I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 10887

A Joint Hazard-Longitudinal Model of the Timing of Migration, Immigrant Quality, and Labor Market Assimilation

Apoorva Jain
Klara Sabirianova Peter

JULY 2017

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 10887

A Joint Hazard-Longitudinal Model of the Timing of Migration, Immigrant Quality, and Labor Market Assimilation

Apoorva Jain
University of North Carolina–Chapel Hill

Klara Sabirianova Peter
University of North Carolina–Chapel Hill, IZA and CEPR

JULY 2017

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

A Joint Hazard-Longitudinal Model of the Timing of Migration, Immigrant Quality, and Labor Market Assimilation*

This paper develops and estimates a joint hazard-longitudinal (JHL) model of the timing of migration and labor market assimilation – two processes that have been assumed to be independent in the existing literature. The JHL model accounts for the endogenous age of entry in estimating the returns to years since migration by allowing cross-equation correlations of random intercepts with individual rates of wage assimilation. Commonly ignored sample selection issues due to non-random survey attrition and missing wages are also addressed. Using German household panel surveys from 1984 to 2014 and home country-level data from 1961, we find large upward bias in the OLS-estimated average rate of wage assimilation. Our estimates suggest that immigrants with lower unobserved skills and with a higher unobserved propensity to migrate early have a faster assimilation rate.

JEL Classification: J24, J31, J61, N30, C41

Keywords: migration, joint hazard-longitudinal model, mixed effects, random slope, individual-specific wage assimilation, unobserved skills, survival analysis, timing of migration, maximum likelihood, selection due to endogenous entry, Germany

Corresponding author:
Klara Sabirianova Peter
Department of Economics
Carolina Population Center
University of North Carolina-Chapel Hill
Chapel Hill, NC 27599
USA

E-mail: kpeter@unc.edu

* We are thankful to Luca Flabbi, Donna Gilleskie, Helen Tauchen, and Ju Hyun Kim for their valuable feedback and to Vishal Verma for very helpful mathematical assistance. We are also grateful to the late Tiago Pires for his inspiration and support. He will be sorely missed and always remembered. This paper has benefited from comments and suggestions received during the Applied Micro Workshop at the UNCChapel Hill and numerous conferences. We are grateful to the Carolina Population Center and its NIH Center grant (P2C HD050924) for general support.

1. Introduction

The current European migration crisis caused a resurgence of interest in the labor market performance of immigrants and their assimilation into the workforce (Kahanes and Zimmerman, 2016). The key statistic that economists often use for tracking the rate of immigrants' assimilation in the labor market is the average wage growth over the time spent in the host country, which is typically measured as the slope coefficient on years since migration in a standard Mincerian wage equation (Chiswick, 1978; Kerr and Kerr, 2011). This seemingly simple estimation procedure is plagued with serious methodological challenges. Our study attempts to address some of the methodological issues by proposing the joint hazard-longitudinal (JHL) model of the timing of migration and wage assimilation.

We first note that the length of stay is age minus age-at-migration, and thus it depends upon the timing (age) of migration. When deciding at what age to migrate, forward-looking individuals are likely to take lifetime earnings into consideration. In other words, the timing of migration is an endogenous choice variable. It is influenced by unobserved ability and other unmeasured wage factors, leading to "ability bias" in the OLS-estimated average rate of labor market assimilation. The problem is further complicated by individual heterogeneity in the slopes of earnings progression, which may be jointly determined with the timing of migration. Anticipated post-migration wage growth may influence the decision of when to migrate, potentially causing "slope-heterogeneity bias" in the OLS estimation. Thus, in the wage equation, both individual-specific intercepts and slopes are correlated with each other and with the unobserved factors that determine the timing of migration.

Previous literature has dealt mainly with ability bias by accounting for initial earnings at the time of entry as a proxy for unobserved immigrant quality (Borjas, 1987), by using arrival cohort fixed effects to control for the average cohort quality (Borjas, 1995; Antecol *et al.*, 2006), by controlling for a variety of macro-level data from the country of origin (Cobb-Clark, 1993), and by including individual fixed effects in the panel data (Fertig and Schurer, 2007).[1] All of these approaches implicitly assume that unobserved immigrant quality affects the wage level but not wage growth. That is, the individual-

[1] It seems that the literature on the endogenous duration of stay has advanced much further with respect to non-random exit than non-random entry; see a very detailed review on selective out-migration in Dustmann and Görlach (2015). Some advanced methods include life-cycle dynamic modelling of out-migration and assimilation (Bellemare, 2007) and joint modelling of migration durations and unemployment spells with correlated random effects (Bijwaard *et al.*, 2014).

specific assimilation slope does not vary with immigrant quality. Another important assumption is that the rate of assimilation is independent from unobserved factors influencing the timing of migration.

The approach that we use to deal with the selective timing of entry is to simultaneously estimate the survival and longitudinal processes by explicitly modelling the correlation between unobserved factors affecting wage level, wage growth, and the timing of migration. The JHL model links the survival and longitudinal processes through three correlated random effects: a random intercept ("immigrant quality" or "unobserved skills") and a random slope on years since migration in the wage equation and the frailty term in the timing of migration equation. The wage equation is a linear mixed-effects model where the log of wage depends on various individual-specific factors, including years since migration and unobserved immigrant quality. A random coefficient on years since migration defines the individual-specific rate of wage progression, also known as the rate of labor market assimilation. The timing of migration equation is a parametric proportional hazard model, in which the hazard of migration at a given age depends upon pre-migration accumulation of human capital, other observed individual characteristics, macro-level factors of migration, and unobserved individual heterogeneity (or the exponential of frailty). Since the survival analysis can be only carried out on individuals who would eventually migrate and earn wages, the study treats the random effect in the hazard model as the unobserved individual propensity to migrate early versus late. This puts certain limitations on the interpretation of model parameters, which we discuss in Section 2 below.

We recognize that the selective timing of migration is not the only selectivity issue in estimating the assimilation slope. Selection bias due to non-random out-migration has long been identified and well researched in the economics of migration (Constant and Massey, 2003; Dustmann and Glitz, 2011; Dustmann and Görlach, 2016). The issue of selection due to survey participation and panel attrition in the wage equation is less studied (Bellemare, 2007; Fertig and Schurer, 2007). As unbelievable as it may seem, the accounting for the non-random selection into employment and wage reporting in predicting the earnings profiles of immigrants is practically non-existent. Since these selection processes are not the focus of our study, we only make simple adjustments for out-migration, panel attrition, survey design, employment participation, and wage response by utilizing the method of inverse propensity weighting with several exclusion restrictions.

This paper uses data on immigrants from the 1984-2014 German Socio-Economic

Panel (GSOEP). Since the 1950s, Germany has had a long and diverse history of immigration, and for this reason it provides an excellent place to study the labor market assimilation of immigrants. As of 2017, Germany hosts more than 12 million immigrants, which is the second highest stock of immigrants in the world after the United States. For more than 8,000 immigrants, the GSOEP provides information on the country of origin, year of migration, and life history calendars between the ages of 15 and 65. Using this information, we construct pre-migration histories that are comprised of individual time-varying records on schooling and employment from before migration as well as macro-level migration factors for each home country from 1961 to 2014. These pre-migration records combined with 31 years of survey observations make for one of the longest panel data on immigrants and allow for estimating long-term assimilation rates, which by itself is a contribution to the existing literature on migration.

Most macro-level factors and pre-migration individual characteristics appear to be statistically significant predictors of the timing of migration. For example, we find that individuals choose to immigrate at a younger age when they experience episodes of warfare and political violence at home, live in a less developed country, come from a geographically close location, have closer ethnic ties with Germany, and expect higher economic growth in the host country. Using a unique measure of the Levenshtein linguistic distance between the primary language(s) of home and host countries, we find that immigrants from countries with closer linguistic proximity to Germany tend to immigrate at a younger age. A particularly strong positive effect on the probability of early-age migration is observed for countries with which Germany has signed guest-worker treaties to meet the rising demand for low-skilled labor. One of the benefits of joint modelling is that it can predict the implied lifetime wage gains/losses from the change in median age-at-migration associated with each covariate, on top of the direct wage returns estimated in the wage equation.

Our results suggest that the exogeneity assumption for years since migration leads to an upward bias in the OLS-estimated average rate of assimilation. After accounting for the endogenous timing of migration, the average rate of assimilation drops from a 1.06 percent to 0.71 percent increase in lifetime wages for each year of stay in the host country. The analysis attributes such a large difference in the estimates to the upward slope-heterogeneity bias outweighing downward ability bias. Upward slope-heterogeneity bias is a result of the positive correlation we find between individual rates of labor market assimilation and the predicted unobserved propensity to migrate early. Ability bias arises from the correlation between unobserved ability and the unobserved propensity for early migration. In the JHL model, this correlation is estimated to be weakly negative, implying that high-ability immigrants tend to postpone their migration decision. The cross-equation correlations we found between the predicted components of unobserved heterogeneity validate the need for the joint estimation of survival and longitudinal processes.

Another important finding from the JHL model estimation is that immigrants with a lower level of predicted unobserved skills tend to have a faster rate of wage assimilation, holding observed skills constant. This finding supports the hypothesis of conditional convergence or catch-up between low-quality and high-quality immigrants of similar levels of education.[2] It is complementary to existing empirical evidence on lower assimilation rates among immigrants with higher entry earnings, which are often used as a proxy for either immigrant quality or transferable skills (Borjas, 1999; Duleep and Regets, 1999). Overall, the study reveals substantial individual-level heterogeneity in all three estimated random effects. For example, the 95 percent range of estimated assimilation slopes is between a −4.5 percent and +5.9 percent wage increase for each additional year of staying in the host country, with negative rates of assimilation being observed more frequently among late arrivals.

The rest of the paper proceeds as follows. Section 2, after presenting a simple theoretical model of the timing of migration, illustrates the endogeneity problem in the years since migration variable and outlines the JHL model of the timing of migration and wage assimilation. Section 3 introduces data sources and variables included in the JHL model. Section 4 discusses results of the JHL model estimates, and Section 5 summarizes the paper.

2. Joint Model of the Timing of Migration and Wage Assimilation

2.1. Theoretical Motivation

To motivate joint estimation of wages and timing of migration, we first present a toy model formalizing main determinants of the timing of migration and highlight the issue of endogeneity with regard to age-at-migration in the wage equation. Consider an individual who decides to migrate to a different country. The individual is forward looking and maximizes expected lifetime earnings. Assume that an individual has an infinite

[2] Positive post-migration wage growth and the wage convergence between low- and high-skilled immigrants do not necessarily result in wage convergence between immigrants and natives. Using the same dataset, Jain and Peter (2017) show how the wage convergence among immigrants can co-exist with widening the native-immigrant wage gap over the life cycle.

horizon and discounts future earnings flows at a constant discount rate r. Migration occurs at age T and incurs a one-time migration cost, ς, at the time of migration. The net present value of lifetime earnings is given by

$$V = \int_0^T \overline{w}_o exp^{-rt}\, dt + \int_T^\infty w_h(t) exp^{-rt}\, dt - \varsigma exp^{-rT} \tag{1}$$

The first term represents the discounted stream of earnings in the country of origin till the age of migration, T. The second term represents the discounted stream of earnings in the host country from the age-at-migration and forward. The last term is the present value of the cost of migration. Assume an individual earns a constant wage $\overline{w}_o$ each year until he emigrates from the home country. After migration, an immigrant receives a starting salary, $\overline{w}_{hT}$, which subsequently grows at a yearly rate g:

$$w_h(t) = \overline{w}_{hT} exp^{g\times(t-T)} = [p(T)\tau K] exp^{g\times(t-T)} \tag{2}$$

The starting salary in the host country depends upon the pre-migration endowment of skills K accumulated in the home country. However, only a fraction τ of pre-migration skills can be transferred to or marketed in the host country. We assume that the skill price per unit of transferable human capital $p(T)$ changes with age-at-migration, such that $p'(T) > 0$ and $p''(T) < 0$.

Substituting Equation (2) in Equation (1) and solving integrals yield:

$$V = \frac{\overline{w}_o(1 - exp^{-rT})}{r} + \frac{p(T)\tau K exp^{-rT}}{r - g} - \varsigma exp^{-rT} \tag{3}$$

The first order condition with respect to T is satisfied with equality:

$$\left(\overline{w}_o + r\varsigma + \frac{p'(T)\tau K}{r - g}\right) exp^{-rT} = \frac{rp(T)\tau K}{r - g} exp^{-rT} \tag{4}$$

The left-hand side of Equation (4) represents the marginal benefits of migrating one year later. Benefits include one-year discounted wage in the origin country, the postponed cost of migration, and an increase in starting wage from later migration. The right-hand side gives the marginal cost of migrating a year later. The marginal cost includes foregone one-year entry earnings in the host country.

Migration does not occur when the marginal cost exceeds the marginal benefit at any age. For the optimal age-at-migration T^* to exist, the marginal benefit must be equal to the marginal cost of migrating, and the following conditions are necessary: $rp(T) >$

$p^{'}(T)$ and $r > g$, since $\overline{w}_o + r\varsigma > 0$.

Assuming that the second order conditions are satisfied, several predictions can be made:

$$\frac{\partial T}{\partial \overline{w}_o} = \frac{(r-g)}{\Delta} > 0 \quad \text{(a)} \qquad \frac{\partial T}{\partial \tau} = \frac{-[rp(T) - p'(T)]K}{\Delta} < 0 \quad \text{(c)}$$
$$\frac{\partial T}{\partial c} = \frac{r(r-g)}{\Delta} > 0 \quad \text{(b)} \qquad \frac{\partial T}{\partial g} = \frac{-(\overline{w}_o + r\varsigma)}{\Delta} < 0 \quad \text{(d)} \tag{5}$$

where $\Delta = [rp'(T) - p''(T)]\tau K > 0$.

The comparative statics predict that an individual chooses to immigrate earlier when they face worse income opportunities in home country ($\overline{w}_o \downarrow$), incur lower costs of migration ($\varsigma \downarrow$), transfer a higher share of preexisting skills ($\tau \uparrow$), and expect better prospects ($g \uparrow$) in the host country. These predictions are intuitive and testable if the right data exist.

The comparative statics prediction with respect to preexisting skills K is more complicated, as K is likely to influence positively home-country earnings $\overline{w}_o$ (which we did not emphasize earlier to keep the model simple). Previous studies show that the sign of the correlation between the initial level of human capital and the rate of wage growth is ambiguous: it is negative (i.e., $g^{'}(K) < 0$) if pre- and post-migration human capital are substitutable or positive if they are complementary (Borjas, 1999, 2015).[3] If all three channels via $\overline{w}_o$, g, and $\overline{w}_{hT}$ are taken into the account, then the overall net effect of K on the timing of migration becomes ambiguous.

The main implication of the theoretical model is that the decision on when to migrate is an endogenous choice. Since the decision depends upon the stream of lifetime earnings, the age of migration is likely to be correlated with the unobserved wage component. We show below that without accounting for the selective timing of entry, the empirical model of immigrants' wage assimilation over the length of stay in the host country is likely to be misspecified.

2.2. Selective Timing of Migration in the Wage Assimilation Model

Next, we illustrate how the failure to account for the selective timing of migration

[3] To keep the theoretical model simple in its message, we do not model the post-migration investment in human capital. In a study with different goals, the model can be extended by including the cost of re-training and new skill acquisition into the cost of migration and allowing the wage growth g to depend on new investment. Alternatively, other classes of theoretical models linking migration and education could be formulated; some of these models are described in Dustmann and Glitz (2011).

can lead to a biased estimate of the rate of wage assimilation. A typical empirical model of the economic assimilation of immigrants is estimated using the following wage equation (see Kerr and Kerr (2011) for a survey):

$$w_{is} = a_0 + \bar{\delta} YSM_{is} + \beta X_{is} + \mu_i + \varepsilon_{is}, \tag{6}$$

where w_{is} is the log of the wage of individual i at survey time s; X_{is} is a vector of immigrant's observed characteristics in the host country that often includes age (or potential labor market experience) and education; time effects are also part of the vector X_{is} when a study employs repeated cross-sections or panel data; YSM_{is} denotes years since migration, calculated as the difference between age at time s and age-at-migration; and μ_i captures time-constant cohort-specific unobserved heterogeneity. A few panel studies like Fertig and Schurer (2007) include time-constant individual heterogeneity instead of μ_i.

$\bar{\delta}$ is the average wage return on spending an additional year in the host country. It represents the rate of assimilation where assimilation is defined in a way similar to LaLonde and Topel (1992): "assimilation occurs, if between two observationally equivalent persons, the one with greater time in the United States typically earns more". Thus, the base group is the immigrant himself, and a positive value of $\bar{\delta}$ does not indicate that immigrant earnings are converging to their native counterparts.[4] This parameter captures the population average of individual wage trajectories in the host country. Individual slopes may deviate from the population average for a variety of reasons, including differences in the timing of migration, initial endowments of skills, efforts put in learning host-country's language and employment training, personality traits influencing the process of integration in a new country, among many others.

Let us rewrite the wage assimilation model in Equation (6) in a more generalizable form, where individual heterogeneity influences both the intercept of the wage equation (via a_i) and the slope of the immigrants' earnings progression (via b_i).

$$w_{is} = a_0 + \left(\bar{\delta} + b_i\right) YSM_{is} + \beta X_{is} + a_i + \varepsilon_{is} \tag{7}$$

where b_i is the individual-specific deviation from the average rate of assimilation with zero mean; a_i is a random intercept capturing the immigrants' unobserved quality with zero mean; $\varepsilon_{is} \sim N(0, \sigma_\varepsilon^2)$ is i.i.d. error independent of b_i's and a_i's. This equation belongs to the class of linear mixed-effects models.

As we saw in subsection 2.1, the decision on when to migrate is not random, and

[4] In a related paper by Jain and Peter (2017), we examine immigrants' rate of assimilation with respect to natives and find a wage divergence between natives and immigrants in the GSOEP data.

wage is an important factor of this decision. The length of stay in the host country is likely to be correlated with both the individual-specific intercept and slope, resulting in a biased estimate of the immigrants' earnings progression. In a similar model with a different application to the returns to schooling, Card (2001) derives the sources of bias in the OLS estimation using simple linear projections of a_i and b_i such as

$$\begin{aligned} a_i &= \lambda_0(YSM_i - \overline{YSM}) + u_i \\ b_i &= \psi_0(YSM_i - \overline{YSM}) + \nu_i \end{aligned} \tag{8}$$

where $\overline{YSM}$ is average years since migration and $E[YSM_i u_i] = E[YSM_i \nu_i] = 0$. Substituting linear projections into (7), the mixed-effects model of wage assimilation can be rewritten as

$$w_i = constant + \left(\bar{\delta} + \lambda_0 - \psi_0\overline{YSM}\right)YSM_i + \psi_0 YSM_i{}^2 + \beta X_{is} + \nu_i YSM_i + u_i + \varepsilon_i \tag{9}$$

In a log wage regression on years since migration, the probability limit of the OLS slope is

$$plim\ b_{OLS} = \bar{\delta} + \lambda_0 + \psi_0\overline{YSM} + \xi \tag{10}$$

where $\xi = E[b_i(YSM_i - \overline{YSM})^2]/Var(YSM_i)$. Card (2001) shows that the last term ξ depends on the third central moment of YSM_i, and it becomes zero when the residual component $\nu_i YSM_i$ is not correlated with YSM_i.

