Timing of Migration, Immigrant Quality and Labor Market Assimilation: Evidence from a Long Panel in Germany*

Apoorva Jain†

Job Market Paper

For the most current version, check my webpage.

Thursday 10th November, 2016

Abstract

This study shows that it is important when estimating immigrants' rate of economic assimilation to account for the timing of migration. The length of stay in the host country, which in the existing literature is commonly assumed to be exogenous, depends on the timing of migration. The optimum timing of migration is based on net expected lifetime earnings, and, thus, the length of stay also is endogenously determined. This paper models and estimates jointly the timing of migration and wage assimilation equations, which is a first in the literature. From German Socio-Economic Panel data that dates from 1984 to 2014, I estimate the individual-specific rate of assimilation while accounting for unobserved immigrant quality and an unobserved propensity to migrate early. Estimates from the joint model reveal four key findings. First, the average rate of assimilation estimated under the exogeneity assumption is biased upwards. Second, the average rate of assimilation hides significant variation in assimilation rates among immigrants who are of different quality. Third, immigrants of low quality have a faster rate of assimilation than their high quality counterparts. Fourth, immigrants who have a high propensity to migrate early have a higher individual rate of assimilation. The joint model allows me to find the interdependence between the timing of migration and the labor market assimilation of immigrants - mechanisms that until now had been assumed in the literature to be independent.

JEL Codes: J24, J31, J61, N30

Keywords: Wage assimilation, Host country specific human capital, Skill-transferability, Immigrant Labor

*I am grateful to my advisor Klara Peter and the committee members Luca Flabbi, Tiago Pires, Helen Tauchen and Charles Becker for their valuable feedback. This paper has also benefited from comments and suggestions received during the Applied Micro Workshop at UNC-Chapel Hill, 85th Annual Conference of the Southern Economic Association and, 91st Annual Conference of the Western Economic Association International. This project was funded by the Williamson Family Excellence Fund, Graduate and Student Professional Federation Travel Grant and, Graduate Student Travel Grant.

†Department of Economics, University of North Carolina-Chapel Hill

1 Introduction

The immigrant assimilation hypothesis (Chiswick, 1978) conjectures that immigrants acquire host country-specific human capital, that this increases with time spent in the host country (henceforth called the length of stay), and that they experience wage growth. The rate of wage growth with respect to the length of stay is known as the rate of assimilation. Existing studies, such as Borjas (1987, 1994), Hu (2000) and Lubotsky (2007), which have assumed that the length of stay is exogenous, have estimated an *average* rate of assimilation.[1] The length of stay depends on the timing of migration, and, thus, it is endogenously determined. This paper relaxes the exogeneity assumption by developing and estimating a joint model of the timing of migration and assimilation in the labor market. The joint model also provides individual-specific rates of assimilation. These individual specific rates can be used to better inform immigration policies, which until now were based solely on the quality of the immigrant at the time of migration and not her future ability to assimilate well.

Forward-looking individuals decide whether or not to migrate on the basis of the net expected utility of migration; in other words, the optimum timing of migration is a choice. Consequently, the length of stay, which is age minus age at migration, is not exogenous. Ignoring the selective timing of migration can lead to an inconsistent estimate of the rate of assimilation. Moreover, the commonly estimated average rate of assimilation neglects the potential differences in the post-migration rate of human capital acquisition between immigrants who migrated at different ages. According to the economic theory of human capital, younger individuals have a higher incentive to acquire human capital. Given that differences in human capital investment result in different rates of assimilation, a twenty year-old and a forty year-old immigrant are likely to have different rates of assimilation.

It is necessary to jointly estimate the timing of migration and economic assimilation to account for unobserved individual factors that affect both the timing of migration and the immigrants performance in the host countrys labor market. Unobserved characteristics, such as risk attitude, personality, and ability, can affect the propensity to migrate and earnings growth in the host country. For instance, immigrants who have a high ability might have a lower cost of migration and, thus, a higher propensity to migrate early. High ability individuals are likely to experience high wage growth after migration.

[1] Rarely, papers estimate assimilation rates for subgroups based on arrival cohorts such as Borjas (2013) and Fertig and Schurer (2007).

In contrast, risk-averse individuals might have a low propensity to migrate early, and they might avoid risky yet profitable job opportunities in the host country. In other words, risk-averse immigrants might have a slower rate of assimilation. In such cases, the rate of assimilation is correlated with both an unobserved propensity to migrate and the timing of migration. To account for the interdependence of the timing of migration and wage assimilation, the two processes should be estimated jointly.

Most papers that estimate the wage assimilation equation, such as Borjas (1987, 1994, 1988) and Antecol et al. (2006), use census data to estimate an average rate of assimilation while controlling for arrival-cohort-specific unobserved immigrant quality.[2] Thus, they assume that immigrants within an arrival cohort have similar unobserved characteristics. A few papers, including Fertig and Schurer (2007) and Cobb-Clark et al. (2012), use longitudinal data and individual fixed effects to account for time-constant individual unobserved heterogeneity. However, both longitudinal and cross-sectional studies implicitly assume that unobserved immigrant quality affects wage level but not wage growth - i.e., they assume that rate of assimilation does not vary with immigrant quality.[3] This paper relaxes these assumptions to estimate individual-specific rates of assimilation that vary with immigrant quality.

While wage assimilation estimates are common in the literature, few papers estimate the timing of the migration equation. The papers that do are limited by the fact that they focus on the effect of a single factor on out-migration and analyze either domestic migration or migration from a single country over a short period of time. For instance, Reed et al. (2010) examine gender differences in mobility in Ghana; Henry et al. (2004) analyze the effect of rainfall on first out-migration in Burkin Faso; Ezra and Kiros (2001) study the effect of drought on rural out-migration within Ethiopia; and Hare (1999) analyzes rural out-migration within China. In contrast, my study examines data on a much larger scale: I estimate the timing of the migration equation for immigrants from over 100 countries during a 53-year period (1961-2014).

To address these shortcomings in the existing models, I develop and estimate a joint

[2] A few papers like Cobb-Clark (1993) account for unobserved immigrant quality using a control function approach by including macro variables indicating socio-economic development at the country of origin.

[3] Borjas (2013) and Fertig and Schurer (2007) are notable exceptions which estimate rates of assimilation for different arrival cohorts by interacting the length of stay variable with the indicator variable for the immigrant's arrival cohort.

model of wage assimilation and the timing of migration. The joint model links the two equations through a correlation between immigrant quality, the individual-specific rate of assimilation (both of which appear in the wage assimilation equation), and the unobserved propensity to migrate early (which appears in the timing of migration equation). The timing of migration equation is a continuous time parametric proportional hazard model in which the hazard of early migration depends on individual characteristics, macro-level factors of migration, and an unobserved individual propensity to migrate early. The wage assimilation equation is a linear mixed model in which the log of the wage depends on various individual-specific factors, including the length of stay and unobserved immigrant quality. The individual-specific rate of assimilation is estimated using a random coefficient on the length of stay variable. Using the parameters estimated in the joint model, I estimate the complete distributions of the individual-specific rate of assimilation, immigrant quality, and the propensity to migrate early. In addition to providing complete distributions, my joint model advances scholarly understanding of the relationship between these components.

