stay, years since immigration and a constant, and the standard deviation of the residual.

All splines have have three intervals: those of experience for $X \in [0, 10], X \in [11, 20]$ and $X > 20$; those for age for $a \in [18, 30], a \in [31, 50]$ and $a > 50$. Year indicators are included in all regressions. Since the empirical moments involving intentions are generated by pooling individuals at ages 23-27 (corresponding to age 25 in the simulation) and ages 33-37 (corresponding to age 35 in the simulation), age indicators are included in all regressions that involve intentions.

5 Estimation Results

The estimated parameters are displayed in Table 3.

6 How Good are Reduced-form Estimation Methods?

7 Conclusion

Table 3: Estimated Parameters

Parameter	Estimate
Value shocks variance parameters:	
s_W	1.72146216
s_U	0.332143666
sP_W	1.4963208
sP_U	0.401586894
Job finding parameters:	
λ_0	-6.49048549
λ_1	0.143188078
λ_2	0.220009001
λ_3	-0.225828813
λ_4	-0.525827027
λ_5	0.245831933
Job loss parameters:	
δ_0	-4.43114785
δ_1	-0.034034386
δ_2	0.0968500064
δ_3	0.0661325563
δ_4	0.0813686408
δ_5	-0.175386552
Locational preference AR(1):	
ρ	1.01631569
μ_i	0.12003367
	0.926784175
σ_Ψ	0.242374851
Utility function parameters:	
ϕ_0	1.00789738
ϕ_1	0.174616413
ϕ_2	0.752099297
h	0.166579951
e	0.532304
Type probabilities:	
$P(\tau_1)$	0.5
$P(\tau_2)$	2.26598206e-013
$P(\tau_3)$	2.26598206e-013

Parameter	Estimate
Income equation parameters:	
α_i	9.52237585
	9.87337292
α_1	0.0419463779
α_2	0.0127829897
α_3	0.00961058029
α_4	0.853083527
Initial stock of assets:	
$A_0(iage = 16)$	0
$A_0(iage = 28)$	943.554161
$A_0(iage = 40)$	30947.4323
Normal shock standard deviations:	
σ_y	0.238505312
σ_L	0.304598491
σ_N	0.800623394
σ_F	0.314542059
Integration outcome parameters:	
γ_0^L	0.225525066
γ_1^L	0.0864990948
γ_0^N	0.765124102
γ_1^N	0.624126457
γ_0^F	0.172273856
γ_1^F	0.0935302631
Starting values and step of integration:	
$\Gamma_0(16)$	1
$\Gamma_0(28)$	0.716989501
$\Gamma_0(40)$	1.09549485
d_Γ	0.0394270807
Efficiency of home country experience:	
ξ	0.365241599