Equation (10) highlights two main sources of bias in the OLS-estimated rate of wage assimilation: (i) ability bias λ_0 due to the correlation of years since migration with unobserved immigrant ability and (ii) slope-heterogeneity bias $\psi_0\overline{YSM}$ arising from the correlation between years since migration and heterogeneous rate of wage assimilation in the host country. It is difficult to predict a priori the sign of bias. Younger immigrants have a generally lower opportunity cost of investing in host country's human capital. This implies that individuals who immigrate at a young age are likely to have a steeper slope of earnings progression in the host country compared to the same-age individuals who immigrate later in life. Hence, if we control for individual age at the time of survey, slope-heterogeneity bias is expected to be positive ($\psi_0 > 0$). However, the sign of ability bias can go either way. If high-ability individuals arrive in the host country at a young age ($\lambda_0 > 0$), then OLS regression would overestimate the wage return on a year spending in the host country. If high-ability immigrants tend to postpone their migration decision ($\lambda_0 < 0$), then there is a theoretical possibility for the OLS-estimated rate of wage assimilation to approach the true population average rate if $\lambda_0 = -\psi_0\overline{YSM}$ and $\xi = 0$.

Previous studies have treated YSM_{is} as exogenous. We note that the fixed-effect estimator does not help in treating bias. First, since age-at-migration is constant for an individual, years since migration and age become perfectly collinear and the two effects are not separable when individual fixed effects are included. Second, the fixed-effect estimator assumes that the rate of assimilation is uncorrelated with both unobserved immigrant quality and unobserved factors affecting the timing of migration. To address the issue of endogeneity, we propose a joint hazard-longitudinal model of the timing of migration and wage assimilation. In this model, the unobserved individual component of the hazard of early migration is allowed to correlate with both the individual-specific intercept and random slope in the wage equation.

2.3. Joint Likelihood

Years since migration in the wage equation is a direct function of the endogenous age-at-migration that we model in a survival framework. Using the language of duration models, let us define age as a continuous random duration variable t, age-at-migration as "the time of failure", and migration itself as a "failure event". The hazard of migration at age t, $\lambda_i(t)$, is defined as the instantaneous probability of migration given an individual i did not immigrate earlier. It captures the whole history of the migration decision process at each age prior to migration.

We model the probability of migration at age t as a continuous-time proportional hazard function given as:[5]

$$\lambda_i(t) = \lambda_0(t)\, exp\{c_0 + \gamma Z_i(t) + c_i\}, \tag{11}$$

where $\lambda_0(t)$ is a baseline hazard function, which is either left unspecified as in the Cox proportional hazard model, or assumed to be a non-linear function of age following some parametrized distribution such as Weibull distribution, $\lambda_0(t) = \varphi t^{\varphi-1}$, or Gompertz distribution, $\lambda_0(t) = exp(\varphi t)$.

$Z_i(t)$ is a vector of observed covariates (both individual and country-level) that shift the hazard of migration. Some of these covariates are time-varying, including pre-migration education history, changes in employment, as well as country-level push-pull factors of migration. Section 3 describes the external covariates in detail. In the proportional hazard model, the effect of external covariates on the hazard of migration is multiplicative.

[5] We treat age as a continuous variable. However, our data are more like interval-censored when the event is observed at some time during a one-year interval. At this point, we do not have a way of handling interval-censored data in the joint modelling framework in the presence of time-varying covariates.

c_i captures time-constant unobserved individual heterogeneity. Hazard functions with the random effects estimator are commonly known as shared-frailty models, $\lambda_i(t|\eta) = \eta_i \lambda_i(t)$, where $\eta_i = exp(c_i)$. The distribution of c_i can be either parametrized or have a non-parametric representation as in Heckman and Singer (1984). To allow for the correlation between c_i and b_i (individual-specific assimilation rates) and between c_i and a_i (immigrant quality), we assume that c_i has a normal distribution with zero mean and finite variance, σ_c^2, $c_i \sim N(0, \sigma_c^2)$; c_i is also assumed to be independent of Z_i, which is a standard assumption in the models with random effects.

We assume that individuals become at risk of migration at age 15. This is partly due to the lack of data on schooling history before age 15, but mainly due to the desire to avoid modelling parental decisions on child migration. Some individuals enter our sample late at ages older than 15. Generally, delayed entry or left truncation can be an issue in the estimation of shared frailty models when frailty c_i is correlated with the truncation point (Van den Berg and Drepper, 2016). However, in our case, it is plausible to assume that the frailty and truncation point are not correlated, since the only reason for left truncation is the lack of macro pull-push factors for years prior to 1961.

We observe only individuals who eventually migrate to Germany. Our sample is right truncated, since it consists solely of individuals who have experienced the failure event by a specified time. In a typical survey from the host country G, information is not available on stayers in the home country, including potential immigrants who at the margin were indifferent between migrating to country G or other options, but chose not to migrate to country G. The latter group is of special interest to decision makers in the destination country, as the migration outcome of potential immigrants could have changed if a different policy is implemented. However, we cannot make inferences with respect to potential immigrants unless the strong assumption of the similarity in both observed and unobserved characteristics between actual and potential immigrants is made. Without information on potential immigrants, we can only speak to the impact of covariates on the timing of migration or the probability of migration at a given age for individuals who eventually migrate. Because of the right truncation, from now on c_i will be interpreted as the unobserved propensity of an individual to migrate early versus late: a higher value of c_i implies a higher probability of early migration.

For convenience and clarity, let us reproduce two equations for wages and the timing of migration below:

$$W_{is} = a_0 + (\delta + \boldsymbol{b_i})(Age_{is} - T_i) + \beta X_{is} + \boldsymbol{a_i} + \varepsilon_{is} \qquad (7')$$

$$\lambda_i(t) = \lambda_0(t)\, exp\{c_0 + \gamma Z_i(t) + \boldsymbol{c_i}\}, \tag{11'}$$

where T_i is age-at-migration, t refers to the time/age before migration, while s denotes the time after migration. These types of models do not require exclusion restrictions for identification (see Lillard, 1999). In principle, past realizations of $Z_i(t)$ could work as exclusion restrictions (if those were necessary) under the assumption that $Z_i(t)$ before migration affects future wages through the timing of migration T_i and/or potentially through future realizations (if $Z_i(s) \in X_{is}$), but not through ε_{is}.

In the joint estimation, we use the Gompertz specification of the baseline hazard $\lambda_0(t)$, as it fits the data slightly better than the Weibull distribution. The reason for choosing a parametric specification is the computational complexity of the semi-parametric Cox model with frailty.[6] Parametrically specified probability models are often better suited for policy simulations and predictions of survival time.

The unobserved propensity of early-age migration c_i, immigrant quality a_i and the individual deviation from the average assimilation rate b_i are all correlated with each other and assumed to have a joint trivariate normal distribution with covariance matrix $\sum_{abc}$. We do not have any prior expectations on the sign and strength of correlation. Immigrants with transferable skills may have a smaller room for growth due to higher starting earnings ($\rho_{ab} < 0$). They may also have lower incentives and a higher opportunity cost in terms of foregone earnings to invest in human capital after migration ($\rho_{ab} < 0$). At the same time, immigrants with better linguistic abilities and other transferable skills could be more efficient in acquiring new skills in the host country and have a steeper wage trajectory ($\rho_{ab} > 0$). Likewise, immigrants who arrive at a young age may start their work career in a new country at lower earnings but experience faster earnings growth ($\rho_{bc} > 0$). Risk averse individuals may postpone their migration decision and be less prone to take risky but more rewarding employment opportunities in the host country ($\rho_{ac} > 0$). Unobserved productivity could be higher among late arrivals if they have acquired portable human capital at home ($\rho_{ac} < 0$). Practically, there is a story behind almost any type of relations, which we can test. Since random effects are likely to be correlated, there is a clear benefit of estimating these two equations jointly.

The joint likelihood function of wages and timing of migration is given by the following expression:

---

[6] In addition, the advantage of the Cox model no longer holds in joint modelling and a completely unspecified baseline hazard will generally lead to underestimation of standard errors (Rizopoulos, 2012; Yuen and Mackinnon, 2016).

$$L(\theta) = \prod_{i=1}^{n} \iiint_{-\infty}^{+\infty} \left\{ \prod_{s=1}^{S} f(W_{is}|X_{is}, T_i, a_i, b_i; \theta_w) \right\} \times f(T_i|Z_{it}, c_i; \theta_t) f(a_i, b_i, c_i; \theta_{abc}) \mathrm{d}a_i \mathrm{d}b_i \mathrm{d}c_i \tag{12}$$

where

$$f(W_{is}|X_{is}, T_i, a_i, b_i; \theta_w) = (2\pi\sigma_\epsilon^2)^{-1/2} \times exp\left\{ \frac{-(W_{is} - a_0 - (\delta + b_i)YSM_{is} - \beta X_{is} - a_i)^2}{2\sigma_\epsilon^2} \right\} \tag{13}$$

$$f(T_i|Z_{it}, c_i; \theta_t) = [\lambda_0(T_i)\, exp(c_0 + \gamma Z_{it} + c_i)] \times exp\left\{ -\int_{T_0}^{T_i} \lambda_0(u)\, exp(c_0 + \gamma Z_{it} + c_i) du \right\} \tag{14}$$

$$f(a_i, b_i, c_i; \theta_{abc}) = ((2\pi)^3 |\textstyle\sum_{abc}|)^{-1/2} exp\left\{ -\frac{1}{2} (a_i b_i c_i)' \textstyle\sum_{abc}{}^{-1} \begin{pmatrix} a_i \\ b_i \\ c_i \end{pmatrix} \right\} \tag{15}$$

$f(W_{is}|X_{is}, T_i, a_i, b_i; \theta_w)$ is the probability density function of wages in the host country conditional on age-at-migration T_i and random effects a_i and b_i. Since YSM_{is} is a linear function of T_i ($YSM_{is} = Age_{is} - T_i$), wages in the host country depend on age-at-migration.

$f(T_i|Z_{it}, c_i; \theta_t)$ is the likelihood of hazard model, where T_i is the failure time. The second expression in Equation (14) $exp\left\{ -\int_{T_0}^{T_i} \lambda_0(u)\, exp(c_0 + \gamma Z_{it} + c_i) du \right\}$ is the survival function from age $T_0 = 15$ till age-at-migration, and the first expression in square brackets $\lambda_0(T_i)\, exp(c_0 + \gamma Z_{it} + c_i)$ is the hazard function at failure time T_i.

$f(a_i, b_i, c_i; \theta_{abc})$ is multivariate normal density for correlated random effects, with $\sum_{abc}$ being the variance-covariance matrix of a_i, b_i, c_i.

$\theta_w, \theta_t, \theta_{abc}$ denote parameters for the random effects in wage equations, timing of migration equation, and the covariance matrix, respectively. They are estimated by maximizing the joint likelihood given by Equation (12). Based on these parameters, we calculate the best linear unbiased prediction (BLUP) of a_i, b_i, c_i at the individual level and recover their complete distributions.

We program our own maximum likelihood routine because the existing software packages written mostly by biostatisticians are not suitable for our research question. In biostatistics, the joint models of longitudinal and time-to-event data typically estimate the biomarker's longitudinal evolution that is terminated by death, the onset of disease or other event. Since the event probability is related to individual-specific parameters of the longitudinal process, joint estimation is required. These joint models either include random effects from the longitudinal equation directly into the survival equation through some association structure (Crowther *et al.*, 2012) or, similar to our approach, allow for longitudinal and survival random effects to be jointly distributed as multivariate normal (Pantazis and Touloumi, 2005). Unfortunately, written programs like JMRE1 in Stata assume only an accelerated failure-time lognormal form and only time-constant covariates for the survival sub-model. More importantly, existing programs do not allow for the failure event to occur prior to or during the longitudinal process, like in our case when the event (i.e., migration) happens before the longitudinal wage assimilation process. The maximum likelihood routine that we develop in Matlab is described in Technical Appendix A1.

As a side note, we found only two published economics papers, in which individual-specific wage trajectories (or random slopes) are correlated with both individual-specific intercepts in the wage equation and constant individual heterogeneity in the proportional hazard duration model (Lillard, 1999; Dostie, 2005). These studies estimate two simultaneously-occurring processes of wage growth and job duration by assuming jointly normally distributed random effects.

These types of models are not without limitations. The reliance on the joint normality of random effects is one of such limitations. Another limitation is the computational complexity of joint modelling. With the size of our dataset (about 100,000 observations), it may take one to two weeks for the full maximum likelihood estimation to converge. In addition, these models assume that schooling and other covariates are not correlated with individual-specific random effects (and with the error), and thus they treat these covariates as exogenous. Finally, while the JHL model is intended to address the selectivity issue with regard to the timing of entry, it has limited capabilities for dealing with other selection issues, which we discuss next.

2.4. Other Selection Issues

The wage equation and subsequently the joint likelihood function can only be estimated for individuals who are employed. This may raise valid concerns over the potential selection bias. Given the complexity of the joint model, we do not have an adequate way of integrating a Heckman-style selection equation into the likelihood function. Instead, we use an alternative estimation procedure, which is based on the inverse propensity weighting (IPW) (Rosenbaum and Rubin, 1983; Hirano *et al*., 2003; Wooldridge, 2007). The procedure involves two steps. In the first step, we estimate a selection model by maximum likelihood and obtain the propensity score of observing a positive wage. In the second step, we apply the weighted least square method to Equation (7), where the weights are given by the inverse propensity score estimated in the first step.

Suppressing subscripts i and s for the compactness of notation, let D_1 be a selection dummy variable, taking the value of one if an immigrant is employed and the value of zero if otherwise. The selection dummy is assumed to be linked through the indicator function $D_1 = I(D_1^* > 0)$ with the following latent index model:

$$D_1^* = \Omega_1\pi_1 + \upsilon_1, \upsilon_1 \sim N(0,1), \tag{16}$$

where Ω_1 is a vector of all explanatory variables in Equation (7), plus the state-level unemployment rate among natives. The latter variable serves as an exclusion restriction that is assumed to influence immigrants' wages only through changes in the employment probabilities of immigrants, but not directly. The propensity score, s_1, is the same as the probability of employment selection given the observed covariates Ω_1:

$$s_1 = Pr(D_1 = 1 \mid \Omega_1) = Pr(\upsilon_1 > -\Omega_1\pi_1). \tag{17}$$

The weight is the inverse of the estimated propensity score that is predicted from a standard probit model (hence the name of the inverse propensity weight).

The IPW method may accommodate cases of double selection. For example, previous literature highlighted the selectivity issue due to out-migration (e.g., Bellemare, 2007; Lubotsky, 2007; Dustmann and Görlach, 2016). The selective nature of out-migration can lead to biased estimates of earnings progression in the host country. A similar selectivity argument can be made with respect to other reasons for leaving the survey, including the refusal to be re-interviewed or death. We do not distinguish between different reasons for the immigrants' survey attrition, although it could be an interesting study on its own. We create a combined dummy variable, $D_2 = I(D_2^* > 0)$, that takes the value of one if the immigrant stays in the survey in one or more subsequent rounds, and the value of zero if otherwise. Then, provided that an exclusion restriction exists, re-interviewing can be modelled as a latent index function, $D_2^* = \Omega_2\pi_2 + \upsilon_2, \upsilon_2 \sim N(0,1)$, with the corresponding propensity score, $s_2 = Pr(D_2 = 1 \mid \Omega_2) = Pr(\upsilon_2 > -\Omega_2\pi_2)$. If two selection processes are independent, that is $corr(\upsilon_1, \upsilon_2) = 0$, then the two probit models can be estimated separately, in which case, the inverse propensity weight is $IPW = 1/(\hat{s}_1 \cdot \hat{s}_2)$. If $corr(\upsilon_1, \upsilon_2) \neq 0$, the inverse propensity weight can be calculated as the inverse of the predicted probability obtained from the joint maximum likelihood estimation of two probit equations, with $IPW = 1/\hat{s}$, where $s = Pr(D_1 = 1, D_2 = 1 \mid \Omega_1, \Omega_2)$. Thus, a higher weight is given to the immigrant worker whose observed characteristics are found to be positively associated with attrition and non-employment. For exclusion restrictions in the second selection equation, we use the mode of interview in survey year s (such as face-to-face, self-written and mailed, and computer assisted), a dummy for the first-time interview, and an indicator for whether the same interviewer surveyed the household in year s as in year s–1. We find interview characteristics to be good predictors of continued survey participation. We also use the average economic growth in the home country in the next three years as a potential factor influencing the decision of immigrants to stay in the host country.

3. Data

3.1. Data Source and Sample

The primary data source for our analysis is the German Socio-Economic Panel (GSOEP). The GSOEP is the longest-running panel of private households and persons in Europe. The survey began in West Germany in 1984. After the fall of the Berlin wall in June 1990, residents of the German Democratic Republic or East Germany were also included in the target population. Over time, several additional samples were drawn to replenish the original sample and to include special sub-populations such as immigrants and high-income households. As of 2014, GSOEP consists of 31 waves and 15 distinct samples.

The GSOEP is widely used in migration research, as it is one of a few national longitudinal surveys with a large representation of immigrants. The immigrant status is defined based on the country of birth. One complication that requires special consideration is that immigrants have different sampling probabilities depending on the sample the immigrants are part of. In several samples (such as the initial sample, the supplementary sample of East Germans, replenishment samples, and the focused sample of high-income earners), the share of immigrants is relatively low and varies from 1 to 12 percent. However, there are three samples that specifically target foreign workers and immigrants: sample B "Foreigners in West Germany", sample D "Immigrants", and sample M "Migration".[7] Over 75 percent of our observations used in the estimation comes

[7] Sample B, which was started in 1984, includes 1393 households where the head of household is from one

from these three samples. Since surveyed immigrants are not randomly drawn from the German immigrant population, their composition does not perfectly match the national composition of immigrants by country of origin. For example, the survey oversamples immigrants from the countries that signed guest-worker agreements and also from Poland and the former Soviet Union. To deal with sampling issues, we use probability sampling weights described in Data Appendix A2.

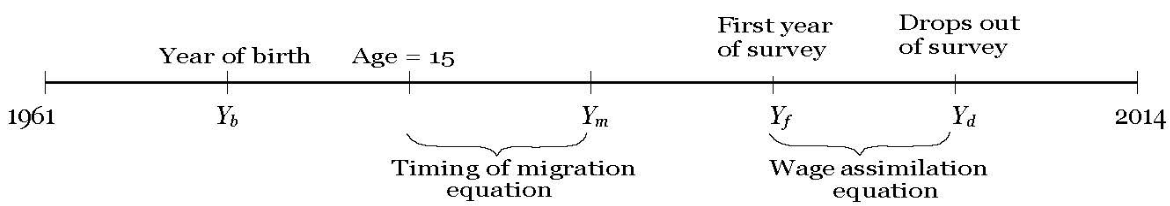
The immigrant status is defined based on the country of birth outside either East or West Germany. Our estimation sample consists of 8,288 immigrants between the ages of 15 and 65 at the time of survey and who arrived to Germany after 1960 at age 15 or older.[8] Child immigrants are excluded from the analysis because they are not likely to make individual migration decisions, and also because previous studies show that the post-migration assimilation experience is very different for child immigrants than for youth and adult immigrants (e.g., Bleakley and Chin, 2004).