I estimate the proposed joint-model using data on immigrants in the German Socio-Economic Panel (henceforth called SOEP) for the period 1984-2014.[4] Since the 1950s, Germany has had a long and diverse history of immigration, and for this reason it provides an excellent location to study immigrants economic assimilation. As of 2015, Germany has hosted more than 12 million immigrants, which is the second highest stock of immigrants in the world. Moreover, for a dense sample of immigrants the SOEP provides information on the country of origin and the year of migration. Using this information, I construct pre-migration histories from which I estimate the timing of the migration equation. To construct pre-migration histories, I collect data on macro-level migration factors from 1961 to 2014, and these are then merged with individual level pre-migration characteristics using the year of migration and country of origin. These long panel data allow me to estimate long-term assimilation rates, which in the literature is a rare achievement.

The joint model is estimated for individuals who migrated after the age of 13 between 1961 and 2014. I limit the sample to youth and adult migrants because: (1) child migrants are not likely to make individual migration decisions and (2) the assimilation experience of child migrants could differ from that of youth and adult migrants. For instance, Bleakley and Chin (2004, 2008, 2010) show that child migrants assimilate economically and socially better than their adult migrant counterparts.

[4] The data used in this paper was made available to us by the German Socio-Economic Panel Study at the German Institute for Economic Research (DIW Berlin), Berlin(Wagner et al., 2007)

The model estimates reveal four key findings. First, the exogeneity assumption of the length of stay results in an upward bias in the average rate of assimilation. After accounting for the selective timing of migration, the average rate of assimilation drops from 1 percent to 0.6 percent. Second, the individual-specific rates of assimilation vary significantly- that is, by 0.021 standard deviation points, or 2.1 percent. Third, the estimates predict a strong negative correlation between the individual-specific rate of assimilation and immigrant quality. This suggests that relative to high quality immigrants, low quality immigrants invest more in human capital after migration and, consequently, they have a higher rate of assimilation. Thus, we observe a catch-up effect between low-quality and high-quality immigrants. These findings are consistent with the theoretical predictions of the Immigrant Human Capital Investment (IHCI) model devised by Duleep and Regets (1999). The model predicts that immigrants who have less transferable skills would have a lower opportunity cost of acquiring human capital in the host country. Thus, immigrants with less transferable skills are more likely to invest in human capital after migration and, thus, they have a higher rate of assimilation. Immigrant quality, in other words, is defined by both ability and the degree of skill-transferability.

Finally, the estimates show a positive correlation between the propensity to migrate early and the individual-specific rate of assimilation. This implies that immigrants who have a higher propensity to migrate early, and, hence, migrate at an early age, invest more in human capital post-migration. It also suggests that expected growth in earnings influences the propensity for early migration and, subsequently, the timing of migration. This, in turn, suggest that the timing of migration and wage assimilation are interdependent - a finding that validates the paper's hypothesis and highlights the need to estimate these equations jointly.

The rest of the paper is organized in five sections. Section 2 presents a simple theoretical model of the timing of migration that explains how individuals decide the optimum period of working life to spend in the host-country so as to maximize expected lifetime earnings. Section 3 illustrates the endogeneity problem in the length of stay variable and develops the joint model of the timing of migration and wage assimilation. In Section 4, I discuss the data and variables used to estimate the joint model. Section 5 discusses model estimates. Finally, Section 6 concludes.

2 Theoretical Model of Timing of Migration

In this section, I present a simple model that explains how the length of working life spent in the host country is determined by the timing of migration. In the model, individuals decide the optimal time of migration that maximizes the net expected lifetime earnings. The migration event is assumed to be an absorbing state, that is, once the individual migrates to the host country, she does not out-migrate until the end of the working life.

A related model was presented by Zimmermann and Constant (2012) that illustrates the role of age in migration decision. My model differs from their model in two major ways. Firstly, I introduce the role of skill-transferability in the migration decision where the degree of skill transferability varies by the age at migration, the time spent in the host country and other exogenous factors. Thus, we gain insight into how differences in skill-transferability affect the marginal cost of migration and the optimum time of migration. Secondly, as the model involves maximization of lifetime earnings (and not utility), I only focus on working age individuals and working life period.

Model Set-up The individual begins working in the origin country at age a_b which is also the first time she decides whether to migrate or not. The working life spans from a_b to A. The individual works in the home country from a_b till the time of migration a_m and in the host country from a_m till A. Thus, the total working life spent in the host country is $A - a_m$. The average wage per unit of human capital H in the origin country is w_o and in the host country is w_h. Wages in both origin and host country are a function of individual's age.

As skills are not perfectly transferable over international borders, individuals can only market a fraction of their pre-migration skills δH in the host country's labor market where $0 < \delta \leq 1$. The degree of skill transferability δ varies with the age at migration a_m, the time in the host country $a - a_m$ and other exogenous factors γ. Country of origin, ethnicity and other exogenous factors captured by γ only contribute an additive shift in the degree of skill transferability and their effect does not change with the time of migration. Migration involves a one time cost C that varies by age at migration.

The present value of net lifetime earnings for an individual who migrates at a_m and discounts future earnings by ρ is given by the following expression:

$$\begin{aligned} E(a_m) = &\int_{a_b}^{a_m} e^{-\rho a} w_o(a) H \mathrm{d}a - C(a_m) e^{-\rho a_m} \\ &+ \int_{a_m}^{A} e^{-\rho a} w_h(a) \delta(a_m, a - a_m, \gamma) H \mathrm{d}a \end{aligned} \tag{1}$$

$$\text{where} \qquad \delta(a_m, a - a_m, \gamma) = \delta_0(a_m, a - a_m) + \alpha\gamma$$

The first term represents the discounted earnings in the origin country from a_b to a_m. The second term is the one-time cost of migration at age a_m and the third term is the discounted earnings in the host country from a_m to A.

The optimal time of migration a_m^* is given by equating the first derivative of Equation 1 with respect to a_m to zero:

$$\begin{aligned} E_1(a_m) = &e^{-\rho a} w_o(a) H + \rho C(a_m) e^{-\rho a_m} - C'(a_m) e^{-\rho a_m} \\ &- e^{-\rho a_m} w_h(a_m) \delta(a_m, 0, \gamma) H + \int_{a_m}^{A} e^{-\rho a} w_h(a)(\delta_1 - \delta_2) H \mathrm{d}a = 0^{5} \end{aligned} \tag{2}$$

where the left hand side gives the marginal benefit and the right hand side gives the marginal cost of migrating a year later. Rearranging Equation 2 yields the following expression:

$$\begin{aligned} &e^{-\rho a} w_o(a) H + \rho C(a_m) e^{-\rho a_m} - C'(a_m) e^{-\rho a_m} \\ &= e^{-\rho a_m} w_h(a_m) \delta(a_m, 0, \gamma) H - \int_{a_m}^{A} e^{-\rho a} w_h(a)(\delta_1 - \delta_2) H \mathrm{d}a \end{aligned} \tag{3}$$

The marginal benefit includes the discounted wage in the origin country for an additional year and the postponed cost of migration minus the change in cost of migration due to the delay. There are two reasons why we might expect the change in cost $C'(a_m)$ to be negative i.e. the cost of migration decreases with age at migration. Firs, selective immigration policies that favor high-skilled immigrants make it easier for older immigrants to obtain a work visa. Second, it might also be easier for older individuals to collect information about migration process and job opportunities in the host country. Under such cases, delaying

[5] δ_1 and δ_2 represent the derivative of δ with respect to a_m and $a - a_m$, respectively.

migration would increase the marginal benefit from reduced cost of migration a year later.

The marginal cost includes the lost earnings in the host country at arrival (i.e. when the length of stay is zero) minus the change in the future stream of earnings in the host country due to migrating a year later. It is assumed that $\delta_1 < 0$ i.e. the degree of skill transferability decreases with an increase in age at migration and $\delta_2 > 0$ i.e. the degree of skill transferability increases with time spent in the host country. Thus, with increase in age at migration, the change in future stream of earnings in the host country decreases and the marginal cost increases.

There is substantial evidence that suggests δ_1 is negative. For instance, Bleakley and Chin (2004, 2008, 2010) show that younger aged migrants are more proficient in host-country's language, thus they perform better in the host-country's labor market and are more socially assimilated than their older counterparts.[6] Similarly, Immigrant Assimilation Hypothesis suggests that δ_2 is positive. Duleep and Regets (1999) show that degree of skill transferability increases with the investment in host-country specific human capital. As investment in human capital post-migration depends on the time spent in the host country, skill transferability is expected to increase with time in the host country.

According to Equation 3, the optimal age at migration a_m^* is chosen when the marginal benefit equals the marginal cost of migrating. Migration does not occur if the marginal benefit from delaying migration is always higher than the marginal cost. On the other hand, if the marginal cost is always higher than the marginal benefit, the individual would choose to migrate at the beginning of working life a_b. The existence of an interior solution depends on the discounting factor, the magnitude of wage loss due to skill transferability and the change in the cost of migration from postponed migration.

Exogenous factors of skill transferability would also affect the optimal time of migration through differences in earnings at arrival. Let γ capture cultural and linguistic similarity between the origin and host country. Thus, the degree of skill transferability increases with an in increase in γ i.e. $\delta_\gamma > 0$. To understand the effect of γ on a_m^*, we take a derivative of Equation 3 with respect to γ:

[6] They also asserted that these findings support Critical Period Hypothesis of language acquisition. Critical Period Hypothesis suggests that early ages are more suitable for language acquisition. So, younger individuals can acquire a new language with less effort and earlier (less time-cost) compared to older individuals.

$$E_{11}(a_m^*(\gamma),\gamma)\frac{\partial a_m^*}{\partial \gamma} + E_{1\gamma}(a_m^*(\gamma),\gamma) = 0 \tag{4}$$

where

$$E_{1\gamma}(a_m^*(\gamma),\gamma) = -e^{-\rho a_m} w_h(a_m)\delta_\gamma H \tag{5}$$

As $E_{11} < 0$, $E_{1\gamma} < 0$ and $\delta_\gamma = \alpha$, this implies

$$\frac{\partial a_m^*}{\partial \gamma} < 0 \tag{6}$$

Thus, the model predicts that migrants from countries similar to the host country would migrate at an earlier age.

From the theoretical model, it is clear that the time of migration is not randomly chosen and hence the length of stay in the host country is also not exogenous. In the next section, I illustrate how the failure to account for selective timing of migration can lead to an inconsistent estimate of the rate of wage assimilation.

3 Joint Model of Wages and Timing of Migration

This section first discusses the implicit assumptions made when the length of stay is treated as an exogenous variable. Next, it develops a joint-model of the timing of migration and wage assimilation that relaxes these assumptions. The joint-model also accounts for selection in unemployment using inverse propensity weighting.

3.1 Problem of Endogeneity in Length of Stay

A typical empirical model of the economic assimilation of immigrants (refer to Chiswick (1978), Borjas (1985, 1987) and Duleep and Regets (2002)) is estimated using the following wage equation :

$$W_{is} = \beta_0 + \delta LOS_{is} + \beta_X X_{is} + \phi(s) + C_i + \tilde{\epsilon_{is}} \tag{7}$$

where W_{is} is the log of wage of individual i at time s, X_{is} is a vector of immigrant's observed characteristics in the host country that often includes age (or experience) and education,

LOS_{is} is the length of stay calculated as the difference between the year of survey and the year of migration $(Y_s - Y_m)$, $\phi(s)$ is a linear time trend capturing the business cycle and C_i captures time-constant cohort-specific unobserved heterogeneity. A few panel studies like Fertig and Schurer (2007) sometimes include time-constant individual heterogeneity α_i instead of C_i.

δ is the average wage return on spending a year in the host country instead of origin country. Thus, δ 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 herself and a positive value of δ does not indicate that immigrant earnings are converging to their native counter-parts. In this paper, I follow a similar definition of assimilation but estimate individual specific rate of assimilation and not just the average.[7] I discuss the estimation of individual-specific rate of assimilation in Section 3.3.

Previous studies have treated LOS_{is} as exogenous. However, there are several reasons why this leads to biased estimates. To understand them, let us consider an individual's migration decision $M_i(a)$ at age a:

$$\begin{aligned} & M_i(a) = 1[MB_i(a) - MC_i(a) = 0] \\ \text{where} \quad MB_i(a) = f(Z_i(a), \nu_i) \quad & \text{and} \quad MC_i(a) = g(Z_i(a), \nu_i) \end{aligned} \tag{8}$$

$$\text{and} \qquad M_i(a) = 1 \implies M_i(a-1) = M_i(a-2) = ...M_i(14) = 0^{8} \tag{9}$$

In Equation 8, individual i migrates at age a if net expected earnings are maximized i.e the marginal benefit of migrating $MB_i(a)$ equals the marginal cost of migrating $MC_i(a)$ where $MB_i(a)$ and $MC_i(a)$ are functions of factors of migration $Z_i(a)$ and the unobserved propensity of migration ν_i. However, as explained in Equation 9, $M_i(a) = 1$ implies that the individual chose not to migrate at an earlier age. Thus, the migration decision at age a not only depends on net expected lifetime earnings but also on past migration decisions.

In equation 7, δ is consistent only under the following assumptions: (1) the decision to migrate at a is random and thus $Cov(LOS_{is}, \tilde{\epsilon_{is}}) = 0$ which means the unobserved propensity

[7] In Jain and Peter (2016), we consider immigrants' rate of assimilation with respect to natives using GSOEP data and find a wage divergence.

[8] I assume, the earliest age an individual decides to migrate is at the age of 14, i.e., the earliest age an individual can begin working.

of migration is uncorrelated with the error in the wage equation, i.e., $Cov(\nu_i, \tilde{\epsilon_{is}}) = 0$ or (2) the timing of migration only has a constant effect on wages (through C_i or α_i) and no effect on the wage growth through δ, i.e., $Cov(\nu_i, \alpha_i) \neq 0$ but $Cov(\nu_i, \delta) = 0$; thus estimating an individual fixed effects model solves the problem of selection bias.