We also drop less than 10 percent of observations with missing values on migration status, country of origin, the year of migration, work experience, and year of schooling and training. Given a very small percent of missing values, we assume that dropped observations are ignorable or missing completely at random.

3.2. Data Structure

The data structure is visualized in Figure 1. To estimate the JHL model, original data is reshaped into a "person-age" format, with each spell being a one year of an immigrant's life.

Figure 1: Data Structure



of the five guest-worker countries, specifically, Turkey, Greece, Ex-Yugoslavia, Spain, or Italy. Sample D "Immigrants" started in 1994/95 with 522 households, which consisted primarily of ethnic German immigrants from the former Soviet Union and Eastern Europe as well as asylum seekers mainly from the parts of Yugoslavia devastated by the war. Finally, sample M "Migration" started in 2013 with 2,723 households. It is designed to account for changes in the composition of migration to Germany since 1995.

[8] We are using the 95 percent of the original data since researchers outside the European Union are not allowed the access to the entire dataset.

There are two major phases of the life course of each surveyed immigrant: before and after the year of migration Y_m. The "timing of migration" hazard model is estimated from the age of 15 until the year of migration. Earnings profiles are estimated from the first year of survey participation Y_f until the last year Y_d, which is either year 2014 or the year before the immigrant drops out of the survey. Less than 1 percent of immigrants are surveyed in the year of their arrival ($Y_m = Y_f$). For most of the estimation sample, the period for which the hazard model is estimated does not overlap with the period for which the wage equation is estimated, as shown in Figure 1. Altogether, we work with four estimation samples: (A) pre-migration full sample of person-age observations from age 15 to the year of arrival (N=108,779), (B) pre-migration "joint model" sample with at least one post-migration observation on wage (N=70,679), (C) post-migration survey sample which includes non-employed immigrants (N=46,713), and (D) post-migration sample with non-missing wage data (N=28,756).

3.3. Variables

Below we briefly describe variables used in this paper. Further details on the construction of variables and data sources are provided in Appendix A3.

Dependent Variable. The log of hourly wage rate is our main outcome variable. The numerator of hourly wage is net income earned from employment last month in constant 2010 prices. Net income means the amount after deduction of taxes, social security, and unemployment and health insurance. The denominator of hourly wage is actual working hours per week times (30/7) weeks in a month. We exclude imputed values of earnings due to potential match bias from earnings imputation (Bollinger and Hirsch, 2006). Instead, we use the inverse propensity weighting to account for missing values in earnings.

In the GSOEP, we observe an increase in real hourly wage with years since migration. However, as depicted in Figure 2, the starting wage and its subsequent growth vary substantially depending on the age at arrival. Immigrants who arrive in Germany early between the age of 15 and 25 start their work career at a lower wage rate, but gain a 57 percent wage growth in the first ten years of living in Germany and an additional increase of 13 percent over the next ten years. Compared to the early arrivals, those who immigrate to Germany between the age of 26 and 35 enjoy a 30 percent higher initial wage, but show a less remarkable rate of subsequent change: an 18 percent wage increase in the first ten years of stay in the host country and no progress afterwards. Finally, the late arrivals at age 36 to 45 years old have a nearly flat wage profile. The right panel of

Figure 2 depicts similar wage trajectories after accounting for constant, both observed and unobserved, individual characteristics. Thus, the empirical relationship between the age-at migration and the slope of wage trajectories appears to be negative.

Individual Covariates. The timing of migration equation and wage equation share several individual-level covariates. Some of these covariates are time-constant such as gender, German ethnicity, place of upbringing in childhood (large or medium-sized city, small city, and rural area), and the level of schooling completed by a parent (basic secondary or lower vocational, general secondary or upper vocational, and higher education). Other shared covariates vary with age and time – namely, years of formal schooling, years of job training, and years of work experience. Using life history calendars between the ages of 15 and 65, we construct time-varying years of schooling, training, and experience at each age both before and after migration (see Appendix Table A3 for further details). Since our experience variable captures actual rather than age-collinear potential years of work experience, we can include age as an additional regressor in the wage equation. Both age and experience enter the wage equation in a quadratic form, following the traditional Mincerian approach. Recall that age enters the baseline hazard in the timing of migration equation by construction.

Apart from the common vector, the wage equation includes years since migration, which are calculated as the difference between age in a survey year a_s and age-at-migration a_m, where $a_m = Y_m - Y_b$ in Figure 1. Another variable that is unique to the wage equation but not included in the hazard model is the type of current residence in a survey year (urban West Germany, rural West Germany, urban East Germany, and rural East Germany).[9] All individual-level covariates in the hazard model enter the wage equation. Pre-migration history of education and employment is assumed to influence the post-migration level of wages through the timing of migration and through the human capital accumulation.

Macro Covariates. Immigrants in our estimation sample come from 122 countries. Having 1,428 unique combinations of the country of origin and year of migration opens an opportunity to exploit country-level sources of variation by linking a variety of external data sources.[10]

[9] By excluding this variable from the hazard model, we assume that immigrants do not foresee perfectly the type of future residence in the host country before migration. If we include the type of future residence into the hazard model assuming a forward-looking agent with perfect foresight, our estimation results would hardly change.

[10] Although it may seem that we have an almost infinite number of possibilities for including external variables, the computational intensity of the JHL model puts constraints on our choice and experimentation

All macro covariates can be grouped into three categories: pull-push factors, proxies for the cost of migration, and controls for time effects. Pull factors attract immigrants to host countries, while push factors force individuals to leave their home country. We use the growth of real PPP-adjusted GDP per capita in Germany as a pull factor that captures better economic prospects in the host country. This variable does not vary across home countries and thus can only be used in conjunction with some sort of a polynomial function of time trend, but not when year fixed effects are included.[11] We approximate push factors with the log of real PPP-adjusted GDP per capita and the level of political instability in the home country.[12]

The costs of migration can be captured by the geographic distance, linguistic distance, and political agreements that ease migration. The geographic distance measures the monetary cost of moving. It also represents the effort cost of collecting information about the host country, which is likely to be higher for prospective immigrants in geographically distant countries. We control for the geographic distance by using fixed effects for the world region of home country and the log distance between capitals in km. The linguistic distance between immigrant's native language and German language reflects the costs of acquiring new language. Immigrants who grew up speaking the language that is distant from German face higher cost in both searching information about the host country and transferring their preexisting skills to the new labor market. We use the Levenshtein linguistic distance, which measures the similarity in 40 basic words between languages (Wichmann *et al.*, 2016); see Appendix Table A3 for further details. Inter-country political agreements may also affect the cost of migration. Specifically for Germany, guest-worker treaties with Turkey, Spain, Italy, Ex-Yugoslavia and several other countries in the 1960-1970s encouraged low-skilled migration from these countries. We include an indicator variable for whether a country signed a guest-worker treaty with Germany in a given year.[13]

Finally, to account for the time effects in both migration and wage equations, we

---

with variables. Each additional variable requires several more hours of the computation time. Thus, we have to be reasonable with the number of covariates.

[11] We tried using the level of GDP per capita in Germany, but had to drop it, as this variable was nearly perfectly collinear with a quadratic time trend.

[12] The Center for Systemic Peace (2015) provides a quantitative assessment of major episodes of international, civil, and ethnic violence and warfare for almost 180 countries worldwide between 1946 and 2014. Based on the total score of political instability, we classify all country-year observations into four categories: no episodes of political violence, limited political violence, serious political violence, and warfare.

[13] Political agreements within Europe are largely captured by world region fixed effects, which include separate parts of Europe such as Western Europe, Eastern Europe, and ex-USSR.

include a quadratic function of the linear time trend and a dummy for the post-unification period after 1990 to distinguish between the two significant periods of German history. In the results section, we will show the sensitivity of estimates to having year fixed effects instead of time trend.

All the above macro covariates enter the migration hazard model. A more debatable question is whether they belong in the wage equation. Host-country GDP growth and time effects certainly do. The former could also be a good way to account for the business cycle effects on real wages. The fact that higher entry costs raise the reservation wage of immigrants suggests that proxies for the cost of migration should be part of the wage equation too. The only exception are the guest-worker programs that were halted in 1973 before the GSOEP has started. Furthermore, the linguistic distance between home- and host-country languages, aside from being a proxy for the cost of migration, also captures skill transferability which directly influences the wage trajectory (Chiswick and Miller, 2012).

The level of GDP per capita and political instability in the home country prior to migration may be reasonably assumed to affect the timing of migration without directly affecting the labor market performance of immigrants in the host country. However, one can argue that immigrants from developed countries with a high level of GDP per capita are likely to have more marketable skills in Germany due to similar levels of schooling quality, qualification requirements, etc. To address these types of concerns, we control for per capita GDP and political instability in the home country *at the time of arrival* in the wage equation, while including past realizations of push factors until the arrival into the timing of migration equation.

To recap the selection of variables, Table 1 classifies them by type and equation. In addition, Table 2 reports the mean and standard deviation of variables for each of the four samples used in the paper.

4. Model Estimates

Given that the full maximum likelihood estimation of the JHL model requires considerable computation time, we first present a sensitivity analysis separately for the hazard model and wage assimilation model. Once we are confident in the final specification, we then proceed to the estimation of the joint model.

4.1. Timing of Migration Model

In survival analysis, the focus is on the age profile of migration events. By looking at the distribution of age-at-migration (Figure 3A below) and the Kaplan-Meier survival curves (Figure 3B), we observe that more individuals immigrate in their 20s than in any other age, and that the rate of immigration diminishes with age. The Kaplan-Meier survival curves show no statistically significant gender difference in the hazard of migration with age (Log Rank, p = 0.904; Wilcoxon, p = 0.180). Based on these findings, we have decided to report subsequent results for the combined sample of males and females.[14]

Table 3 presents the reduced-form estimates of Equation (11). The first three columns show estimates of Cox, Weibull and Gompertz hazard models without individual unobserved heterogeneity (or frailty). The differences between the models come from different distributional assumptions of the baseline hazard.[15] The last two columns report estimates of the shared-frailty hazard models (both Weibull and Gompertz specifications) with individual-specific random effects, $\eta_i = \exp(c_i)$, as in Equation (11), where η_i is assumed to have an inverse Gaussian distribution.[16]

Estimates in Table 3 are largely consistent with theoretical predictions outlined in Section 2. The theoretical model predicts that individuals with higher skill transferability would immigrate at an early age. We find that ethnic German resettlers, who are likely to be familiar with the culture and language in Germany, have a higher hazard of early migration. Likewise, individuals from countries that are linguistically similar or spatially close to Germany tend to migrate at a younger age. Immigrants from these countries not only face a lower cost of migration, but also have more transferable skills, as they are more likely to be familiar with host country culture and customs.

Push-pull factors of migration also appear to behave in a predictable way. The inter-country guest-worker treaties that ease labor migration are found to draw more immigrants of a younger age. Higher growth of GDP per capita in the host country attract immigrants, while lower levels of GDP per capita in the home country push people to leave their home country early in a working career. Predictably, the hazard of early migration is considerably lower for immigrants from stable countries with no episodes of political violence. However, no statistically significant differences are found in the risk of early

[14] Separate estimates for each gender can be provided if requested.

[15] The shape of the baseline hazard follows either Weibull distribution, $\lambda_0(t) = \varphi t^{\varphi-1}$, or Gompertz distribution, $\lambda_0(t) = exp(\varphi t)$. The shape is not defined in the Cox proportional hazard model. In the Cox model, the standard test of the proportional-hazards assumption on the basis of Schoenfeld residuals rejected the null of proportionality of hazards.

[16] An alternative assumption of gamma-distributed shared frailties is also tested. Due to the similarity of obtained results to the ones reported in the last two columns of Table 3, the estimates with gamma-distributed frailties are not shown to preserve space.

migration across three categories of political violence (limited, serious, and warfare). The estimates suggest that even the moderate levels of civil unrest and political instability push people to migrate at a younger age.

We do not find statistically significant gender differences in the timing of migration in four out of five specifications. The exception is the Gompertz frailty model, where females are found to immigrate at a later age compared to male immigrants. Most of the measures of pre-migration human capital indicate that previous investment in human capital postpones the migration decision. The hazard of early migration decreases with more years of formal schooling, longer work experience in the home country, and more parental investment (measured via the highest level of education completed by a parent). At the same time, the amount of job training in the home country does not appear to be a relevant factor in the decisions of when to migrate, apart from a statistically significant negative effect in the Weibull frailty model. Finally, no significant variation in the hazard of early migration is found across the location types in childhood (large city/small city/village).

Overall, estimates in Table 3 do not seem to be very sensitive to different distributional assumptions. The shared-frailty hazard models with individual unobserved heterogeneity generally show stronger (in absolute terms) effects of pre-migration human capital on the timing of migration. The estimated coefficients on measures of the cost of migration are also larger in magnitude in shared-frailty models than in non-frailty models. A standard likelihood ratio test for the presence of heterogeneity indicates significant heterogeneity, which likely arises from some unobserved individual-specific effect ($\widehat{Var}(\eta_i) = 0.285, \chi^2 = 111.6$ in Gompertz specification).

In Table 3, we restrict the sample size to be the same as in the joint model by selecting immigrants with at least one observation of wages in the post-migration period. To check how this restriction impacts our estimates, we have re-estimated the same five specifications of Equation (11) on a larger sample irrespective of whether wage is observed in post-migration years. The results of this exercise are reported in Appendix Table A3-1. They are largely consistent with results from the restricted sample, with a few notable exceptions. The gender gap in the timing of migration is more pronounced in the full sample, as non-working females tend to postpone their migration decision. When the non-working immigrant population is included, the effect of pre-migration accumulation of formal schooling and work experience on the probability of early migration becomes either zero or slightly negative. At the same time, the effect of pre-migration years of training changes from zero to positive in four out of five specifications. We postpone the interpretation of the last result until our discussion of the JHL model estimation where we also find a positive effect of training on early-age migration. With regard to other variables, the estimates are essentially unaffected by including immigrants who never worked in Germany.

4.2. Selection Model

Before presenting the wage equation, we first show the selection model estimates that go into the calculation of inverse propensity weights, as described in Section 2.4. We estimate two selection equations, one for non-missing wages and another one for survey re-interviewing. The dependent variable in the first equation takes the value of one if a positive wage is observed, and the value of zero if otherwise. In the post-migration survey sample, wages are not observed for 39 percent of immigrants between the ages of 15 and 65, of which 35 percentage points are due to non-employment and only 3 percentage points are due to non-reporting of wages by those who are employed. Because of a much smaller share of the latter group, we call this equation in a traditional way as employment participation equation. Apart from the wage equation covariates discussed above, the employment participation equation includes the unemployment rate among natives calculated at the state-year level.

The dependent variable in the second selection equation takes the value of one if the immigrant stays in the survey in one or more subsequent rounds, and the value of zero if otherwise. Observations in the last survey round in year 2014 are all coded as zero. The average survey attrition rate among immigrants before 2014 is about 10 percent per year. In addition to wage equation covariates, the survey re-interviewing equation includes the average economic growth in the home country in the next 3 years as a proxy for the attractiveness of return migration. A study by Kroh et al. (2015) finds interview characteristics to be a good predictor of survey attrition in GSOEP. Following this study, we use the mode of interview (such as face-to-face, self-written and mailed, and computer assisted), a dummy for the first-time interview, and a dummy for having the same interviewer in two consecutive survey rounds as additional exclusion restrictions. Summary statistics for variables from two selection equations can be found in Appendix Table A4-2.

The two probit equations are estimated jointly by maximum likelihood. Table 4 reports the marginal effects evaluated at the mean of all variables. The Wald likelihood-ratio test of independent equations rejects the null of $corr(v_1, v_2) = 0$ at the 0.001 percent level of significance, suggesting that the joint estimation of two probit equations is preferred to single-equation estimation. The estimated positive correlation between residuals in two equations indicates that immigrants with a lower unobserved propensity to be employed also tend to leave the survey in the next period.

The estimates of the employment participation equation are mostly in line with our priors. Higher levels of unemployment among natives significantly reduce employment outcomes for immigrants. We also find that the probability of employment increases with work experience, years of formal schooling and training, and the length of stay in the host country. The employment probability is expectedly higher for males by 8 percentage points, ethnic German resettlers, and immigrants from Europe and other developed countries with higher levels of GDP per capita. Yet, once the world regions are controlled for, immigrants who came from geographically distant countries have a higher probability of employment. This result is interesting, as it may suggest that immigrants facing a higher cost of migration are more likely to invest in job search. Among other results, the urban-rural gap in the employment participation of immigrants appears to be much larger in East Germany than in West Germany. No statistically significant employment effect is found with respect to linguistic distance and GDP growth in Germany. The estimated negative coefficient on higher levels of parents' education was counter-intuitive at first. Yet, further examination reveals that this result is driven largely by a higher wage non-response among immigrant workers with better-educated parents. We also find that immigrants who fled the warfare zones are more likely to work but less likely to report their wages compared to immigrants from politically stable countries. This finding could be attributed to higher dependency of refugees on informal employment (Tumen, 2016).[17]

In the survey re-interviewing probit, many covariates appear to be statistically insignificant determinants of survey participation. The probability of remaining in the study (and in the host country) in the future period increases with years since migration, work experience, and linguistic proximity; but it decreases with better prospects in the home country. The likelihood of successful follow-up is larger for females, ethnic Germans, and immigrants from more developed countries. Compared to traditional in-person interviews, surveys by mail reduce the likelihood of future survey participation. Immigrants are also less likely to participate after their first interview and after a change of interviewer. All these findings are quite intuitive and do not require much discussion.

[17] The positive employment effect is not found for immigrants from countries with serious political violence. As a result, the negative wage reporting effect predetermined the negative sign of the estimated coefficient for this group of immigrants.

We use the bivariate probit estimates from Table 4 to obtain the predicted joint probability of working and future survey participation. The inverse of this probability times the sampling weight introduced in Appendix A2 is our final weight.

4.3. Wage Assimilation Model

In Table 5, we provide reduced-form estimates of the linear mixed-effects model of wage assimilation, which corresponds to Equation (7). We also show an OLS specification of wage equation in column 1 as a reference point for the reader. Column 3 reports our preferred specification, as it uses the inverse propensity weight to account for the potential non-random selection into the wage sample.

Our main parameter of interest is the average wage return to an additional year of stay in the host country, δ in Equation (7). It is estimated to be about 1.06 percent. This estimate falls within the range of the estimates that have been reported for various countries. Here is just a small snippet of the previously estimated returns to an additional year since migration:[18]

- 1.17-1.69 for males and 0.04-1.63 for females depending on the level of education (Bratsberg *et al.*, 2006, the U.S. Current Population Survey 1979-2003, YSM=10);
- 1.20 for males and 0.94 for females (Basilio *et al.*, 2009, GSOEP 1984-2005, YSM=10);
- 1.18 for males from specification with the highest *R* squared (Borjas, 1995, U.S. Census of 1970, 1980, and 1990, YSM=10);
- 1.85 for males with non-U.S. schooling and 0.71 for males with U.S. schooling (Bratsberg and Ragan, 2002, (U.S. Census of 1970, 1980, and 1990, YSM=10);
- 1.14 for white males (Chiswick, 1978, U.S. Census of 1970, YSM=10);
- 1.21 for male immigrants from non-English speaking country with the highest linguistic distance from English; the returns go towards zero with smaller linguistic distance (Chiswick and Miller, 2012, U.S. Census of 2000);
- 0.81 for males (Friedberg, 2000, Israel Census of 1972 and 1984);
- 1.66 for all (Sanroma, 2015, The National Immigrant Survey of Spain 2006-2007).