The first assumption is unrealistic and contradicts the theoretical evidence provided in Section 2. The second assumption implies that timing of migration does not influence the level of human capital acquired in the host-country and subsequently does not affect the rate of assimilation. As per this assumption, a twenty year old and a thirty year old would have the same incentive to invest in host-country's human capital. In light of empirical evidence that people tend to invest more in human capital during the early period of life-cycle, this assumption is quite restrictive.

Moreover, as depicted in the theoretical model, the degree of skill-transferability is a function of age at migration and influences the stream of earnings in the host country. In fact, Schaafsma and Sweetman (2001) and Friedberg (1992) have shown that age at migration affects earnings level. This effect possibly reflects differences in post-migration education. Furthermore, if we believe that a forward-looking rational individual not only cares about the wage level but also the growth in wages (δ), the migration propensity ν_i in equation 8 would be correlated with δ in equation 7.

With this in mind, I explicitly model the timing of migration and account for selective timing of migration in estimating the rate of wage assimilation. Subsection 3.2 presents the hazard model used to estimate the timing of migration and Section 4 discusses the push-pull factors of migration included in the hazard model.

3.2 Timing of Migration

I model the timing of migration using a parametric continuous-time proportional hazard model for future immigrants, i.e., those individuals who eventually migrate.[9] Since the aim

[9] Although, the data is available in yearly intervals, I treat time as continuous as the hazard model is estimated over a long period of time, specifically from 1960 to 2013. Another reason for choosing a proportional hazard model over a discrete time logistic regression is the benefit of defining a flexible baseline hazard. However, I refrain from choosing a non-parametric baseline hazard (as is the case in traditional cox model) as parametric models perform better when data suffers from left-truncation (Hancock and Mueller, 2010).

of the paper is to obtain a consistent estimate of the rate of assimilation and the wage assimilation equation is only estimated for immigrants, the timing of migration equation is also estimated only for individuals who eventually migrate. The equation is given as:

$$\begin{aligned} &\lambda_i(t) = \lambda_0(t)\exp(X_i(t)'\beta^X + E_i(t)'\beta^E + c_i)^{10} \\ &\text{where} \qquad c_i \sim \mathcal{N}(0, \sigma_c) \end{aligned} \tag{10}$$

where $\lambda_i(t)$ is the instantaneous rate of migration *given* the individual did not migrate earlier. Thus, it captures the whole history of migration decision process as well as the conditional dependence. $X_i(t)$ is a vector of individual's observed characteristics (both time -constant and -varying) and $E_i(t)$ is a vector of country-level push-pull factors of migration. As the name suggests, push-pull factors of migration are exogenous factors of migration that push the individuals out of the origin country and pull towards the host country. In Section 4, I describe the chosen factors of migration included in $E_i(t)$ in detail.

$\lambda_0(t)$ is the baseline hazard and is assumed to be a linear function of working life period (which begins from the age of 14 to 65). c_i is the individual's unobserved propensity of early-migration and captures time-constant individual unobserved heterogeneity. As equation 10 is only estimated for future immigrants, c_i measures the unobserved propensity to migrate early versus later. A higher value of c_i would mean the individual has a higher propensity of migrating early and thus a higher rate of hazard $\lambda_i(t)$.

Equation 10 is estimated for the years 1960 - 2014. Thus, individuals who turn 14 years of age before 1960 enter late in the hazard model estimation, specifically at 1960.[11] Such delayed entry or left truncation can be an issue in estimation of shared frailty models if frailty c_i is correlated with the truncation point. In estimation of equation 11, I assume c_i to be uncorrelated with the truncation point. Given that delayed entry of individuals (for individuals who turn 14 years of age before year 1960) is only due to lack of data on migration push-pull factors for years prior to 1960, assuming no correlation between

[10]This is a generalized form of Proportional Hazard model. Notice that $E_i(t)$ and $X_i(t)$ include time-varying explanatory variables, thus the hazard ratio will not be constant over time as is the case with the traditional Proportional Hazard model. This random effects hazard model is commonly represented in the following manner:

$$\lambda_i(t) = \lambda_0(t)\eta_i \exp(X_i(t)'\beta^X + E_i(t)'\beta^E)$$

where $\eta_i = \exp(c_i)$ is the shared frailty.

[11]Entry year in hazard model estimation is given by the rule: max (year when age is fourteen, 1960).

frailty and truncation point is not restrictive. I also assume that $E(c_i, X_i(t)) = 0$ and $E(c_i, E_i(t)) = 0$ which are standard assumptions in estimation of random effects model.

Also, it is assumed that every immigrant maximizes her individual utility. This is a common assumption in empirical literature on migration. However, in case of children, the migration decision is necessarily made at household-level and not on per individual basis. Thus, I restrict the estimation sample to youth and adult immigrants, i.e., individuals who migrated after the age of 14. So, the the first time individual faces the risk of migration is at age 14. The waiting time from age 14 until the age at migration is the failure time of migration T_i and the dependent variable in equation 10.

3.3 Wage Assimilation

The wage assimilation equation estimated in this paper differs from equation 7 in three respects. Firstly, I allow the rate of assimilation to vary by individual. Secondly, I account for time-constant individual unobserved heterogeneity a_i instead of cohort-specific unobserved heterogeneity C_i. And thirdly, I allow correlation between time-constant individual unobserved heterogeneity a_i and individual specific rate of assimilation. Thus, equation 7 transforms into the following linear mixed model:

$$W_{is} = \beta_0 + (\delta + b_i)LOS_{is} + \beta_X X_{is} + \phi(s) + a_i + \epsilon_{is} \tag{11}$$

where δ is the fixed coefficient on LOS_{is} and gives the average rate of assimilation. b_i is the random slope on LOS_{is} and gives the individual-specific variation from the average rate of assimilation rate. Thus, individual i's rate of assimilation is $(\delta + b_i)$. a_i, the random intercept, captures time-constant unobserved individual heterogeneity and is correlated with b_i.

Unlike equation 7, the linear mixed model given by equation 11 can estimate fixed as well as random coefficient of independent variables, specifically LOS_{is}. It also allows the random intercept a_i to be correlated with the random slope b_i. Moreover, we can estimate the fixed coefficient on time-constant observable characteristics on wages which is not possible with fixed effects model.

The individual-specific random intercept a_i captures the heterogeneity in unobserved quality of immigrants. It allows each immigrant to have her own initial point of wage trajectory. Borjas (1987) found significant differences in initial earnings between arrival cohorts and argued that these differences reflect differences in unobserved ability between cohorts. However, a competing argument was put forward by Duleep and Regets (2002) who showed a negative correlation between cohort-specific initial earnings and assimilation rates, i.e., cohorts with higher initial earnings had slower rates of assimilation and vice versa. Thus, they suggested that initial earnings indicate the degree of skill-transferability rather than ability. As a higher degree of skill-transferability is associated with a higher opportunity cost of investing in host country specific human capital (in terms of foregone earnings), the rate of assimilation is expected to be slower for immigrants who can easily market their pre-existing skills in the host country's labor market. I discuss the interpretation of a_i in depth in Section 5 and for now refer to a_i as immigrant quality.

The random slope b_i on LOS_{is} captures unobserved differences in post-migration investment in human capital between immigrants. It allows each immigrant to have her own wage trajectory. Unobserved differences in post-migration investment in human capital can be due to differences in the level of effort or due to other unobserved individual characteristics. For instance, some immigrants might enroll in language training or employment training after migration which helps them perform better in the host-country's labor market. Thus, they receive a higher wage return on an additional year of stay in the host-country. However, a higher rate of assimilation could also imply that the immigrant has a people-friendly personality and a great deal of perseverance which helps her progress in the workplace. b_i allows to capture such differences unlike the average rate of assimilation δ.

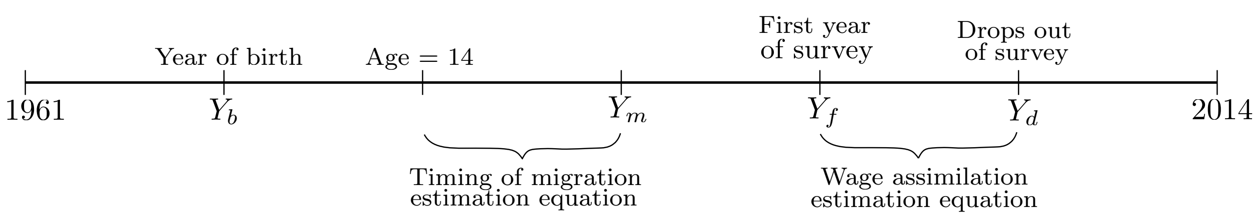
The correlation between the random slope b_i and the random intercept a_i accounts for unobserved individual characteristics which affect both the wage level and the wage growth. I do not have any prior expectation of the relationship between a_i and b_i. Their correlation can be positive or negative depending on what immigrant quality signifies. For instance, high-ability individuals would have a higher wage level and are also expected to easily acquire the host country specific human capital. Thus, we expect a positive correlation between a_i and b_i. However, if a high value of a_i means a higher degree of skill-transferability, then we expect a_i and b_i to have a negative correlation, i.e., people with high degree of skill-transferability have a lower incentive (and also a higher opportunity cost in terms of foregone earnings) to invest in human capital after migration. Regardless of the direction, we need to account for the correlation between a_i and b_i.

3.4 Joint Estimation

In this section, I present the joint model of the wage assimilation and timing of migration equations. I first explain the timeline of a typical immigrant and then develop the likelihood function of the joint distributions of wages and timing of migration $\{log(W_{is}), T_i\}$.[12]

Timeline The timeline of a typical immigrant is given in Figure 1. It specifies the portion of migrant's life cycle estimated using the timing of migration equation and wage assimilation equation. Every prospective migrant at age of 14 decides to migrate or not for the first time. Timing of migration equation is estimated from the age of 14 up till the year of migration Y_m. Wage equation is estimated for the years the individual participates in the survey. Notice that it is not necessary each migrant is surveyed in the year of arrival. Thus, estimation of wage assimilation is from the first year of survey Y_f until the migrant drops out of the survey Y_d. The length of stay is calculated as the difference between the age at survey year a_s and age at migration a_m. Note that age at survey year is a function of birth year Y_b and survey year Y_s i.e $a_s = Y_s - Y_b$. Similarly, age at migration a_m is a function of year of birth and year of migration Y_m, i.e., $a_m = Y_m - Y_b$. Thus, length of stay $a_s - a_m$ equals $(Y_s - Y_b) - (Y_m - Y_b) = Y_s - Y_m$.

Figure 1: A Typical Immigrant's Timeline



Notes: For the majority of the estimation sample, the period for which hazard model is estimated does not overlap with the period for which wage equation is estimated. The only exception are cases where the migrant is surveyed in the year of arrival, so Y_m and Y_f is the same year. Only .67 percent of the sample was surveyed in the year of migration.

[12] T_i is the waiting peiod from age 14 till migration, i.e., $T_i = a_m - 14$ where a_m is the age at migration.

Joint Likelihood Function For convenience and clarity, I reproduce the timing of migration and wage equation below with random effects a, b, and c:

$$\lambda_i(t) = \lambda_0(t)\exp(X_i(t)'\beta^X + E_i(t)'\beta^E + \mathbf{c_i}) \tag{12}$$

$$W_{is} = \beta_0 + (\delta + \mathbf{b_i})LOS_{is} + \beta_X X_{is} + \phi(s) + \mathbf{a_i} + \epsilon_{is} \tag{13}$$

In the above equations, the unobserved propensity of early-age migration c_i, quality of immigrant a_i and individual deviation from the average assimilation rate b_i are correlated. It is clear from Section 3.3 that we need to account for correlation between a_i and b_i. In a case where the propensity of early age migration c_i is uncorrelated with a_i and b_i, equation 13 can be estimated alone to get a consistent estimate of δ, b_i and a_i.

However, as seen in Section 2 and further explained in Section 3.1, forward looking individuals consider lifetime earnings when taking the migration decision, i.e., they care about both the wage level and the wage growth in the host country. Thus, it would be unrealistic to assume that the propensity of early-migration is independent of a_i and b_i. The correlation between them allows me to capture time-constant unobserved individual characteristics that affect a_i, b_i as well as c_i. For instance, a risk averse individual would have a low propensity to migrate early and would also be less likely to take risky employment opportunities or job projects in the host country. Thus, both the labor market performance in the host country and migration decision are affected by the level of risk averse nature. Similarly, innate ability of individuals could affect both the unobserved propensity to migrate early and earnings in the host country. A high ability individual might have a lower cost of migration and at the same time have higher earnings in the host country relative to her low-ability counterparts. In such cases, wages in the host country depend on the timing of migration and the random effects.

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

$$L(\theta) = \prod_{i=1}^{n} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ \prod_{s=1}^{S} f(W_{is}|T_i, a_i, b_i; \theta_w) \right\} \times f(T_i|c_i; \theta_t) f(a_i, b_i, c_i; \theta_{abc}) \mathrm{d}a_i \mathrm{d}b_i \mathrm{d}c_i \tag{14}$$

where

$$f(W_{is}|T_i, a_i, b_i;\theta_w) = (2\pi\sigma_\epsilon^2)^{-1/2} \times \exp\{\frac{-(W_{is} - \beta_0 - \beta_X X_{is} - (\delta + b_i)LOS_{is} - \phi(s) - a_i)^2\}}{2\sigma_\epsilon^2} \tag{15}$$

$$f(a_,b_ic_i;\theta_{abc}) = ((2\pi)^3|\Sigma_{abc}|)^{-1/2} \exp\{-\frac{1}{2}(a_i\ b_i\ c_i)'\Sigma_{abc}{}^{-1} \begin{pmatrix} a_i \\ b_i \\ c_i \end{pmatrix}\} \tag{16}$$

$$f(T_i|c_i;\theta_t) = [\lambda_0(T_i) \exp(X_i(T_i)'\beta^X + E_i(T_i)'\beta^E + c_i)] \times \exp\{-\int_0^{T_i} \lambda_0(u) \exp(X_i(u)'\beta^X + E_i(u)'\beta^E + c_i)\ \mathrm{d}u\} \tag{17}$$

$f(W_{is}|T_i, a_i, b_i;\theta_w)$ is the probability density function of wages in the host country conditional on the timing of migration T_i and random effects a_i, b_i. As LOS_{is} is a linear function of T_i ($LOS_{is} = Age_{is} - 14 - T_i$), wages in the host country depend on the timing of migration. However, the conditional distribution of wages and the conditional distribution of timing of migration are independent. This function can also be modified to include inverse propensity weights to estimate weighted least square estimates. I include inverse propensity weights for selection in employment in one of the specifications. I discuss in detail how the weights are calculated in Section 3.5.