Therefore, our reduced-form estimate of the average rate of assimilation is comparable to the estimates found by other studies. The mixed-effects model has an additional benefit of estimating the entire distribution of individual-specific returns to the years-since-migration variable, $(\delta + b_i)$ in Equation (7). The standard deviation of $\hat{b}_i$

[18] In cases when researchers used a polynomial function with respect to years since migration, we recalculated the one-year return to YSM at the tenth year of the immigrant's stay in the host country (YSM=10) for comparability purposes.

in our preferred specification is 0.028. In 95 percent of our estimation sample, the assimilation slope lies between −4.5 percent and +6.6 percent. Negative assimilation is not very common; about 6 percent of surveyed immigrants experiences a wage loss by more than 1 percent per year after migration. Another important finding from the mixed-effects model estimation is a very high negative correlation between $\hat{a}_i$ and $\hat{b}_i$, which is estimated to be −0.826. This result implies a lower assimilation rate among immigrants with higher unobserved skills. We will get back to discussing this result after estimating the joint model.

Among other results from the wage model, we find that female immigrants in Germany earn less than male immigrants do. The gender wage gap is estimated to be about 0.25-0.28 log points or 28-32 percent, *ceteris paribus*. Immigrants enjoy relatively small returns to education: 2.6 percent of a wage increase for one year of formal schooling and 0.7 percent increase for a year of job training. But having a parent with a college degree helps an immigrant to earn 12 percent more in wages compared to an immigrant with less educated parents. Estimates also suggest a concave age-wage profile reaching its peak at age 37. Controlling for the wage trajectory over age, an average immigrant receives an additional 1 percent premium for each year of actual work experience.

Most measures of skill transferability tend to be positively associated with the level of wages. Immigrants from Europe, more developed countries, and countries whose primary language is linguistically close to German earn higher hourly wage, on average. There is also a 2.9 percent wage premium for ethnic German resettlers. No statistically significant wage returns are found with respect to the place of upbringing, political violence in home country, GDP growth in the host country, and geographic distance between countries in mixed-effects specifications. At the same time, wage differences are substantial across locations in Germany. Immigrants in East Germany receive considerably less than their counterparts in West Germany. The regional gap is 26 percent (0.23 log points) in urban areas and 37 percent (0.32 log points) in rural areas (in column 3).

In column 4 of Table 5, we estimate a mixed-effects model with year fixed effects. Our estimates barely change from column 3 to column 4. The quadratic function of time trend serves as a good representation of time effects. Based on this result, we opted for the parametric function of time in the joint maximum likelihood estimation to reduce the number of estimated parameters.

4.4. Joint Model Estimates

Covariance Structure of Individual Heterogeneity Components. Table 6 reports maximum likelihood estimates of the JHL model given by Equations (7') and (11'). We begin with interpreting the variance-covariance structure of individual-level heterogeneity components. The structure of predicted random effects is reported at the bottom of Table 6. The key result is a positive and statistically significant correlation between predicted individual earnings progression $\hat{b}_i$ and predicted unobserved propensity to migrate early $\hat{c}_i$ ($\hat{\rho}_{bc} = 0.204$). This positive correlation is consistent with our earlier conjecture made in Section 2.2 that individuals with a higher unobserved propensity for early migration may have a lower cost of acquiring host-country specific human capital and, as a result, assimilate faster in a new labor market. However, the reverse association could also be true: individuals who expect high wage growth in the host country choose to migrate earlier. Although we cannot separately identify the relative importance of the two explanations, the timing of migration and wage assimilation are positively related, hence indicating an upward slope-heterogeneity bias in the OLS-estimated returns to years since migration according to Equation (10).

If the positive sign of slope-heterogeneity bias is expected, the sign of ability bias arising from the correlation between a_i and c_i is less clear *a priori*; see discussion in Section 2.2. Table 6 shows that the estimated correlation between unobserved skills and the predicted unobserved propensity for early migration is weakly negative but statistically significant ($\hat{\rho}_{ac} = -0.067$). The negative sign implies that high-ability immigrants tend to postpone their migration decision, or conversely, those who delay their entry tend to have higher unobserved portable skills.

Because of a smaller magnitude of downward ability bias, total bias in the OLS-estimated average rate of assimilation is largely influenced by significant upward slope-heterogeneity bias. In concordance with obtained correlations, the JHL model estimate of the average rate of assimilation is 33 percent lower than the reduced-form estimate. If the linear mixed-effects wage model predicts the average rate of assimilation to be 1.06 percent of earnings growth for each additional year of stay in the host country, the joint model predicts it to be only 0.71 percent. Thereby, the failure to account for the selective timing of migration overestimates the average rate of assimilation.

The last correlation coefficient shown in Table 6 is between unobserved skills and the slope of earnings progression. The estimated correlation between $\hat{a}_i$ and $\hat{b}_i$ is very large ($\hat{\rho}_{ab} = -0.828$), and it is practically identical to a mixed-effects wage model estimate. This negative correlation implies that immigrants with higher unobserved skills have a smaller rate of wage assimilation, holding observed skills (such as formal schooling, training, and work experience) constant. This finding supports the presence of conditional convergence between low- and high-quality immigrants. Some parallels can be drawn here with earlier discussion in the migration literature on the inverse relationship between initial earnings and assimilation rate. Initial earnings are often used as a proxy for the unobserved quality of arrival cohorts (Borjas, 1987) or for transferable skills (Duleep and Regets, 1999). Borjas's (1999) theoretical framework derives two conditions for the negative unconditional correlation between entry wages and the rate of wage growth: (1) substitutability between pre- and post-migration human capital when skilled immigrants with transferable human capital invest less, earn more at the time of arrival, and experience slower wage growth; and (2) strong relative complementarity in human capital when skilled immigrants invest a lot at the time of entry, so that they earn less initially but have a higher post-investment wage growth. Furthermore, Borjas (1999) shows that, holding initial skills constant, the conditional correlation between entry wages and the rate of wage growth is always negative. Likewise, Duleep and Regets (1999) predict the inverse relationship between immigrant initial wages and wage growth. They argue that immigrants with higher skill transferability have lower investment in human capital and hence lower wage growth. Thus, the negative correlation between unobserved skills and assimilation slope in our joint model is consistent with existing theoretical models and empirical evidence on lower assimilation rates among immigrants with higher entry wage.

Figure 4 shows that the JHL method reduces the overall dispersion of predicted individual assimilation rates compared to the mixed-effects model estimates. However, substantial individual-level heterogeneity in earnings progression remains. 95 percent of all estimated assimilation slopes fall between −4.5 percent and +5.9 percent. In extreme cases, the rate of wage assimilation goes to as low as −20.6 percent and as high as +15.3 percent increase in hourly wage rate per year of stay in Germany. About 18 percent of the sample has a negative rate of wage assimilation, but only 7 percent experiences a wage loss by more than 1 percent per year after migration. In previous studies, the negative rates of assimilation are found among immigrants with highly transferable skills (Chiswick and Miller, 2012). We find negative rates to be more common not only among immigrants with higher unobserved transferable skills but also among late arrivals.

Effects of Covariates. The estimates of the hazard sub-model in Table 6 are consistent with the comparative statics predictions derived in Section 2.1. Individuals are found to immigrate earlier when they face worse conditions in the home country (political violence '+', GDP per capita '-'), incur lower costs of migration (geographic distance to Germany '–', guest-worker treaties '+'), have higher transferability of preexisting skills (ethnic Germans '+', linguistic distance '–'), and expect better prospects in the host country (economic growth in Germany '–'). At the same time, immigrants who are raised by better-educated parents and accumulated more formal schooling and work experience at home tend to delay their migration decision. According to the timing of migration model presented in Section 2.1, this last result could be attributed to higher earnings at home and thus larger opportunity costs of migration for individuals with more human capital ($\overline{w}_o{}'(K) > 0$) and/or to greater substitutability between pre- and post-migration human capital ($g'(K) < 0$). It is less obvious, however, why the estimated coefficient on job training is positive and statistically significant, indicating that immigrants with vocational training enter Germany at a younger age. Job training is presumably less costly than formal schooling. The positive sign on pre-migration years of job training may be indicative of the potential complementarity between pre- and post-migration training investment if, for example, an easier entry for immigrant workers with certain job qualifications is augmented with post-migration training programs.

To give an idea of the magnitude of the effects, we calculate the change in predicted median age-at-migration ($\hat{t}^{med}$) associated with a one-unit change in covariate $z \in Z$:

$$\frac{\partial \hat{t}^{med}}{\partial z} = -\widehat{\gamma_z} \frac{(ln2)exp(-\bar{Z}\hat{\gamma} - \hat{\sigma}_c^2/2)}{[\hat{\varphi}(ln2)exp(-\bar{Z}\hat{\gamma} - \hat{\sigma}_c^2/2) + 1]}, \tag{18}$$

where $\hat{\sigma}_c^2$ is the estimated variance of c_i.[19]

The median age-at-migration effects shown in Table 6 are evaluated at the mean values of covariates at the time of entry with the parameter estimates plugged in. This exercise highlights the two effects of migration factors on earnings stream in the host country: direct productivity effect and indirect wage effect through the change in the length of stay. For example, linguistic proximity of 0.7 (which is equivalent to the distance between German and English languages) not only earns 21 percent wage premium per additional year of stay, but also increases the duration of stay by almost 2 years, thus increasing total earnings received in the host country. For some migration factors, the two wage effects work in opposite directions. For example, an increase in the level of GDP per

[19] Predicted median age-at-migration is derived from the predicted survival time being set to one half: $\hat{S}(t) = exp\left\{-\frac{1}{\hat{\varphi}}\left(e^{\bar{Z}\hat{\gamma}+\frac{\hat{\sigma}_c^2}{2}}\right)\left(e^{\hat{\varphi}t} - 1\right)\right\} = \frac{1}{2}$. See Liu (2014) for the explanation of why the adjustment $\hat{\sigma}_c^2/2$ is needed when the random effect is normally distributed.

capita in the country of origin by 1 log point (which approximately corresponds to the difference in real GDP per capita between Turkey and Switzerland in 2014) increases wages by 3.8 percent per year of stay, but postpones the median age-at-migration by 3.6 years. Some migration factors do not matter in the wage equation, but they do change the length of stay (for example, political instability or geographic distance).

5. Conclusions

This paper develops and estimates the joint hazard-longitudinal model of the timing of migration and wage assimilation with correlated random intercepts and slopes. The JHL model is designed to account for the selective timing of migration in estimating the returns to years since migration. The main benefit of this model is that it allows for the variation in individual rates of wage assimilation and their correlation with unobserved skills and with the unobserved propensity to migrate early.

So far, the economic literature has not devoted much attention to the link between selective timing of migration and the process of labor market assimilation. We find that the predicted unobserved propensity to migrate early and individual rates of assimilation are positively correlated. This result supports the need for the two processes – survival and longitudinal – to be estimated jointly. We also observe that immigrants with a lower level of unobserved skills tend to have a faster rate of wage assimilation, holding observed skills constant. This finding supports the hypothesis of conditional convergence or catch-up between low-quality and high-quality immigrants.

Our analysis shows both theoretically and empirically that the length of stay is an endogenous variable in the wage equation due to the selective timing of entry, and thus the commonly estimated average rate of labor market assimilation is likely to be biased. In the German data, the estimated average returns to years since migration is about 33 percent lower in the JHL model compared to the OLS estimation. This difference stems from relatively large upward slope-heterogeneity bias, that is not enough compensated by small downward ability bias. We also observe that individual-specific rates of wage assimilation vary a great deal across immigrants. About 18 percent of all immigrants, especially those who arrive late with higher levels of unobserved skills, appear to experience negative wage growth after migration.

Among other results, the JHL model reveals that individuals tend to immigrate earlier when they encounter political violence at home, live in a less developed country, have closer ethnic ties with the host country, acquire short-term vocational skills, and expect higher economic growth in the host country. Immigrants from countries with closer linguistic proximity, nearer geographic location, and guest-worker agreements also tend to arrive at a younger age. At the same time, pre-migration investment in formal schooling and work experience seems to postpone the migration decision.

As to the wage returns, most observed indicators of skill transferability are found to increase post-migration earnings both directly per hour and indirectly through the length of stay. Ethnic German immigrants and immigrants from Europe and countries whose primary language is linguistically close to German enjoy a higher stream of earnings that come from both higher hourly wage rate and longer duration of stay. Similar complementarity between the two effects is observed with respect to job training. Even though some factors such as political instability in the home country or geographic proximity may not seem matter for the wage per hour, they impact the total post-migration earnings by increasing the duration of stay in the host country. This again shows the benefit of the joint model.

Thus, the evidence we obtain highlights the importance of joint modelling, although further applications of this method in other institutional settings will be needed to verify if our conclusions regarding the magnitude of bias are generalizable.

6. References

Antecol, Heather, Peter Kuhn, and Stephen J. Trejo, 2006. Assimilation via Prices or Quantities? Sources of Immigrant Earnings Growth in Australia, Canada, and the United States," *Journal of Human Resources* XLI (4): 821-840.

Basilio, Leilanie, Thomas K. Bauer, and Mathias Sinning, 2009. "Analyzing the Labor Market Activity of Immigrant Families in Germany," *Labour Economics* 16: 510-520.

Bellemare, Charles, 2007. "A Life-Cycle Model of Outmigration and Economic Assimilation of Immigrants in Germany," *European Economic Review* 51: 553-576.

Bijwaard, Govert E., Christian Schluter, and Jackline Wahba, 2014. "The Impact of Labor Market Dynamics on the Return Migration of Immigrants," *Review of Economics and Statistics* 96(3): 483–494.

Bleakley, Hoyt and Aimee Chin, 2004. "Language Skills and Earnings: Evidence from Childhood Immigrants," *Review of Economics and Statistics* 86(2): 481–496.

Bollinger, Christopher R. and Barry T. Hirsch, 2006. "Match Bias from Earnings Imputation in the Current Population Survey: The Case of Imperfect Matching," *Journal of Labor Economics* 24(3): 483-519.

Borjas, George J., 1987. "Self-Selection and the Earnings of Immigrants," *American Economic Review* 77(4): 531-553.

Borjas, George J., 1995. "Assimilation and Changes in Cohort Quality Revisited: What Happened to Immigrant Earnings in the 1980s?" *Journal of Labor Economics* 13(2): 201-245.

Borjas, George J., 1999. "Economic Analysis of Immigration," in *Handbook of Labor Economics,* edited by Orley Ashenfelter and David Card, Elsevier Science, Vol. 3: 1697-1760.

Borjas, George J., 2015. "The Slowdown in the Economic Assimilation of Immigrants: Aging and Cohort Effects Revisited Again," *Journal of Human Capital* 9(4): 483-517.

Bratsberg, Bernt, Erling Barth, and Oddbjørn Raaum, 2006. "Local Unemployment and the Relative Wages of Immigrants: Evidence from the Current Population Surveys," *Review of Economics and Statistics* 88(2): 243-263.

Bratsberg, Bernt and James F. Ragan, 2002. "The Impact of Host-Country Schooling on Earnings – A Study of Male Immigrants in the United States," *Journal of Human Resources* 37(1): 63-105.

Card, David, 2001. "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica* 69(5): 1127-1160.

Center for Systemic Peace, 2015. *Major Episodes of Political Violence, 1946-2014*, http://www.systemicpeace.org/inscrdata.html

Chiswick, Barry R., 1978. "The Effect of Americanization on the Earnings of Foreign-Born Men," *Journal of Political Economy* 86(5): 897-921.

Chiswick, Barry R. and Paul W. Miller, 2012. "Negative and Positive Assimilation, Skill Transferability, and Linguistic Distance," *Journal of Human Capital* 6(1): 35-55.

Cobb-Clark, Deborah, 1993. "Immigrant Selectivity and Wages: The Evidence for Women," *American Economic Review* 83(4): 986-993.

Constant, Amelie and Douglas S. Massey, 2003. "Self-selection, Earnings, and Out-migration: A Longitudinal Study of Immigrants to Germany," *Journal of Population Economics* 16: 631-653.

Crowther, Michael J., Keith R. Abrams, and Paul C. Lambert. 2012. "Flexible Parametric Joint Modelling of Longitudinal and Survival Data," *Statistics in Medicine* 31: 4456–4471.

Dostie, Benoit, 2005. "Job Turnover and the Returns to Seniority," *Journal of Business and Economic Statistics*, 23(2), 192-199.

Duleep, Harriet Orcutt and Regets, Mark C., 1999. "Immigrants and Human-Capital Investment," *American Economic Review* 89(2): 186-191.

Dustmann, Christian and Albrecht Glitz, 2011. "Migration and Education," in *Handbook of the Economics of Education,* edited by Eric A. Hanushek, Stephen Machin and Ludger Woessmann, Elsevier, Vol.4: 327-439.

Dustmann, Christian and Joseph-Simon Görlach, 2015. "Selective Out-Migration and the Estimation of Immigrants' Earnings Profiles," in *Handbook of the Economics of*

International Migration, edited by Barry R. Chiswick and Paul W. Miller, Elsevier, Volume 1A: 489-533.

Dustmann, Christian and Joseph-Simon Görlach, 2016. "The Economics of Temporary Migrations," *Journal of Economic Literature*, 54(1), March 2016: 98-136(39).

Fertig, Michael and Stefanie Schurer, 2007. "Labour Market Outcomes of Immigrants in Germany: The Importance of Heterogeneity and Attrition Bias," *IZA Discussion Paper*, No. 2915.

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

Heckman, James and Burton Singer, 1984. "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica* 52(2): 271-320.

Hirano, Keisuke, Guido Imbens, and Geert Ridder, 2003. "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica* 71, 1161-1189.

Jain, Apoorva and Klara Sabirianova Peter, 2017. "Limits to Wage Growth: Understanding the Wage Divergence between Immigrants and Natives" *IZA Discussion Paper*, No.

Kahanec, Martin and Klaus F. Zimmermann (editors), 2016. *Labor Migration, EU Enlargement, and the Great Recession*, Springer, Berlin, 419-445.

Kerr, Sari Pekkala and William R. Kerr, 2011. "Economic Impacts of Immigration: A Survey," *NBER Working Paper Series*, No. 16736.

Kroh, Martin, Simon Kühne, Rainer Siegers, 2015. "Documentation of Sample Sizes and Panel Attrition in the German Socio-Economic Panel (SOEP) (1984 until 2014)," *SOEP Survey Papers*, No. 297: Series C. Berlin: DIW / SOEP.

LaLonde, Robert J. and Topel, Robert H., 1992. "The Assimilation of Immigrants in the U.S. Labor Market," in *Immigration and the Workforce: Economic Consequences for the United States and Source Areas*, edited by George Borjas and Richard Freeman, University of Chicago Press: 67–92.