$f(T_i|c_i;\theta_t)$ is the likelihood of the hazard model where T_i is the failure time. The second expression in equation 17 $\exp\{-\int_0^{T_i}\lambda_0(u) \exp(X_i(u)'\beta^X + E_i(u)'\beta^E + c_i)\ \mathrm{d}u\}$ is the survival function from age 14 up till the age before migration and the first expression in square brackets $\lambda_0(T_i) \exp(X_i(T_i)'\beta^X + E_i(T_i)'\beta^E + c_i)$ is the hazard function at the failure time, i.e., T_i. $f(a_i, b_i, c_i; \theta_{abc})$ is the multivariate normal density for the correlated random effects. $\theta_{abc}, \theta_t, \theta_w$ denote parameters for random effects covariance matrix, timing of migration equation and wage assimilation equation respectively.

I estimate $\theta_{abc}, \theta_t, \theta_w$ of equation 14 using maximum likelihood estimation. I then use these parameters to calculate Best Linear Unbiased Predictions (BLUPs) of a_i, b_i, c_i at a per individual basis and recover their complete distributions. The exogenous factors of migration included in equation 12 serve as exclusion restrictions for identification of parameters. I discuss these factors in detail in the next section. Table 2 also illustrates the key variables included in each equation.

3.5 Correcting Bias from Selection into Employment

Wage assimilation equation and subsequently the joint likelihood function can only be estimated for individuals who are employed and report positive wages. To account for non-random selection in employment, I utilize two approaches. In the first approach, I exploit the fact that unemployment in working-age male immigrants is relatively low compared to their female counterparts and estimate the model only for males where the employment selection is not a severe issue.[13]

In the second approach, I estimate the joint-model for both males and females while using inverse probability weights (IPW) in the wage equation. The weights are calculated from estimating the following Probit equation:

$$Pr(S_{it} = 1 | X_{it}, Z_{it}) = Pr(\epsilon > -\alpha_0 - \alpha_1 X_{it} - \alpha_2 Z_{it}) \tag{18}$$

where

$$S_{it} = \begin{cases} 1[S^*_{it} > 0] \\ 0[S^*_{it} < 0] \end{cases}$$

and

$$S^*_{it} = \alpha_0 + \alpha_1 X_{it} + \alpha_2 Z_{it} + \epsilon; \qquad \epsilon \sim \mathcal{N}(0, 1)$$

S^*_{it} is the latent variable which represents the utility from employment for individual i. Thus, when $S^*_{it} > 0$, the individual is employed and reports a positive wage, i.e., $S_{it} = 1$. X_{it} is a vector of all individual characteristics included in the wage equation. Z_{it} is the exclusion restriction that affects the decision of employment but not the individual's wage. I discuss the exclusion restriction for selection into employment in Section 4.2.

[13]Out of a total of 8,000 migrants, 47.6 percent are men, of which 85 percent report positive wages. The same number for women is only 65 percent.

4 Data and Variables

The joint model explained in the previous section is estimated using the data on immigrants in German Socio-Economic Panel. It is the longest-running panel of private households and persons in the Federal Republic of Germany. It is collected and distributed by the German Institute for Economic Research, DIW Berlin. The survey began in year 1984 and consists of 31 waves so far. After the fall of Berlin wall in June 1990, residents of German Democratic Republic were also included in the target population. Thus, Germany represents Federal Republic of Germany, commonly called West Germany, from 1984 - 1989 and unified Germany from 1990 - 2014. In total, there are 15 samples and each sample was created using multistage random sampling clustered by region-level. As SOEP over-samples immigrants, it is one of the very few panel dataset that can be used in migration studies.

I next discuss the selection of final estimation sample, the creation of pre-migration histories to estimate the timing of migration equation, and the variables included in wage assimilation and employment selection equation.

Sample Selection The major share of the estimation sample is drawn from three samples of SOEP, specifically sample B of foreigners in Federal Republic of Germany (FRG, commonly called West Germany), sample D of immigrants in FRG and sample M of immigrants in Unified Germany. Together, sample B, D and M constitute over 75 percent of the estimation sample. Sample B, which was started in 1984, includes households where the head of household is from either of the five Guest-worker countries, specifically Turkey, Greece, Ex-Yugoslavia, Spain or Italy. There are 1,393 households in sample B. Sample D includes households with at least one member who migrated from abroad to FRG after 1984. Sample M was started in 2013 and covers 2,723 households. It includes immigrants who migrated after 1995 to unified Germany.

Of the initial survey sample, 5.1 percent were dropped due to missing values in the key variables such as country of origin, birth year and year of migration. I also drop immigrants who were living in German Democratic Republic (GDR, commonly called East Germany) before 1989 which constituted 0.37 percent because macro-level data on pre-migration history is not available for GDR. Also, as mentioned earlier, the final estimation sample excludes child migrants and only includes those who migrated at age 14 or higher. Table 1 presents the summary statistics of the estimation sample.

4.1 Variables in Timing of Migration Equation

The timing of migration equation given in Equation 10 estimates the hazard rate of migrating to Germany. The dependent variable is the waiting time till migration occurs, i.e., age at migration minus 14. Figure 2 presents the distribution of age at migrating for males and females. For both males and females, the highest proportion of immigrants migrated between the ages of 20 and 30 i.e during the early years of working life. However, the distribution of female immigrants is more dispersed than male immigrants. Non-parametric Kaplan-Meier survival estimates, Nelson-Aalen cumulative hazard and smoothed hazard rate are also given in Figure 4. These estimates suggest that females and males have a similar survival and hazard experience.

Since GSOEP surveys immigrants after they have moved to Germany, estimation of Equation 10 requires creating pre-migration history of each migrant using the information on year of migration and country of origin. The timing of migration equation has four set of covariates: time-constant individual variables, time-varying individual variables, macro-level time-constant variables and macro-level time-varying variables. The time-constant individual characteristics include: an indicator variable if the immigrant is of German ethnicity, a categorical variable of the type of birthplace (city, small city, or rural), a categorical variable of the highest education level of a parent, and the linguistic distance between immigrant's native language and Standard German language.[14] The only individual time-varying variable included in Equation 10 is pre-migration years of schooling. As this variable cannot be directly obtained from the survey, I construct this variable using information on total years of pre-migration schooling and assume continuous education from the age of six.