Lillard, Lee, 1999. "Job Turnover Heterogeneity and Person-Job-Specific Time-Series Wages," *Annales d'Économie et de Statistique* 55/56: 183-210.

Liu, Xian, 2014. "Survival Models on Unobserved Heterogeneity and their Applications in Analyzing Large-Scale Survey Data," *J Biomet Biostat* 5: 191. doi:10.4172/2155-6180.1000191.

Lubotsky, Darren, 2007. "Chutes or Ladders? A Longitudinal Analysis of Immigrant Earnings," *Journal of Political Economy* 115(5): 820–867.

Pantazis, Nikos and Giota Touloumi. 2005. "Bivariate Modelling of Longitudinal Measurements of Two Human Immunodeficiency Type 1 Disease Progression Markers In the Presence of Informative Drop-Outs," *Journal of the Royal Statistical Society*, Series C 54: 405–423.

Rizopoulos Dimitris, 2012. *Joint Models for Longitudinal and Time-To-Event Data: With Applications in R*. Boca Raton: CRC Press.

Rosenbaum, Paul R. and Donald B. Rubin, 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika* 70(1): 41-55.

Sanroma, Esteban, Raul Ramos, and Hipolito Simon, 2015. "How Relevant is the Origin of Human Capital for Immigrant Wages? Evidence from Spain," *Journal of Applied Economics* 18(1): 149-172

Tumen, Semih, 2016. "The Economic Impact of Syrian Refugees on Host Countries: Quasi-Experimental Evidence from Turkey," *American Economic Review* 106 (5): 456-460.

Van den Berg, Gerard and Bettina Drepper, 2016. "Inference for Shared-Frailty Survival Models with Left-Truncated Data," *Econometric Reviews* 35(6): 1075-1098.

Wichmann, Søren, Eric W. Holman, and Cecil H. Brown (editors). 2016. *The ASJP Database* (version 17).

Wooldridge, Jeffrey M., 2007. "Inverse Probability Weighted M-Estimation for General Missing Data Problems," *Journal of Econometrics* 141(2): 1281-1301.

Yuen, Hok Pan and Andrew Mackinnon, 2016. "Performance of Joint Modelling of Time-to-Event Data with Time-Dependent Predictors: An Assessment Based on Transition to Psychosis Data," *PeerJ* 4: e2582.

7. Tables

Table 1: Key Variables in Each Equation

	Type	*Time-varying?*	*Migration Hazard*	*Wage Assimilation*
Individual-Level Variables				
Years since migration	Years	Yes	–	☑
Female	Binary	No	☑	☑
Years of formal schooling	Years	Yes	☑	☑
Years of job training	Years	Yes	☑	☑
Years of work experience	Years	Yes	☑	☑ (quadratic)
Age	Years	Yes	in baseline hazard	☑ (quadratic)
Ethnic German	Binary	No	☑	☑
Place of upbringing	Categorical	No	☑	☑
Parents' education	Categorical	No	☑	☑
Current residence	Categorical	Yes	–	☑
Macro-Level Variables				
GDP growth in Germany	Continuous	Yes	☑	☑
GDP per capita in home country	Continuous	Yes	☑	☑ (at arrival)
Instability in home country	Continuous	Yes	☑	☑ (at arrival)
Linguistic distance	Continuous	No	☑	☑
Distance between capitals	Continuous	No	☑	☑
Home country's region	Categorical	No	☑	☑
Guest worker treaty	Binary	Yes	☑	–
Time trend	Years	Yes	☑ (quadratic)	☑ (quadratic)
Unified Germany	Binary	Yes	☑	☑

Table 2: Summary Statistics, Main Variables

Variables	*Sample A*	*Sample B*	*Sample C*	*Sample D*
Female	0.533	0.479	0.517	0.419
Ethnic German	0.323	0.316	0.237	0.252
Place of upbringing				
Large or medium city	0.402	0.405	0.355	0.345
Small city	0.202	0.212	0.225	0.234
Unknown	0.033	0.030	0.029	0.029
Parents' education				
General sec and upper vocational	0.310	0.321	0.169	0.175
Higher education	0.125	0.129	0.074	0.074
Unknown	0.091	0.086	0.103	0.096
Current residence				
West Germany - urban	...	...	0.830	0.830
East Germany - urban	...	...	0.010	0.010
West Germany - rural	...	...	0.150	0.153
Instability in home country				
Limited political violence	0.062	0.060	0.102[a]	0.100[a]
Serious political violence	0.043	0.038	0.058[a]	0.040[a]
Warfare	0.075	0.069	0.087[a]	0.074[a]
Home country's region				
America	0.018	0.018	0.015	0.014
Asia	0.176	0.161	0.261	0.207
Ex-USSR	0.317	0.303	0.143	0.148
Eastern Europe	0.315	0.328	0.312	0.340
Western Europe	0.151	0.168	0.259	0.283
Guest worker treaty	0.157	0.175	...	...
Unified Germany	0.327	0.335	0.746	0.725
Years since migration	...	...	17.411	17.408
			(9.175)	(8.379)
Years of formal schooling	9.504	9.670	9.343	9.474
	(2.754)	(2.771)	(3.200)	(3.270)
Years of job training	0.860	0.909	1.173	1.296
	(1.439)	(1.463)	(1.800)	(1.880)
Years of work experience	6.030	5.110	19.214	21.351
	(8.162)	(6.730)	(12.190)	(10.757)
Age	25.954	24.465	44.473	43.747
	(9.174)	(7.689)	(10.979)	(9.767)
GDP growth in Germany	2.281	2.270	1.634	1.654
	(1.920)	(1.906)	(1.658)	(1.660)
Log GDP per capita in home country	9.282	9.302	9.260[a]	9.288[a]
	(0.606)	(0.604)	(0.529)	(0.526)
Linguistic distance	0.886	0.884	0.885	0.880
	(0.108)	(0.110)	(0.119)	(0.120)
Log of distance between capitals	7.367	7.358	7.304	7.270
	(0.707)	(0.716)	(0.625)	(0.626)
Log wage per hour	...	...	...	2.110
				(0.411)
N of observations	108,779	70,679	46,713	28,711

Notes: Table shows the mean and standard deviation of variables in the four samples used in the paper. Standard deviations are in parenthesis and not reported for dummy variables. The four samples are (A) pre-migration full sample, (B) pre-migration estimation sample with at least one post-migration observation on wage, (C) post-migration survey sample which includes non-employed immigrants, and (D) post-migration sample with non-missing wage data. Superscript *a* indicates the year of arrival for samples C and D. Omitted categories are described in notes to Table 3.

Table 3: Timing of Migration Equation

	Cox	*Weibull*	*Gompertz*	*Weibull Frailty*	*Gompertz Frailty*
Female	-0.008	-0.023	-0.044	-0.030	-0.082**
	(0.030)	(0.030)	(0.034)	(0.029)	(0.032)
Years of formal schooling	-0.027***	-0.040***	-0.022***	-0.048***	-0.038***
	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)
Years of job training	0.001	-0.016	0.003	-0.021**	-0.005
	(0.010)	(0.010)	(0.012)	(0.010)	(0.010)
Years of work experience	-0.015***	-0.021***	-0.032***	-0.022***	-0.049***
	(0.004)	(0.003)	(0.005)	(0.002)	(0.003)
Ethnic German	0.358***	0.358***	0.352***	0.386***	0.408***
	(0.042)	(0.042)	(0.044)	(0.047)	(0.051)
Place of upbringing					
Large or medium city	-0.017	-0.000	-0.023	0.012	-0.000
	(0.035)	(0.036)	(0.039)	(0.035)	(0.039)
Small city	0.020	0.031	0.018	0.034	0.026
	(0.037)	(0.038)	(0.039)	(0.039)	(0.043)
Parents' education					
General sec and upper vocational	-0.269***	-0.248***	-0.267***	-0.277***	-0.330***
	(0.038)	(0.039)	(0.041)	(0.038)	(0.043)
Higher education	-0.195***	-0.163***	-0.186***	-0.185***	-0.238***
	(0.051)	(0.051)	(0.052)	(0.052)	(0.058)
GDP growth in Germany	0.054***	0.052***	0.051***	0.049***	0.043***
	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)
Log GDP per capita in home country	-0.323***	-0.326***	-0.318***	-0.338***	-0.342***
	(0.030)	(0.031)	(0.032)	(0.035)	(0.038)
Instability in home country					
Limited political violence	0.522***	0.540***	0.507***	0.539***	0.505***
	(0.047)	(0.048)	(0.048)	(0.052)	(0.054)
Serious political violence	0.438***	0.452***	0.449***	0.441***	0.427***
	(0.065)	(0.066)	(0.066)	(0.073)	(0.076)
Warfare	0.487***	0.504***	0.446***	0.525***	0.478***
	(0.048)	(0.048)	(0.050)	(0.051)	(0.054)
Linguistic distance	-0.402**	-0.393**	-0.353**	-0.411**	-0.370**
	(0.160)	(0.162)	(0.170)	(0.169)	(0.183)
Log of distance between capitals	-0.108***	-0.113***	-0.084**	-0.125***	-0.103**
	(0.032)	(0.032)	(0.034)	(0.038)	(0.041)
Home country's region					
America	0.737***	0.757***	0.697***	0.815***	0.807***
	(0.151)	(0.155)	(0.167)	(0.164)	(0.181)
Asia	0.437***	0.428***	0.460***	0.451***	0.522***
	(0.099)	(0.102)	(0.108)	(0.113)	(0.125)
Ex-USSR	-0.021	-0.022	-0.015	-0.053	-0.099
	(0.101)	(0.104)	(0.110)	(0.117)	(0.128)
Eastern Europe	0.449***	0.443***	0.480***	0.452***	0.504***
	(0.106)	(0.109)	(0.116)	(0.121)	(0.133)
Western Europe	0.483***	0.504***	0.513***	0.558***	0.649***
	(0.110)	(0.113)	(0.118)	(0.127)	(0.140)
Guest worker treaty	1.387***	1.341***	1.366***	1.444***	1.547***
	(0.071)	(0.072)	(0.073)	(0.062)	(0.065)
Time trend, 1960=1	0.077***	0.075***	0.074***	0.082***	0.086***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
Time trend squared / 100	-0.011	-0.008	-0.011	-0.008	-0.009
	(0.008)	(0.008)	(0.008)	(0.009)	(0.010)
Unified Germany	-0.153***	-0.170***	-0.147***	-0.194***	-0.187***
	(0.049)	(0.050)	(0.049)	(0.055)	(0.056)

Intercept	…	-2.411***	-1.697***	-2.423***	-1.583***
		(0.458)	(0.479)	(0.510)	(0.555)
$\hat{\varphi}$ (parameter in Gompertz hazard)	…	…	0.068***	…	0.098***
			(0.004)		(0.004)
Test for presence of heterogeneity					
$\widehat{Var}(\eta_i)$ (frailty variance)	…	…	…	0.086	0.285
χ^2 (LR test of $Var(\eta_i) = 0$)	…	…	…	22.840	111.595
p-value	…	…	…	0.000	0.000

Notes: Table presents estimates of the proportional hazard model that corresponds to Equation (8). Cox, Weibull, and Gompertz denote distributional assumptions of the baseline hazard. The estimates in the first three columns do not account for individual unobserved heterogeneity (or frailty), while estimates in the last two columns do. Individual-specific random effects in the last two columns are assumed to have inverse Gaussian distribution. The shared-frailty models are estimated using streg with the frailty option in Stata.

N of observations=70,679; N of immigrants=5,820. The sample is limited to immigrants with at least one wage observation after migration. See Table A3-1 for the full sample results. Standard errors are in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Robust standard errors are only allowed and reported for the models without frailty.

Base/omitted categories are rural area for the place of upbringing, “basic secondary and lower vocational” for parents’ education, “no episodes of political violence” for instability in home country, and Africa for home country’s region. Unknown place of upbringing and unknown parents’ education are also included in the estimates but not shown here.

Table 4: Selection Equations, Bivariate Probit

Variables	Reports wage	Stays in panel	Variables	Reports wage	Stays in panel
Years since migration	0.003***	0.001***	Log of distance b/w	0.031***	-0.004
	(0.000)	(0.000)	capitals	(0.006)	(0.004)
Female	-0.080***	0.016***	Instability in home country at arrival		
	(0.004)	(0.003)	Limited political violence	0.013*	-0.004
Years of formal schooling	0.013***	0.000		(0.007)	(0.005)
	(0.001)	(0.000)	Serious political violence	-0.019**	0.004
Years of job training	0.014***	-0.000		(0.010)	(0.006)
	(0.001)	(0.001)	Warfare	0.014*	-0.007
Years of work experience	0.020***	0.001***		(0.008)	(0.005)
	(0.000)	(0.000)	Home country's region		
Age	-0.021***	-0.001***	America	0.008	0.004
	(0.000)	(0.000)		(0.027)	(0.018)
Ethnic German	0.021***	0.014***	Asia	0.025	0.013
	(0.006)	(0.004)		(0.020)	(0.013)
Place of upbringing			Ex-USSR	0.066***	0.011
Large or medium city	-0.010**	0.002		(0.020)	(0.013)
	(0.005)	(0.003)	Eastern Europe	0.124***	0.009
Small city	0.009*	-0.001		(0.021)	(0.013)
	(0.005)	(0.004)	Western Europe	0.089***	-0.011
Parents' education				(0.021)	(0.014)
Secondary education	-0.008	-0.003	Time trend, 1960=1	-0.001*	-0.004***
	(0.006)	(0.004)		(0.000)	(0.000)
Higher education	-0.022***	0.004	Unified Germany	0.013	0.021***
	(0.008)	(0.006)		(0.010)	(0.008)
Current residence			Unemployment rate	-0.013***	...
West Germany – urban	0.094***	-0.014	among natives	(0.001)	
	(0.020)	(0.013)	Mode of interview		
East Germany - urban	0.153***	0.004	Self-written and mailed	...	-0.065***
	(0.027)	(0.018)			(0.005)
West Germany - rural	0.080***	-0.014	Computer assisted	...	-0.004
	(0.021)	(0.013)			(0.004)
GDP growth in Germany	-0.002	-0.000	Interviewer		
	(0.001)	(0.001)	First interview	...	-0.069***
Log GDP per capita in	0.021***	0.008**			(0.003)
home country at arrival	(0.006)	(0.004)	Different interviewer	...	-0.030***
Linguistic distance	0.011	-0.040***			(0.005)
	(0.019)	(0.015)	Predicted 3-year growth	...	-0.001***
			in home country		(0.000)

Notes: N=46,713. Table presents the joint maximum likelihood estimates of two probit equations: one for non-missing wage (column 1) and another one for survey re-interviewing (column 2). Reported are the marginal effects (MEs) evaluated at sample means. Robust standard errors are in parentheses; *** p<0.01, ** p<0.05, * p<0.1. The Wald LR test of independent equations of $corr(\upsilon_1, \upsilon_2) = 0$: chi-squared = 27.0134***; $\widehat{corr}(\upsilon_1, \upsilon_2) =$ 0.060. Omitted/base categories are face-to-face interviews for the mode of interview and the same interviewer; other omitted categories are listed in notes to Table 3. Work experience, age, and linear trend enter the estimation in a quadratic form, but MEs are at the mean. The means are reported in column "Sample C" of Table 2.

Table 5: Wage Equation with Random Assimilation Slope and Random Intercept

Variables	*OLS, No IPW*	*Mixed Model, No IPW*	*Mixed Model, IPW, Baseline*	*Mixed Model, IPW, Year FE*
	(1)	*(2)*	*(3)*	*(4)*
Years since migration	0.009***	0.011***	0.011***	0.010***
	(0.000)	(0.001)	(0.001)	(0.001)
Female	-0.275***	-0.246***	-0.252***	-0.253***
	(0.005)	(0.010)	(0.011)	(0.011)
Years of formal schooling	0.030***	0.028***	0.026***	0.026***
	(0.001)	(0.002)	(0.002)	(0.002)
Years of job training	0.010***	0.011***	0.007**	0.008**
	(0.001)	(0.003)	(0.003)	(0.004)
Years of work experience	0.009***	0.007***	0.010***	0.009***
	(0.001)	(0.002)	(0.003)	(0.003)
Years of work experience squared / 100	-0.003	-0.000	-0.010	-0.007
	(0.002)	(0.004)	(0.007)	(0.006)
Age	0.032***	0.036***	0.025***	0.026***
	(0.003)	(0.004)	(0.006)	(0.006)
Age squared / 100	-0.043***	-0.048***	-0.034***	-0.036***
	(0.003)	(0.005)	(0.007)	(0.007)
Ethnic German	0.030***	0.030**	0.029*	0.033*
	(0.007)	(0.014)	(0.016)	(0.017)
Place of upbringing				
Large or medium city	0.017***	0.011	0.006	0.005
	(0.005)	(0.011)	(0.012)	(0.012)
Small city	0.020***	-0.000	-0.003	-0.004
	(0.005)	(0.012)	(0.013)	(0.013)
Parents' education				
General sec and upper vocational	0.028***	0.027**	0.021	0.020
	(0.007)	(0.013)	(0.015)	(0.015)
Higher education	0.097***	0.110***	0.123***	0.122***
	(0.012)	(0.022)	(0.025)	(0.025)
Current residence				
West Germany – urban	0.390***	0.374***	0.372***	0.380***
	(0.037)	(0.053)	(0.059)	(0.059)
East Germany – urban	0.107**	0.137*	0.140*	0.145*
	(0.049)	(0.071)	(0.075)	(0.076)
West Germany – rural	0.344***	0.324***	0.318***	0.328***
	(0.038)	(0.053)	(0.059)	(0.060)
GDP growth in Germany	0.000	0.000	0.000	...
	(0.001)	(0.001)	(0.002)	
Log GDP per capita in home country at arrival	0.046***	0.043***	0.046***	0.043***
	(0.008)	(0.014)	(0.015)	(0.015)
Instability in home country at arrival				
Limited political violence	-0.039***	-0.023	-0.022	-0.023
	(0.007)	(0.016)	(0.018)	(0.018)
Serious political violence	-0.023*	-0.019	-0.019	-0.018
	(0.014)	(0.027)	(0.030)	(0.030)
Warfare	-0.035***	-0.016	-0.015	-0.014
	(0.009)	(0.017)	(0.020)	(0.019)
Linguistic distance	-0.246***	-0.302***	-0.307***	-0.307***
	(0.024)	(0.057)	(0.063)	(0.063)
Log of distance between capitals	0.003	0.016	0.013	0.014
	(0.008)	(0.014)	(0.016)	(0.016)
Home country's region				
America	0.118***	0.109	0.082	0.086

	(0.038)	(0.069)	(0.076)	(0.076)
Asia	0.189***	0.124***	0.105**	0.106**
	(0.030)	(0.047)	(0.051)	(0.051)
Ex-USSR	0.097***	0.069	0.058	0.058
	(0.030)	(0.047)	(0.050)	(0.050)
Eastern Europe	0.159***	0.136***	0.122**	0.123**
	(0.031)	(0.047)	(0.050)	(0.050)
Western Europe	0.136***	0.120**	0.101*	0.104*
	(0.032)	(0.051)	(0.055)	(0.055)
Time trend, 1960=1	0.032***	0.023***	0.021**	...
	(0.003)	(0.004)	(0.008)	
Time trend squared / 100	-0.041***	-0.030***	-0.027***	...
	(0.004)	(0.004)	(0.010)	
Unified Germany	-0.023**	-0.007	-0.018	...
	(0.009)	(0.008)	(0.015)	
Intercept	-0.257*	-0.170	0.107	0.439*
	(0.139)	(0.224)	(0.315)	(0.257)
Year FE	No	No	No	Yes
R^2	0.269	...	...	...
Standard deviation of $\hat{a}_i$, $\hat{\sigma}_a$	...	0.449	0.528	0.526
		(0.011)	(0.021)	(0.021)
Standard deviation of $\hat{b}_i$, $\hat{\sigma}_b$	...	0.021	0.028	0.027
		(0.001)	(0.002)	(0.002)
Correlation ($\hat{a}_i$, $\hat{b}_i$), $\hat{\rho}_{ab}$	...	-0.776	-0.826	-0.825
		(0.015)	(0.012)	(0.012)

Notes: Table presents the estimates of the wage model with a random intercept a_i and a random slope b_i on years since migration. The dependent variable is the log of hourly wage rate. Estimates correspond to Equation (7). IPW stands for inverse propensity weighting. Weights are calculated as the product of sampling weights and the inverse of the predicted joint probability of observing a positive wage and staying in the survey. Column 4 includes 30 year fixed effects instead of a quadratic polynomial of time trend. N of observations=28,711. Robust standard errors are in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Omitted categories are listed in the notes to Table 3.

Table 6: Joint Hazard-Longitudinal Model Estimates

Variables	***Hazard Equation***	***Change in Median Age-at-Migration***	***Wage Equation***
	(1)	***(2)***	***(3)***
Years since migration	...	...	$\bar{\delta}$ =0.0071***
			(0.000)
Female	-0.023	0.153	-0.252***
	(0.033)		(0.008)
Years of formal schooling	-0.017***	0.116	0.026***
	(0.006)		(0.002)
Years of job training	0.027***	-0.182	0.007***
	(0.010)		(0.003)
Years of work experience	-0.013***	0.085	0.008***
	(0.004)		(0.002)
Years of work experience squared / 100	...	...	-0.005***
			(0.001)
Age	...	...	0.029***
			(0.001)
Age squared / 100	...	...	-0.038***
			(0.001)
Ethnic German	0.511***	-3.442	0.037**
	(0.052)		(0.016)
Place of upbringing			
Large or medium city	-0.026	0.176	0.003
	(0.040)		(0.012)
Small city	0.023	-0.153	-0.005
	(0.044)		(0.014)
Parents' education			
General sec and upper vocational	-0.409***	2.754	0.018
	(0.044)		(0.014)
Higher education	-0.327***	2.204	0.119***
	(0.060)		(0.019)
Current residence			
West Germany - urban	...	...	0.375***
			(0.042)
East Germany - urban	...	...	0.133**
			(0.053)
West Germany - rural	...	...	0.313***
			(0.043)
GDP growth in Germany	0.041***	-0.276	0.001
	(0.008)		(0.001)
Log GDP per capita in home country	-0.450***	3.031	0.038***
	(0.011)		(0.004)
Instability in home country			
Limited political violence	0.557***	-3.751	-0.019
	(0.055)		(0.018)
Serious political violence	0.501***	-3.372	-0.023
	(0.076)		(0.023)
Warfare	0.553***	-3.720	-0.017
	(0.055)		(0.018)
Linguistic distance	-0.370**	2.494	-0.296***
	(0.173)		(0.050)
Log of distance between capitals	-0.167***	1.126	0.010
	(0.014)		(0.008)
Home country's region			
America	0.834***	-5.616	0.087***

	(0.148)		(0.025)
Asia	0.520***	-3.502	0.114**
	(0.111)		(0.052)
Ex-USSR	-0.176	1.182	0.049*
	(0.111)		(0.025)
Eastern Europe	0.415***	-2.792	0.120***
	(0.119)		(0.040)
Western Europe	0.755***	-5.080	0.110**
	(0.118)		(0.045)
Guest worker treaty	1.642***	-11.052	...
	(0.056)		
Time trend, 1960=1	0.027***	-0.195	0.022***
	(0.005)		(0.002)
Time trend squared / 100	0.123***	...	-0.028***
	(0.009)		(0.002)
Unified Germany	-0.251***	1.689	-0.018***
	(0.056)		(0.004)
Intercept	0.274	...	0.100***
	(0.254)		(0.030)

Parameter Estimates

Parameter Description	***Parameter***	***Estimate***	***Std. Error***
Slope in Gompertz baseline hazard	$\hat{\varphi}$	0.070	(0.003)
Standard deviation of $\hat{a}_i$	$\hat{\sigma}_a$	0.508	(0.021)
Standard deviation of $\hat{b}_i$	$\hat{\sigma}_b$	0.026	(0.001)
Standard deviation of $\hat{c}_i$	$\hat{\sigma}_c$	0.606	(0.006)
Correlation ($\hat{a}_i$, $\hat{b}_i$)	$\hat{\rho}_{ab}$	-0.828	(0.022)
Correlation ($\hat{b}_i$, $\hat{c}_i$)	$\hat{\rho}_{bc}$	0.204	(0.036)
Correlation ($\hat{a}_i$, $\hat{c}_i$)	$\hat{\rho}_{ac}$	-0.067	(0.014)

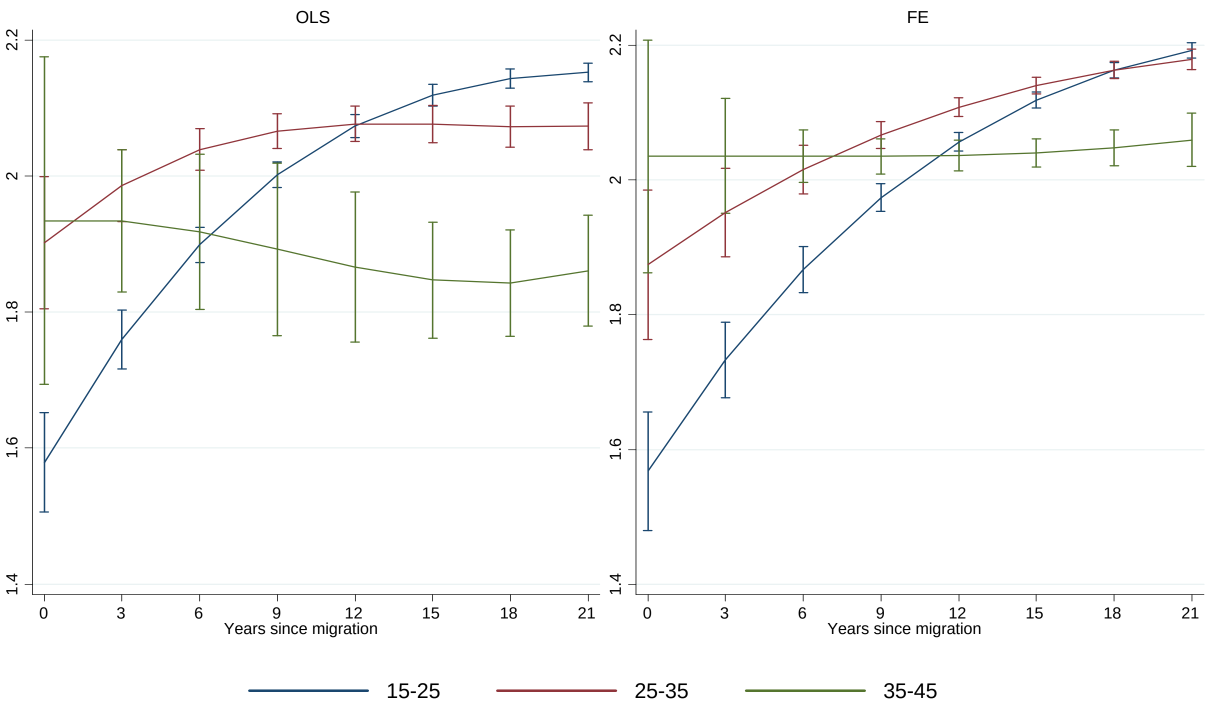
Notes: Table presents joint maximum likelihood estimates of the timing of migration equation (column 1) and wage equation (column 3). The timing of migration equation corresponds to Equation (11'). It is a parametric proportional hazard model with shared frailty and Gompertz-form baseline hazard. The wage equation is described by Equation (7') as a linear mixed-effects model with random intercept a_i and random slope $\bar{\delta} + b_i$; the dependent variable is the log of the hourly wage rate. Estimates are weighted using sample weights and the inverse of the predicted joint probability of observing positive wage and staying in the survey. Individual-specific random effects a_i, b_i, c_i are assumed to be jointly normally distributed. The change in predicted median age-at-migration in column 2 is based on Equation (18).

Number of observations=70,679; number of immigrants=5,820. Standard errors (in parentheses) are computed using the numerical Hessian; *** p<0.01, ** p<0.05, * p<0.1. Omitted categories are listed in the notes to Table 3.

8. Figures

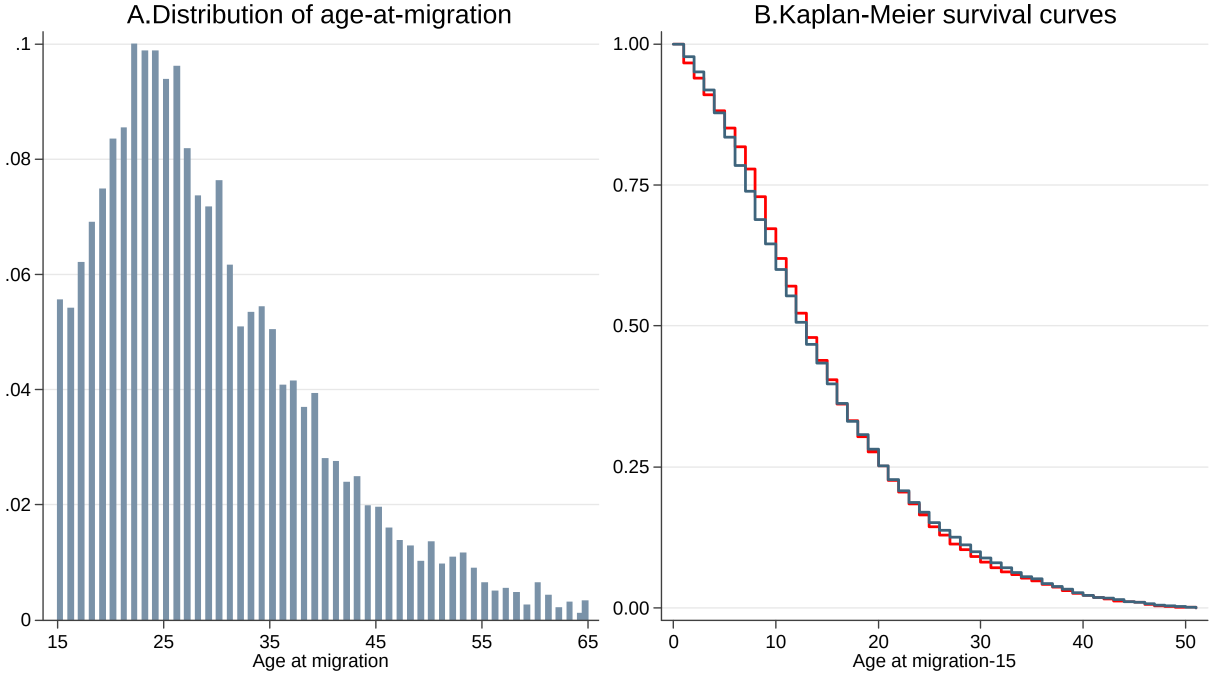
Note: Figure 1 is shown in the text in Section 3.2

Figure 2: Wage Trajectory by Age-at-Migration



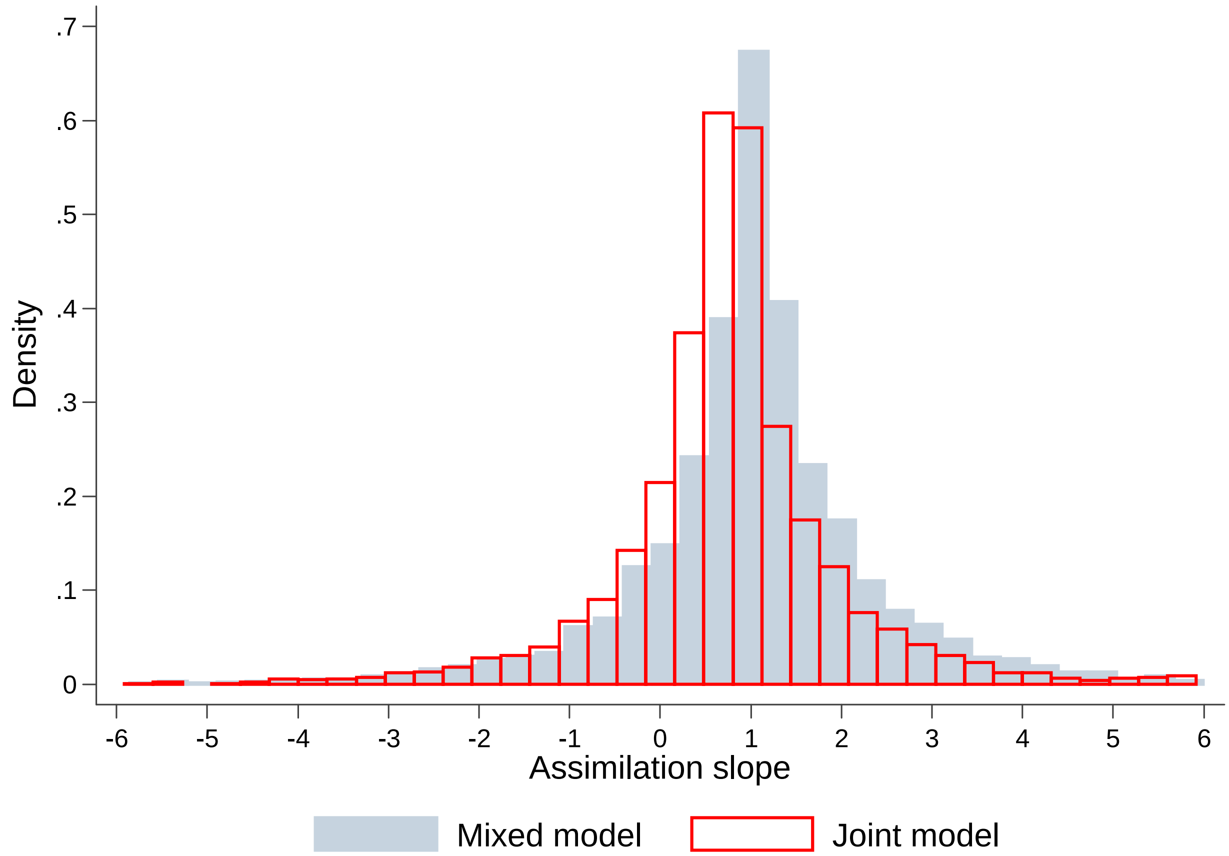
Notes: The wage profiles are predicted from the regression of the log of hourly wage on the full two-way interaction between the categorical variable for age-at-migration and a cubic polynomial in years since migration with and without individual fixed effects (FE) and with robust standard errors. All immigrants are categorized into three groups based on the age of arrival to Germany: migrated at age 15 to 25, 26 to 35, and 36 to 45. The 95 percent confidence interval for each point estimate is also shown.

Figure 3: Age-at-Migration



Notes: Panel A depicts the histogram of age-at-migration across all immigrants in the data. Panel B plots the Kaplan–Meier estimate of the survival function by gender for the period before migration. The zero on the horizontal axis is equivalent to the age of migration being 15 years old. The standard log-rank test fails to reject the null of the equality of survivor functions between males (red line) and females (blue line). The sample in both panels includes individuals who immigrated to Germany between the ages of 15 and 65.

Figure 4: Distribution of Individual-Specific Rates of Wage Assimilation



Notes: Figure depicts two distributions of predicted individual-specific assimilation rates, $(\hat{\delta} + \hat{b}_i)\times 100$. The first distribution is estimated from the mixed-effects wage model with random intercept and random slope (mean=1.06, sd=2.77). The second distribution is obtained from the joint hazard-longitudinal model with three random effects (mean=0.71, sd=2.59). Only values within a 95 percent range are shown.

Appendix A1. Technical Appendix

A. Joint Estimation

The parameter estimates of the joint model (i.e., $\theta_{abc}, \theta_m, \theta_w$) are obtained by maximizing the joint likelihood function given in Equation (12). The joint estimation requires integrating over the three random effects a_i, b_i and c_i, which is computationally intensive. Thus, a modified joint likelihood is maximized. There are three key steps in calculating the modified likelihood. First, the multivariate normal distribution of a_i, b_i and c_i is expressed as the conditional distribution of a_i and bivariate distribution of b_i and c_i, i.e., $f(a_i, b_i, c_i; \theta_{abc}) = f(a_i|b_i, c_i; \theta_{a|b,c}) \times f(b_i, c_i; \theta_{bc})$. Second, the part of the log likelihood independent of a_i (i.e., the probability distribution function of the hazard model and the bivariate distribution of b_i and c_i) are taken out of the integral over a_i. Third, with a few mathematical manipulations, the part of the log likelihood dependent on a_i (i.e., the probability distribution of wages and the conditional distribution of a_i) is expressed as a normalized Gaussian function whose integral equals 1. Thus, the integral over a_i can be eliminated. The detailed mathematical steps are given in the next subsection.

The final log likelihood function is given by Equation (A2). The negative of the log likelihood function is minimized using the nonlinear multivariate constrained optimization and the interior point algorithm available in Matlab. To make the computation efficient, the feasible ranges of b_i and c_i over which the function is integrated is computed using Singular Value Decomposition. The detailed calculation of the range (which covers approximately 99.97 percent of the probability mass of b_i and c_i) is given in Subsection C.

The standard errors of the parameter estimates are calculated by taking the square root of the diagonal elements of the inverted Hessian matrix. The Hessian is numerically computed.