The time-varying and time-constant macro-level variables are the exogenous push-pull factors of migration. These are collected from several sources and merged with the above mentioned individual variables. Push factors are those which force individuals to leave their home country such as lack of opportunities, unstable political environment and unsatisfactory social development at origin. Pull factors, on the other hand, are factors which attract immigrants to host countries such as better employment opportunities and higher standard of living. The one-time cost of migration such as geographic distance can also influence the decision to migrate. These factors of migration affect the decision to migrate without directly affecting the labor market performance of immigrants in the host country.

[14] The measure of linguistic distance was constructed using the program provided by Max Planck Society for the Advancement of Science and information.

Hence, they are the exclusion restrictions for the joint model given by Equation 14.

The lack of socio-economic development at origin can push individuals to migrate whereas better conditions at host country can attract migrants. To measure the time-varying relative differences in economic development between origin and host country, I use GDP per capita (constant in 2010 dollars) in both the countries. Also, as forward looking individuals care about not only the current but future economic growth in the host country, I include predicted growth of GDP per capita in next five years. As an indicator of the level of social development, I also include life expectancy at origin country.

An unstable political environment, the risk of government collapse or wars can push individuals to relocate at safer destinations. To include these push factors, I use a categorical variable of political instability. At the same time, political factors such as inter-country treaties can also facilitate migration between countries. Specifically for Germany, Guest-worker treaties with Turkey, Spain, Italy, Ex-Yugoslavia etc. encouraged low-skilled migration from these countries. Similarly, Schengen agreements and the formation of the European Union has attracted immigrants to Germany from several European countries. With this in mind, I include indicator variables for Guestworker programs and whether a country is a member of European Union in a given year. To capture differences in Germany pre- and post-unification, I also include an indicator if Germany is unified in the specific year or not.

Apart from the already included linguistic distance, I include an indicator if the origin and host country share a border. The geographic distance measures the monetary cost of moving. Moreover, it also represents the effort cost of collecting information about the host country, which is likely higher for prospective immigrants in geographically distant countries.

4.2 Variables in Wage Assimilation Equation

The wage assimilation equation, given in Equation 11, estimates the rate of wage assimilation, i.e., the wage return on an additional year of stay in the host country. The dependent variable is the log of the net hourly wage rate. As can be observed from Figure 3, the age-earnings profile of male and female immigrants are surprisingly similar. Although the earnings gap between male and female immigrants never decreases, the wage trajectory is similar for both males and females.

The wage assimilation equation is of a standard Mincerian form. Apart from commonly included years of education and work experience, it also includes the length of stay in the host country and pre-migration characteristics. These pre-migration characteristics are the time-constant individual characteristics included in the timing of migration equation. The wage assimilation equation also includes an indicator for the current urban residence and a quadratic polynomial of the time trend.

I do not distinguish between the years of education or experience acquired in origin and home country. Although the returns on schooling and experience attained in the host country is likely to be more valuable relative to that from origin country, I refrain from making such a distinction in the wage equation due to the lack of information on actual schooling in host-country. While it is a common practice in literature to use approximated measures of pre- and post-migration schooling using total years of schooling, age at migration and assuming continuous school attendance from the age of 6 (refer to Friedberg (2000), Bratsberg and Ragan Jr (2002) and Sanromá et al. (2015)), including such measures creates measurement error and bias as pointed out by Duleep (2015). However, the joint model does account for time-constant individual unobserved heterogeneity to account for endogenous total years of schooling.

Exclusion Restriction for Selection in Employment As explained in Section 3.5, the joint model uses inverse propensity weights to account for selection in employment when the model is estimated for both males and females. Estimation of the employment selection equation (refer to Equation 18) requires an exclusion restriction that affects the decision of employment but not the earnings once the individual is employed. I use average commuting distance from home to workplace as an exclusion restriction similar to Jain and Peter (2016). Average commuting distance represents a fixed cost of employment. Long commuting distance can discourage employment, however it is unlikely to affect wages earned by the individual if employed. The average commuting distance varies by state of residence and survey year. It is computed using individual-level responses to three questions regarding commuting distance (in kilometers) to the place of work. The detailed construction of the variable is given in Data appendix.

5 Model Estimates

In this section, I present the estimates for: (1) separately estimated timing of migration migration, (2) separately estimated wage assimilation equation, (3) selection in employment equation and (4) the joint model of (1) and (2). I also discuss the distributions and correlation of immigrant quality, individual-specific rate of assimilation and the unobserved propensity to migrate early.

5.1 Reduced Form Estimates

Timing of Migration The timing of migration estimates for only males and for the full sample are presented in Tables 3 and 4 respectively. I estimate three specifications of the timing of migration equation: a Weibull proportional hazard model, a Gompertz proportional hazard model and a Cox proportional hazard model. The difference between the specifications is due to different distributional assumptions of the baseline hazard. The first specification assumes a Weibull distribution, i.e., $\lambda_0(t) = pt^{p-1}$, the second assumes Gompertz distribution, i.e., $\lambda_0(t) = \exp \gamma t$ and in the third $\lambda_0(t)$ is left unspecified. Based on the estimates in Tables 3 and 4, the different distributional assumptions do not seem to affect the estimates.

Estimates are also consistent with the theoretical predictions of the model given in Section 2. The model predicts that individuals with a higher skill-transferability would migrate at an early age. We observe that ethnic Germans, who are likely to be familiar with the culture and language in Germany and have a higher degree of skill-transferability have a higher hazard of early migration. On the other hand, individuals with a high linguistic distance and hence a low degree of skill-transferability migrate late. Estimates also show that immigrants from geographically distant countries such as in Asia and Africa have a lower hazard of early migration compared to immigrants from Europe. Apart from the cost of migration, immigrants from Asia and Africa also less likely to be familiar with German culture and customs. Hence, immigrants from these continents have a lower degree of skill-transferability and consequently migrate late.

Among the pre-migration individual characteristics, the education of both the immigrant and her parent decreases the hazard of early migration. Thus, individuals from better socio-economic family background and with higher education choose to migrate late. On the other hand, the type (rural versus urban) birthplace does not have a statistically significant effect. The estimates of country-level factors of migration are statistically significant and in line with our expectations: immigrants from countries that share a border with Germany (contiguity) or those with a Guestworker treaty have a higher hazard of early migration whereas immigrants from origin countries with a higher annual GDP per capita have a lower hazard of early migration.

Interestingly, at any given point of time, the GDP per capita in Germany increases the hazard of early migration but the future economic growth in Germany decreases the hazard. This indicates that prospective immigrants postpone migration if they expect a higher economic growth in Germany in the next five years. The estimate of political violence at origin is a bit surprising. The hazard of early migration does not monotonically increases with an increase in the level of political violence. Immigrants migrate early when either there is a low level of political violence or there is a war outbreak. The recent Syrian refugee crisis is a testament to how a war can push people to migrate. However, it is puzzling to find that a medium intensity of political violence at origin decreases the hazard of early migration. A possible explanation is that individuals keep postponing the migration in the hope that the situation at origin will improve in the near future.

Wage Assimilation The estimates of the wage assimilation equation for males and for the full sample are presented in Tables 6 and 7, respectively. I present three specifications of the wage assimilation equation: ordinary least squares, random effects and a linear mixed model. For the full sample, the wage assimilation equation uses inverse propensity weights obtained from estimating the selection into employment equation.