B. Modified Joint Likelihood Function

This section explains the detailed mathematical steps taken to reduce the likelihood function's dependence on a_i and to eliminate the integral over a_i. Let us express the joint likelihood given in Equation (12) as the following:

$$L(\theta) = \prod_{i=1}^{n} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left\{ \prod_{s=1}^{S} f(W_{is}|T_i, a_i, b_i; \theta_w) \right\} f(T_i|Z_{it}, c_i; \theta_m) f(a_i|b_i, c_i; \theta_{a|b,c}) \times f(b_i, c_i; \theta_{bc}) \mathrm{d}a_i \mathrm{d}b_i \mathrm{d}c_i \tag{A1}$$

where $f(a_i, b_i, c_i; \theta_{abc})$ is broken into $f(a_i|b_i, c_i; \theta_{a|b,c}) f(b_i, c_i; \theta_{bc})$ In the above equation, only $\prod_{s=1}^{S} f(W_{is}|a_i, b_i; \theta_w)$ and $f(a_i|b_i, c_i; \theta_{a|b,c})$ depend on a_i. Let us focus on these two expressions:

$$\begin{aligned}
&\int_{-\infty}^{\infty} f(W_{is}|a_i, b_i; \theta_w) f(a_i|b_i, c_i; \theta_{a|b,c}) \mathrm{d}a_i \\
&\quad = \int_{-\infty}^{\infty} \left\{ \prod_{s=1}^{S} (2\pi\sigma_\epsilon^2)^{-1/2} \exp\{ \frac{-(W_{is} - a_0 - (\delta + b_i) YSM_{is} - \beta X_{is} - a_i)^2}{2\sigma_\epsilon^2} \} \right\} \\
&\quad \times (2\pi {\sigma_{a|b,c}}^2)^{-1/2} \exp \frac{-(a_i - \Sigma_{12}\Sigma_{22}^{-1} \begin{pmatrix} b_i \\ c_i \end{pmatrix})^2}{2{\sigma_{a|b,c}}^2} \mathrm{d}a_i
\end{aligned}$$

Next, let us take the constant terms outside of the integral, express $W_{is} - a_0 - (\delta + b_i) YSM_{is} - \beta X_{is}$ as D_s, apply the formula that $\mu_{a|b,c} = \Sigma_{12}\Sigma_{22}^{-1} \begin{pmatrix} b_i \\ c_i \end{pmatrix}$, and bring the product inside the exponential function, which converts into a sum.

$$= ((2\pi\sigma_\epsilon^2)^{-1/2})^S \times (2\pi {\sigma_{a|b,c}}^2)^{-1/2} \int_{-\infty}^{\infty} \exp \frac{-\sum_{s=1}^{S} (D_s - a_i)^2}{2\sigma_\epsilon^2} \times \exp \frac{-(a_i - \mu_{a|b,c})^2}{2{\sigma_{a|b,c}}^2} \mathrm{d}a_i$$

Next, the product of the constants $((2\pi\sigma_\epsilon^2)^{-1/2})^S \times (2\pi {\sigma_{a|b,c}}^2)^{-1/2}$ is expressed as C, and a standard formula of the square of a difference is applied to $(D_s - a_i)^2$ and $(a_i - \mu_{a|b,c})^2$.

$$= C \int_{-\infty}^{\infty} \exp\Big(\frac{-\sum_{s=1}^{S} ({D_s}^2 + {a_i}^2 - 2a_i D_s)}{2\sigma_\epsilon^2} \Big) \times \exp\Big(\frac{-({a_i}^2 + {\mu_{a|b,c}}^2 - 2a_i \mu_{a|b,c})}{2{\sigma_{a|b,c}}^2} \Big) \mathrm{d}a_i$$

Next, we can take the other constant terms out of the integral and express their product with C as C_1.

$$= C \times \exp\{ -\frac{\sum_{s=1}^{S} {D_s}^2}{2\sigma_\epsilon^2} \} \times \exp\{ -\frac{{\mu_{a|b,c}}^2}{2{\sigma_{a|b,c}}^2} \} \int_{-\infty}^{\infty} \exp\Big(\frac{-\sum_{s=1}^{S} ({a_i}^2 - 2a_i D_s)}{2\sigma_\epsilon^2} - \frac{({a_i}^2 - 2a_i \mu_{a|b,c})}{2{\sigma_{a|b,c}}^2} \Big) \mathrm{d}a_i$$

$$= C_1 \int_{-\infty}^{\infty} \exp\Big(-\frac{{a_i}^2}{2}\{\frac{S}{\sigma_\epsilon^2} + \frac{1}{{\sigma_{a|b,c}}^2}\} + a_i\{\frac{\sum_{s=1}^{S} D_s}{\sigma_\epsilon^2} + \frac{\mu_{a|b,c}}{{\sigma_{a|b,c}}^2}\}\Big) \mathrm{d}a_i$$

In the following step, $\{\frac{S}{\sigma_\epsilon^2} + \frac{1}{{\sigma_{a|b,c}}^2}\}$ is expressed as F and $\{\frac{\sum_{s=1}^{S} D_s}{\sigma_\epsilon^2} + \frac{\mu_{a|b,c}}{{\sigma_{a|b,c}}^2}\}$ as E.

$$= C_1 \int_{-\infty}^{\infty} \exp\Big(-\frac{{a_i}^2}{2}\{F\} + a_i\{E\}\Big) \mathrm{d}a_i$$

In the next two steps, $\frac{E^2}{F^2}$ is added and subtracted to the expression inside the exponential. Further, the first three terms inside the small round bracket ${a_i}^2 - \frac{2a_iE}{F} + \frac{E^2}{F^2}$ are combined together as $(a_i - \frac{E}{F})^2$.

$$= C_1 \int_{-\infty}^{\infty} \exp\Big(\frac{-F}{2}({a_i}^2 - \frac{2a_iE}{F} + \frac{E^2}{F^2} - \frac{E^2}{F^2})\Big) \mathrm{d}a_i$$
$$= C_1 \int_{-\infty}^{\infty} \exp\Big(\frac{-F}{2}{(a_i - \frac{E}{F})}^2 + \frac{E^2}{2F}\Big) \mathrm{d}a_i$$

Now, we can multiply and divide the expression inside the integral by $2\pi\frac{1}{F}^{1/2}$ to get the following expression:

$$= C_1 \times \exp\frac{E^2}{2F} \int_{-\infty}^{\infty} 2\pi\frac{1}{F}^{1/2} \left\{ 2\pi\frac{1}{F}^{-1/2} \exp(\frac{-1}{2\frac{1}{F}}{(a_i - \frac{E}{F})}^2) \right\} \mathrm{d}a_i$$

Note, the integrand is the normalized Gaussian function and hence equals 1.

$$= C_1 \times \exp\frac{E^2}{2F} \times (2\pi\frac{1}{F})^{1/2} \int_{-\infty}^{\infty} \left\{ (2\pi\frac{1}{F})^{-1/2} \exp(\frac{-1}{2\frac{1}{F}}{(a_i - \frac{E}{F})}^2) \right\} \mathrm{d}a_i$$
$$= C_1 \times \exp\frac{E^2}{2F} \times (2\pi\frac{1}{F})^{1/2}$$

Thus, the final modified joint likelihood function is:

$$L(\theta) = \prod_{i=1}^{n} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ C_1 \times \exp\frac{E^2}{2F} \times (2\pi\frac{1}{F})^{1/2} \right\} f(b_i, c_i; \theta_{bc}) f(T_i|c_i; \theta_m) \mathrm{d}b_i \mathrm{d}c_i \tag{A2}$$

where

$$
\begin{aligned}
C1 &= C \times \exp\{-\frac{\sum_{s=1}^{S} {D_s}^2}{2\sigma_\epsilon^2}\} \times \exp\{-\frac{{\mu_{a|b,c}}^2}{2\sigma_{a|b,c}^2}\} \\
C &= ((2\pi\sigma_\epsilon^2)^{-1/2})^S \times (2\pi\sigma_{a|b,c}^2)^{-1/2} \\
E &= \{\frac{\sum_{s=1}^{S} D_s}{\sigma_\epsilon^2} + \frac{\mu_{a|b,c}}{\sigma_{a|b,c}^2}\} \\
F &= \{\frac{S}{\sigma_\epsilon^2} + \frac{1}{\sigma_{a|b,c}^2}\} \\
D_s &= W_{is} - a_0 - (\delta + b_i)YSM_{is} - \beta X_{is} \\
\mu_{a|b,c} &= \Sigma_{12}\Sigma_{22}^{-1}\begin{pmatrix} b_i \\ c_i \end{pmatrix} \\
&= \begin{pmatrix} \sigma_{ab} & \sigma_{ac} \end{pmatrix}\begin{pmatrix} \sigma_b^2 & \sigma_{bc} \\ \sigma_{bc} & \sigma_c^2 \end{pmatrix}\begin{pmatrix} b_i \\ c_i \end{pmatrix} \\
&= \begin{pmatrix} \sigma_{ab}\sigma_b^2 + \sigma_{ac}\sigma_{bc} & \sigma_{ab}\sigma_{bc} + \sigma_{ac}\sigma_c^2 \end{pmatrix}\begin{pmatrix} b_i \\ c_i \end{pmatrix} \\
&= (\sigma_{ab}\sigma_b^2 + \sigma_{ac}\sigma_{bc})b_i + (\sigma_{ab}\sigma_{bc} + \sigma_{ac}\sigma_c^2)c_i \\
\Sigma_{12} &= \begin{pmatrix} Cov(a_i, b_i) & Cov(a_i, c_i) \end{pmatrix} \\
&= \begin{pmatrix} \sigma_{ab} & \sigma_{ac} \end{pmatrix} \\
\Sigma_{22} &= \begin{pmatrix} Var(b_i) & Cov(b_i, c_i) \\ Cov(b_i, c_i) & Var(c_i) \end{pmatrix} \\
&= \begin{pmatrix} \sigma_b^2 & \sigma_{bc} \\ \sigma_{bc} & \sigma_c^2 \end{pmatrix} \\
\sigma_{a|b,c}^2 &= \sigma_a^2 - \begin{pmatrix} \sigma_{ab}\sigma_b^2 + \sigma_{ac}\sigma_{bc} & \sigma_{ab}\sigma_{bc} + \sigma_{ac}\sigma_c^2 \end{pmatrix}\begin{pmatrix} \sigma_{ab} \\ \sigma_{ac} \end{pmatrix} \\
&= \sigma_a^2 - \begin{pmatrix} \sigma_{ab}^2\sigma_b^2 + \sigma_{ab}\sigma_{bc}\sigma_{ac} + \sigma_{ab}\sigma_{bc}\sigma_{ac} + \sigma_{ac}^2\sigma_c^2 \end{pmatrix} \\
&= \sigma_a^2 - (\sigma_{ab}^2\sigma_b^2 + 2\sigma_{ab}\sigma_{bc}\sigma_{ac} + \sigma_{ac}^2\sigma_c^2) \\
\sigma_{b|c}^2 &= \sigma_b^2 - \frac{\sigma_{bc}^2}{\sigma_c^2}
\end{aligned}
$$

$$
\begin{aligned}
\mu_{b|c} &= \mu_b + \Sigma_{12}\Sigma_{22}^{-1}(c_i - \mu_c) \\
&= 0 + \frac{\sigma_{bc}}{\sigma_c^2}(c_i - 0) \\
&= \frac{\sigma_{bc}}{\sigma_c^2}c_i
\end{aligned}
$$

C. Calculating the Range for Numerical Integration

We need to numerically integrate over b_i and c_i . As b_i and c_i are correlated and follow a bivariate normal $f(b_i, c_i) \sim \mathcal{N}(0, \Sigma_{bc})$, their distribution is given by:

$$f(b_i, c_i; \theta_{bc}) = \frac{1}{2\pi\sigma_b\sigma_c\sqrt{1-\rho^2}} \exp[-\frac{1}{2(1-\rho^2)} \left\{ \frac{b_i^2}{\sigma_b^2} - \frac{2\rho\sigma_b\sigma_c}{\sigma_b\sigma_c} + \frac{c_i^2}{\sigma_c^2} \right\}] \tag{A3}$$

The expression in exponential can be expressed in the matrix form as :

$$-\frac{1}{2(1-\rho^2)} \begin{bmatrix} b_i & c_i \end{bmatrix} A \begin{bmatrix} b_i \\ c_i \end{bmatrix} \tag{A4}$$

where A is the covariance matrix given by:

$$\begin{bmatrix} \frac{1}{\sigma_b^2} & -\frac{\rho}{\sigma_b\sigma_c} \\ -\frac{\rho}{\sigma_b\sigma_c} & \frac{1}{\sigma_c^2} \end{bmatrix} \tag{A5}$$

We can write

$$\begin{bmatrix} b_i \\ c_i \end{bmatrix} = R \begin{bmatrix} w \\ z \end{bmatrix} \tag{A6}$$

which gives

$$-\frac{1}{2(1-\rho^2)} \begin{bmatrix} w & z \end{bmatrix} R^T G R \begin{bmatrix} w \\ z \end{bmatrix} \tag{A7}$$

We use singular value decomposition to decompose A into R (rotation matrix) and G (singular matrix). This process of basis transformation ensures that w and z have covariance 0. The obtained G gives σ_w and σ_z as

$$G = \begin{bmatrix} \frac{1}{\sigma_w^2} & 0 \\ 0 & \frac{1}{\sigma_z^2} \end{bmatrix} \tag{A8}$$

Finally the range for b_i and c_i is given by the maximum of b_1, b_2 and c_1, c_2 obtained using the following equation:

$$\begin{bmatrix} |b_1| \\ |c_1| \end{bmatrix} = k\sqrt{1-\rho^2} R \begin{bmatrix} \sigma_w \\ \sigma_z \end{bmatrix} \tag{A9}$$

$$\begin{bmatrix} |b_2| \\ |c_2| \end{bmatrix} = k\sqrt{1-\rho^2} R \begin{bmatrix} \sigma_w \\ -\sigma_z \end{bmatrix} \tag{A10}$$

k is chosen to be 4 which ensures that 99.97 percent of probability mass of b and c is covered.

Appendix A2. Data

This appendix is largely replicated from the Data Appendix in Jain and Peter (2017). The German Socio-Economic Panel (GSOEP) is the longest-running panel of private households and persons in Europe. It is widely used in migration research, as it is one of a few national longitudinal surveys with a large representation of immigrants and substantial information on immigrants. Some examples of published migration studies based on GSOEP include Brücker *et al.* (2014b), Constant et al. (2009), Jaeger *et al.* (2010), Zimmermann (2007), among others.

A. Samples

GSOEP is collected and distributed by the German Institute for Economic Research, DIW Berlin. The survey started in 1984 and includes 31 survey waves as of 2014. In 1990, residents of the German Democratic Republic (East Germany) were included in the target population. Later, several additional samples were drawn to replenish the original sample and to include special sub-populations such as immigrants and high-income households.

With respect to immigrants, all samples can be divided into two large groups. In the first group (samples, A, C, E, G, H, J, and K), immigrants are sampled as part of the total population or subpopulation. In the random samples of the total population, the share of immigrants is relatively low: about 4 percent in the initial sample A and 7 to 12 percent in replenishment samples E, H, J, and K. Immigrants constitute about 5 percent of high-income earners in sample G and a mere 1.3 percent of East Germans in sample C. The second group of GSOEP samples includes samples B, D, and M, which focused specifically on immigrants. Sample B "Foreigners in West Germany" started in 1984 with 1393 households whose head came from one of the five largest guest-worker countries (Turkey, Italy, the former Yugoslavia, Greece, and Spain). Sample D "Immigrants" started in 1994/95 with 522 households, which consisted primarily of ethnic German immigrants from the former Soviet Union and Eastern Europe as well as asylum seekers mainly from the parts of Yugoslavia devastated by the war. Finally, sample M "Migration" started in 2013 with 2,723 households. It is designed to account for changes in the composition of migration to Germany since 1995 (Brücker *et al.*, 2014b). All three migration-focused samples have a substantial share of native-born population (27 percent) since some members of households are born in Germany.

The immigrant status in GSOEP is defined based on the country of birth or in earlier waves based on the country of residence since 1949. Immigrants in GSOEP come from more than 130 countries. We are using 95 percent of the original data since researchers outside the European Union are not allowed to access the entire dataset. We limit the sample of immigrants to those who were between the ages of 17 and 65 at the time of survey and who arrived to Germany after 1960 at age 15 or older. The sample of natives is constrained by age 17-65. We drop observations with missing values on migration status, country of origin, the year of migration, work experience, and the level of schooling. In total, we drop less than 10 percent of observations with missing values in the constrained sample. Given a very small percent of missing values, we assume that dropped observations are ignorable or missing completely at random.

B. Weights

Immigrant-focused GSOEP samples are not a random draw from the German immigrant population, and their composition in the GSOEP does not match the national composition of immigrants by country of origin. Due to the sampling design, there is a substantial oversampling of immigrants from the countries that signed guest-worker agreements and also from Poland and former Soviet Union. At the same time, immigrants from Asia, Africa, Middle East and other geographic areas are under-sampled. It is apparent that sample re-weighting is required to match the sample moments to the population moments. The GSOEP provides researchers with cross-sectional weights **phrf*, which we renamed as *CWEIGHT*. However, since many immigrants are sampled outside the main sampling frame, their cross-sectional weight is often set to zero; for example, more than 40 percent of sample D “Immigrants” have zero sampling weight.

To keep as many surveyed immigrants as possible in our estimation sample, we develop immigrant sampling weights (***IWEIGHT***) based on the annual share of each home country in the total German population. The OECD International Migration Database (OECD, 2016) and the German Central Register of Foreign Nationals (Ausländerzentralregisters) report the annual composition of foreign population by origin, which covers more than 99 percent of the foreign population from 1990 to 2015 and 91 to 96 percent from 1984 to 1989. *IWEIGHT* is obtained as a ratio of the country share in total German population to the country share in GSOEP sample for each year separately. The *IWEIGHT* for German-born respondents is above 1 due to oversampling of the immigrant population in GSOEP; it ranges between 1.01 and 1.24, with mean=1.08. Most oversampled home countries with *IWEIGHT* below 0.3 are countries of the former Soviet Union and Poland, while the top under-sampled countries with weights above 2 are Israel, Australia, and countries of East Asia and the Pacific. Since *IWEIGHT* does not account for the sample demographic composition, we control for standard weighting factors such as gender, age, and urban residence in any estimates that involve immigrant composition weights.

C. References for GSOEP data description

Ausländerzentralregisters (Central Register of Foreign Nationals). Ausländische Bevölkerung (Foreign Population Statistics), Statistisches Bundesamt (Federal Statistical Office), annual.

Brücker, Herbert, Andreas Hauptmann, Elke J. Jahn, and Richard Upward, 2014a. "Migration and Imperfect Labor Markets: Theory and Cross-country Evidence from Denmark, Germany and the UK," *European Economic Review*, Volume 66, February 2014: 205-225.

Brücker, Herbert, Martin Kroh, Simone Bartsch, Jan Goebel, Simon Kühne, Elisabeth Liebau, Parvati Trübswetter, Ingrid Tucci, and Jürgen Schupp, 2014b. "The New IAB-SOEP Migration Sample: An Introduction into the Methodology and the Contents," *SOEP Survey Papers*, No. 216: Series C. Berlin: DIW/SOEP.

Constant, Amelie F., Liliya Gataullina, and Klaus F. Zimmermann, 2009. "Ethnosizing Immigrants," *Journal of Economic Behavior & Organization*, 69(3), March 2009: 274-287.

Jaeger, David A., Thomas Dohmen, Armin Falk, David Huffman, Uwe Sunde, and Holger Bonin, 2010. "Direct Evidence on Risk Attitudes and Migration," *Review of Economics and Statistics*, August 2010, 92(3): 684-689.

OECD, 2016. OECD International Migration Database, http://www.oecd.org/els/mig/keystat.htm

Zimmermann, Klaus F., 2007. "The Economics of Migrant Ethnicity," *Journal of Population Economics* 20: 487-494.

Appendix A3. Variables

This appendix is largely replicated from the Data Appendix in Jain and Peter (2017).

A. Individual-level variables

Country of origin and year of immigration

Country of origin is defined as Germany if a person is born in Germany or immigrated before 1949. Another 130+ countries of origin are re-coded according to the UN country classification to link individual observations with macro indicators. Kurdistan is coded as Turkey, Benelux as Netherlands, and the Free City of Gdansk as Poland. Categories for "No nationality", "Africa", "Other unspecified foreign country", and "Unspecified country within EU" are coded as missing. The category "unspecified Eastern Europe", which mostly includes immigrants from former German territories of Eastern Europe, is kept separately, but linked with macro indicators from Poland. Year of immigration is the calendar year in which the first immigration to territories of the Federal Republic of

Germany occurred. Both variables are provided for public use as part of the biography and life history data; see documentation of biography variables in SOEP (2014a).

Years since migration (YSM)

Number of years since immigration, or the length of stay in the host country, is calculated as year of survey minus year of immigration.

Female, age, year of survey

Self-explanatory.

Years of formal schooling, job training, and work experience

These variables are constructed using the spell dataset on activity status between the ages of 15 and 65; see description in SOEP (2014a). We start with six main activities that include formal schooling, job training, full-time employment, part-time employment, military/civil service, and unemployment. If more than one activity is reported in a given year, then each activity gets a corresponding share of one year. Work experience at a given age is calculated by summing up all the spells of "full-time employment" and "part-time employment" prior to that age, assuming zero work experience before age 15. We also assume continuous schooling from age 7 to age 14 and no job training before age 15. Then, years of formal schooling and job training at each age are calculated as a running sum of corresponding spells up to a given age.

Ethnic German

A dummy variable indicating if an immigrant is of German descent from Eastern Europe.

Place of upbringing in childhood

Four categories are created to characterize the place of upbringing in childhood: [1] Medium or large city, [2] Small city, [3] Rural area, and [4] Unknown. The third category is chosen as a base category. The share of immigrants in the unknown category is about 3 percent. There is no definition of the city size in the questionnaire.

Parents' education

This variable represents the highest level of schooling completed by a parent: [1] Level I "Basic secondary, lower vocational or less", [2] Level II "General secondary or upper vocational", [3] Level III "Higher education or more", and [4] "Unknown level of parents' education". The first category is chosen as a base category. This variable is constructed based the level of general schooling and the level of professional education provided for each parent in the biography dataset *BIOPAREN* (SOEP, 2014a). First, we aggregate all levels of schooling into three categories. Level III includes degrees from technical engineering school, college, university, and foreign college. Level II includes degrees from intermediate school, technical school, upper secondary school, vocational school, foreign vocational school, health care school, and special technical school. Level I consists of other types of schooling which are not in Level II or III and include basic secondary school degree, incomplete secondary school, no schooling, apprenticeship, and on-the-job training. Then, we choose the highest level completed among parents. If information is only available for one parent, only that parent's data is used. If the level of schooling is missing for both parents, then these respondents are combined into the fourth category "Unknown level of parents' education". The share of immigrants in the unknown category is about 10 percent.

Type of location in Germany

Four categories for the type of location in Germany: [1] Urban area of West Germany, [2] Urban area of East Germany, [3] Rural area of West Germany, and [4] Rural area of East Germany. The last category is used as a base category.

Monthly earnings

Log of net income earned from employment last month in constant 2010 prices (in Euro). Net income means the amount after deduction of taxes, social security, and unemployment and health insurance. The amount excludes vacation pay or back pay. Net labor earnings last month in current prices are part of the dataset of generated variables (SOEP, 2014b). We make two adjustments to this variable:

- First, we exclude imputed values due to potential match bias from earnings imputation (Bollinger and Hirsch, 2006). Instead, we use inverse propensity weighting to account for missing values in earnings.
- Second, we adjust labor earnings to 2010 Euros using annual CPI for Germany (West Germany until 1990) (OECD, 2016).

Hourly wage

Log of net wage per hour last month in constant 2010 prices (in Euro). The hourly wage is calculated as total net income earned from employment last month in constant 2010 prices (in Euro) divided by the product of actual working hours per week and (30/7) number of weeks in a month. Contractual hours are not used because they are not available for the self-employed and exclude over-time work.

Remains in the panel

A binary indicator that takes up the value of one if a respondent in year *t* remains in the GSOEP panel in one of the subsequent survey rounds, and the value of zero if otherwise. Observations in the last survey round in year 2014 are coded as zero.

Reports wage

A binary indicator that takes up the value of one if positive wage is observed, and the value of zero if otherwise.

Unemployment rate among natives by region

The unemployment rate is measured in a traditional way as a percent share of the unemployed respondents in the labor force for each state and year. Immigrants are excluded from the calculation of the unemployment rate to satisfy the conditions of exclusion restrictions (see Section 3.6).

Mode of interview

The GSOEP uses several different modes of interviews, which we classify into 3 categories: [1] face-to-face, [2] self-written and mailed, and [3] computer assisted. Web-based interviews are combined with computer assisted personal interviews into one category.

Interviewer

All household interviews are classified into three categories: [1] first-time interview, [2] recurring interview with the same interviewer as in the previous round, and [3] recurring interview with a different interviewer compared to the previous round.

B. Macro variables

GDP per capita

GDP numbers are taken from multiple sources. To make numbers consistent across sources, we first build an annual growth series for GDP per capita in constant prices. In 98 percent of our sample, we use the Conference Board Total Economy Database (TED, 2015), from which we extract the growth rate of PPP-adjusted GDP per capita in 1990 international dollars between 1960 and 2014. Missing values are replaced with real growth rates obtained from the Maddison Project (2013) and the World Development Indicators (WDI, 2016). The former source employs the same definition of GDP per capita as in TED (2015), while the latter source reports PPP-adjusted real GDP per capita in constant 2011 international dollars.[1]

For some countries that split apart (e.g., Czechoslovakia, Yugoslavia), the Maddison Project publishes the growth series for country parts before the breakup. However, GDP per capita is not available in any source for ex-USSR republics before 1980. Since some immigrants came to Germany from the former Soviet Union before 1980, we use real wage growth instead of GDP per capita growth for the Soviet republics between 1960 and 1980. Real wage growth is obtained from inflation-adjusted monthly wage series reported by the Central Statistical Board of the USSR.

The above four sources provide a complete time series on real growth of GDP per capita for all countries in GSOEP sample between 1960 and 2014. By using this growth series and the PPP-adjusted GDP per capita values in 2011 as a baseline (WDI, 2016), we construct a time-series of PPP-adjusted GDP per capita in constant 2011 international dollars.

Political instability

We capture political instability in a home country by using the dataset on Major Episodes of Political Violence (1946-2014) published by the Center for Systemic Peace (2015). This dataset assigns an integer score between 0 and 10 to each major episode of a war for independence, international violence/warfare, civil violence/warfare, and ethnic violence/warfare, where 0 indicates no episodes of political violence, 1 denotes sporadic political violence, and 10 stands for extermination and annihilation. All these scores are summed up into a combined index of political violence, which in our sample varies from 0 (74 percent of all immigrants) to 14 (Iraq in 1986). The original source does not provide scores for parts of former unified countries. Since many immigrants came from the former Soviet Union and ex-Yugoslavia, we use a variety of web sources to create the index of political violence for each republic before the breakup.

This variable is highly skewed, with only 2 percent of immigrants coming from countries with the index higher than 4. Instead of treating it as a continuous variable, we aggregate scores into four distinct categories: 0="no episodes of political violence", 1 or 2="limited political violence", 3="serious political violence", 4 and above="warfare". In the category of limited political violence, events are confined to short periods or specific areas; some population dislocation may occur; attributable deaths are up to ten thousand. Some examples from our sample include Czech Republic 1968, Turkey 1981-1983, Russia 1990, and China 1998. In the category of serious political violence, events are longer and

[1] A simple coefficient of correlation between the TED and Maddison series of per capita GDP growth is 0.92 and between the TED and WDI series is 0.91.

involve a limited use of destruction technologies; population dislocations are in the tens of thousands; attributable deaths range from ten to fifty thousand. Examples include Syria 1973, Croatia 1992-1995, Tajikistan 1993-1995, and Kosovo 1996-1999. In the last category of warfare, events involve a broad use of destruction technologies and large dislocations of people; attributable deaths exceed 50,000 people. Examples include Afghanistan 1978-2001, Iran-Iraq 1980-1988, Armenia-Azerbaijan 1991-1994, Bosnia and Herzegovina 1992-1995, and Syria 2011 to present.

Linguistic distance

See Section C below on the construction of linguistic distance between the primary language(s) of home countries and German language.

Geographic distance

The geographic distances between capitals of the home country and Germany are taken from the GeoDist database made available by CEPII (2016) and described in Mayer and Zignago (2011). Geodesic distances in the dataset are calculated using the great circle formula and geographic coordinates of the capital centers.

World region of home country

We use six broad categories of world regions: [1] Africa, [2] America, [3] Asia and Pacific, [4] countries of former USSR, [5] Eastern Europe, and [6] Western Europe. Turkey is included in the Asia category, while the Baltic countries are classified as Eastern Europe. The rest is self-explanatory. Africa is the base for comparison.

Guest worker treaty

A binary indicator that equals 1 if the home country has a guest worker treaty with West Germany in a given year.

Time trend

Year of survey minus 1960.

Unified Germany

This dummy variable equals 1 if the survey year is higher than or equal to 1990.

C. Construction of linguistic distance

The Levenshtein linguistic distance between the primary language(s) of home countries and Standard German is constructed using the Automated Similarity Judgement Program (ASJP) provided by the German Max Planck Institute for Evolutionary

Anthropology (Wichmann *et al.*, 2016).[2] The ASJP program uses a list of 40 words for all languages to calculate the distance matrix. The words selected in the list have no cultural context and are present in all languages. These words are first transcribed into a standardized orthography, which uses only the symbols from QWERTY keyboard and has 7 vowel symbols and 34 consonant symbols.

Once the words are transcribed, the calculation of the Levenshtein linguistic distance involves three steps (Bakker et al., 2009):

1. For each word pair of 40 words, calculate the minimum number of consecutive additions, deletions or substitutions of a symbol necessary to convert one word into another.
2. Divide the above number of successive changes by the length of the longest of compared words to account for the differences in word length. This produces normalized linguistic distance (LDN).
3. Take the average LDN across all word pairs and then divide the average LDN by the average LDN of N(N-1)/2 word pairs with different meaning. This second normalization is intended to correct for chance resemblances due to overlap in phoneme inventories or shared phonotactic preferences in the two languages.

This measure is continuous, and it provides variation between languages even if they belong to the same language families. The table below shows languages with highest and lowest linguistic distance from Standard German in GSOEP.

Highest		Lowest	
Language	LD (ASJP)	Language	LD (ASJP)
Korean	1.0468	Luxembourgish	0.4083
Palestinian Arabic	1.0332	Dutch	0.4883
Malay	1.03	Afrikaans	0.595
Arabic Gulf Spoken	1.024	Norwegian Bokmaal	0.6438
Maltese	1.0227	Swedish	0.6979

For the primary language, we chose either official language or the most spoken language in countries with multiple official languages. For example, we chose Hindi for India even though English is a second official language. The information on the number of people who speaks each language by country is provided in Ethnologue (2016). If the country does not have a dominant language (e.g., there are two equally spoken languages), then the linguistic proximity is calculated for each language separately, and the final score is averaged (some examples include Chad, Cyprus, Kenya, Switzerland, etc.).

[2] Refer to http://asjp.clld.org/ for more information.

D. References for data sources

Bakker, Dik, André Müller, Viveka Velupillai, Søren Wichmann, Cecil H. Brown, Pamela Brown, Dmitry Egorov, Robert Mailhammer, Anthony Grant & Eric W. Holman, 2009. "Adding Typology to Lexicostatistics: a Combined Approach to Language Classification". *Linguistic Typology* 13: 167–179.

Bollinger, Christopher R. and Barry T. Hirsch, 2006. "Match Bias from Earnings Imputation in the Current Population Survey: The Case of Imperfect Matching," *Journal of Labor Economics* 24(3): 483-519.

CEPII, 2016. GeoDist Distance Measures, http://www.cepii.fr/

Center for Systemic Peace, 2015. *Major Episodes of Political Violence, 1946-2014*, http://www.systemicpeace.org/inscrdata.html

Ethnologue, 2016. *Languages of the World*, 19th edition http://www.ethnologue.com

Maddison-Project, 2013 version, http://www.ggdc.net/maddison/maddison-project/home.htm

Mayer, Thierry and Soledad Zignago, 2011 "Notes on CEPII's Distances Measures: The GeoDist Database," *CEPII Working Paper 2011-25*

OECD, 2016. OECD.Stat (database). http://stats.oecd.org/

SOEP, 2014a. *Documentation on Biography and Life History Data for SOEP v31 and v31.1*, edited by Jan Goebel. SOEP Survey Papers 312: Series D. Berlin: DIW/SOEP

SOEP, 2014b *Documentation of Person-related Status and Generated Variables in $PGEN for SOEP v31.1*. SOEP Survey Papers 307: Series D. Berlin: DIW Berlin/SOEP

TED, 2015. The Conference Board Total Economy Database, May 2015, http://www.conference-board.org/data/economydatabase/

Wichmann, Søren, Eric W. Holman, and Cecil H. Brown (eds.), 2016. The ASJP Database (version 17).

WDI, 2016. World Development Indicators, World Bank, http://data.worldbank.org/

Appendix A4. Supplementary Tables

Table A4-1: Timing of Migration Equation, Full Sample

	Cox	***Weibull***	***Gompertz***	***Weibull Frailty***	***Gompertz Frailty***
Female	-0.022	-0.045*	-0.051**	-0.054**	-0.075***
	(0.024)	(0.025)	(0.025)	(0.025)	(0.026)
Years of formal schooling	-0.008	-0.014***	0.004	-0.022***	0.001
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Years of job training	0.025***	0.015*	0.037***	0.012	0.040***
	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)
Years of work experience	-0.006***	-0.013***	-0.017***	-0.015***	-0.026***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Ethnic German	0.334***	0.332***	0.326***	0.370***	0.364***
	(0.033)	(0.034)	(0.035)	(0.040)	(0.041)
Place of upbringing					
Large or medium city	-0.047	-0.037	-0.052	-0.019	-0.034
	(0.029)	(0.031)	(0.031)	(0.031)	(0.032)
Small city	0.013	0.022	0.017	0.028	0.023
	(0.032)	(0.033)	(0.033)	(0.035)	(0.036)
Parents' education					
General sec and upper vocational	-0.200***	-0.187***	-0.202***	-0.227***	-0.251***
	(0.030)	(0.032)	(0.032)	(0.033)	(0.034)
Higher education	-0.183***	-0.159***	-0.189***	-0.198***	-0.248***
	(0.041)	(0.042)	(0.042)	(0.046)	(0.047)
GDP growth in Germany	0.049***	0.049***	0.049***	0.045***	0.044***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Log GDP per capita in home country	-0.261***	-0.265***	-0.245***	-0.291***	-0.271***
	(0.028)	(0.029)	(0.030)	(0.029)	(0.029)
Instability in home country					
Limited political violence	0.489***	0.502***	0.473***	0.511***	0.479***
	(0.039)	(0.040)	(0.040)	(0.044)	(0.045)
Serious political violence	0.431***	0.448***	0.429***	0.450***	0.435***
	(0.052)	(0.054)	(0.053)	(0.057)	(0.058)
Warfare	0.493***	0.494***	0.461***	0.534***	0.498***
	(0.038)	(0.039)	(0.039)	(0.042)	(0.043)
Linguistic distance	-0.386***	-0.387***	-0.355**	-0.431***	-0.397***
	(0.136)	(0.139)	(0.142)	(0.146)	(0.150)
Log of distance between capitals	-0.111***	-0.110***	-0.082***	-0.130***	-0.094***
	(0.026)	(0.027)	(0.028)	(0.033)	(0.034)
Home country's region					
America	0.578***	0.591***	0.528***	0.670***	0.599***
	(0.112)	(0.117)	(0.118)	(0.138)	(0.142)
Asia	0.373***	0.367***	0.379***	0.433***	0.457***
	(0.077)	(0.082)	(0.081)	(0.092)	(0.094)
Ex-USSR	-0.132*	-0.130	-0.140*	-0.161*	-0.183*
	(0.077)	(0.081)	(0.082)	(0.095)	(0.097)
Eastern Europe	0.327***	0.331***	0.345***	0.362***	0.389***
	(0.081)	(0.086)	(0.086)	(0.099)	(0.101)
Western Europe	0.383***	0.401***	0.374***	0.519***	0.518***

	(0.089)	(0.094)	(0.095)	(0.105)	(0.107)
Guest worker treaty (dummy)	1.328***	1.302***	1.337***	1.455***	1.513***
	(0.060)	(0.061)	(0.062)	(0.053)	(0.054)
Linear trend, 1960=1	0.074***	0.072***	0.070***	0.081***	0.078***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Linear trend squared / 100	-0.001	0.000	0.003	0.000	0.006
	(0.006)	(0.007)	(0.006)	(0.007)	(0.008)
Unified Germany	-0.143***	-0.151***	-0.130***	-0.185***	-0.162***
	(0.041)	(0.041)	(0.041)	(0.046)	(0.046)
Intercept		-2.945***	-2.468***	-2.937***	-2.421***
		(0.397)	(0.404)	(0.429)	(0.439)
Test for presence of heterogeneity					
$\hat{\theta}$ (frailty variance)	...	...	...	0.154	0.213
χ^2 (likelihood ratio test of $\hat{\theta} = 0$)	...	...	...	67.244	102.922
p-value	...	...	...	0.000	0.000

Notes: Table presents estimates of the proportional hazard model that corresponds to Equation (8). Cox, Weibull, and Gompertz denote distributional assumptions of the baseline hazard. The estimates in the first three columns do not account for individual unobserved heterogeneity (or frailty), while estimates in the last two columns do. Individual-specific random effects in the last two columns are assumed to have inverse Gaussian distribution. The shared-frailty models are estimated using streg with the frailty option in Stata.

N of observations=108,779; *N* of immigrants=8,288. Compared to Table 3, the sample in this table also includes immigrants who had never reported positive wage in GSOEP. Standard errors are in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Robust standard errors are only allowed and reported for the models without frailty.

Base/omitted categories are rural area for the place of upbringing, “basic secondary and lower vocational” for parents’ education, “no episodes of political violence” for instability in home country, Africa for home country’s region. Unknown type of place of upbringing and unknown parents’ education are also included in the estimates but not shown here.

Table A4-2: Summary Statistics, Selection Variables

Variables	*Sample C*
Reports wage	0.615
Remains in the panel in future rounds	0.903
Mode of interview	
Self-written and mailed	0.057
Computer-assisted	0.251
Interviewer	
First interview	0.163
Different interviewer	0.088
Predicted 3-year economic growth in home country	2.100
	(3.688)
Unemployment rate among natives	5.059
	(2.116)
N of observations	46,713

Notes: Table shows the mean and standard deviation of the variables used in two selection equations. Standard deviations are in parenthesis and not reported for dummy variables. Omitted/base categories are face-to-face interviews for the mode of interview and the same interviewer