Table 5 presents the results for the Probit equation for the selection into employment. The equation estimates are presented for both the full sample (specification I) and only males (specification II). The effect of average commuting distance to workplace is negative and statistically significant in both the specifications. As expected, women have a lower probability and ethnic Germans have a higher probability of employment. The employment probability increases with the number of years of education and work experience, but decreases with the length of stay. Immigrants with a high linguistic distance compared to those with a zero linguistic distance have a lower employment probability. A surprising result is that immigrants whose parents had higher education are less likely to be employed relative to immigrants whose parents had little or no education.

The estimates of the rate of assimilation in Table 6 and 7 are quite similar across the three specifications. The linear mixed model, however, has a slightly higher rate of assimilation than the ordinary least squares and random effects specifications. An additional year of stay increases the hourly wage by less than one percent. This estimate might seem low, however, it falls within the wide range of estimates that have been reported for various countries. There is no consensus on assimilation rates in either country. For instance, Borjas (1988) using census data reports an earnings growth of over 2 percent for immigrants in both the United States and Canada. However, Borjas (1989) used longitudinal data and found negligible earnings growth of highly skilled immigrants in the United States. Antecol et al. (2006) found the rate of assimilations to be in the range of 5-16 percent and 7-27 percent for Canada and United States respectively. The only consensus between Borjas (1988) and Antecol et al. (2006) is evidence of no assimilation for immigrants in Australia.

For the case of Germany, Dustmann (1993) found a 1.4 percent earnings return on the length of stay for temporary migrants whereas Constant and Massey (2003) found the rates of assimilation in the range of 0.5-0.7 percent. Basilio et al. (2014) also found a return of 0.8 and 0.5 percent on the length of stay in Germany for males and females, respectively. Thus, the reduced form estimate of the average rate of assimilation is comparable to the estimates found by other studies on Germany.

5.2 Joint Model Estimates

The joint model estimates, shown in Table 8, resemble the reduced form estimates of the timing of migration. We observe that the hazard of early migration increases when the host and origin countries share a border. This estimate validates the high volume of migration observed between neighboring countries in the world such as between the United States and Mexico. The immigrants from origin countries which had a Guest worker treaty with Germany have a higher hazard of early migration. On the other hand, immigrants from member countries of European Union migrate late. Also, immigrants are more likely to migrate after the fall of Berlin wall in a unified Germany.

Similar to the reduced form estimates, we observe that migrants from Asia and Africa are less likely to migrate early relative to European migrants. Also, a low level of political violence and war outbreak increases the hazard of early migration, but a medium level of political violence does not. Among individual characteristics, high linguistic distance, a rural place of ubpringing, and pre-migration years of schooling decrease the hazard of early migration. Being an ethnic German, on the other hand, increases the hazard.

The estimates from the wage assimilation equation in the joint model (see Table 9) show that the reduced form estimates of the rate of assimilation are biased upward. We observe that the joint-estimate of the average rate of assimilation in nearly half of what is suggested by the linear mixed model and is closer to the ordinary least square estimates (refer to Table 6). The average rate of assimilation in the joint model is 0.6 percent whereas the linear model predicts it to be 1 percent. Thus, the failure to account for selective timing of migration overestimates the actual rate of assimilation. As the unobserved propensity to migrate early and individual rates of assimilation are positively correlated (refer to Table 10), individuals who have a high propensity to migrate early and also have high rates of assimilation appear for a longer duration of time in the data. The upward bias is potentially a result of such sample selection.

As one would expect, years of schooling and work experience have a positive effect on the hourly wage, by 3 and 2.5 percent respectively. On the other hand, a high linguistic distance has a negative effect. Among pre-migration individual characteristics, a rural place of upbringing and German ethnicity have a weak negative effect on the wage. However, a better socio-economic status (as indicated by parent's education) positively affects an immigrant's wage.

Immigrant quality, Individual Rate of Assimilation and Propensity of Early Migration The distribution of the individual-specific rate of assimilation in Figure 6 shows individual rates of assimilation vary significantly between immigrants. This suggests that the rate of human capital acquisition after migration varies widely between immigrants. Even though the average assimilation rate is 0.6 percent , the individual rate of assimilation can be as high as 5 percent and as low as a negative 3.5 percent. Thus, it is clear that an average rate of assimilation hides a remarkable degree of variation in the individual-specific rates of assimilation. Although a negative rate of assimilation seems puzzling, Chiswick and Miller (2011) have found "negative" assimilation to occur for immigrants with highly transferable skills. Consistent with Chiswick and Miller's finding, we observe that negative assimilation mostly occurs for immigrants from Guest-worker countries who could easily use their pre-migration skills in Germany.

Unsurprisingly, the immigrant quality and the unobserved propensity to migrate early vary considerably between immigrants. A closer look at the variation of immigrant quality by continent reveals an interesting feature (refer to Figure 10). Both the mean and the median immigrant quality is high for immigrants from America and Europe, the two continents with the developed countries and a geographical proximity to Germany. However, the mean and median immigrant quality of immigrants from Africa and Asia is significantly lower. This suggests European and American immigrants on average have higher wages due to a higher unobserved quality.

Further analysis of the immigrant quality by country of origin reveals: (1) immigrants from countries which had a Guest worker treaty with Germany (specifically Turkey, ex-Yugoslavia, Spain, Italy and Greece) have on average a higher immigrant quality and (2) immigrants from countries where most migration was due to wars and political unrest (such as Russia, Kazakhstan, Ukraine, Bosnia-Herzegovina and Kosovo-Albania) have on average a lower immigrant quality. As a large proportion of immigrants from Guest worker countries consist of low skilled workers who worked in blue collar jobs, these immigrants have a high degree of skill-transferability. In contrast, refugees, for whom migration is an unplanned event, lack transferable skills. Such findings show that skill-transferability has a large bearing on immigrant quality.

The covariance structure (see Table 10 and Figures 8, 9, 13 and, 14) shows that immigrants of high quality (skill-transferability) have a lower rate of assimilation and vice-versa. This finding is consistent with the theoretical predictions of the Immigrant Human Capital Investment (IHCI) model in Duleep and Regets (1999). Their model predicts that immigrants with less-transferable skills would have a lower opportunity cost of acquiring human capital in the host country. Thus, immigrants with less-transferable skills are more likely to invest in human capital after migration and consequently have a higher rate of assimilation. Empirically, Duleep and Regets (2002) have also shown an inverse relationship between the growth of immigrants' earnings and immigrants' entry earnings, where entry earnings are used to proxy the degree of skill-transferability.[15] However, a few papers in literature such as Borjas (1987, 1994) argue that immigrant quality indicates the level of innate ability. Estimates from the joint model reinforce the theoretical and empirical evidence on the inverse relationship between skill-transferability and wage growth by calculating the

[15] Similar findings were earlier reported by LaLonde and Topel (1991).

residuals that measure the degree of skill-transferability (instead of relying on a contentious proxy).

As suggested by Borjas (1998), a negative correlation could also indicate that there exists a “relative substitutability” between pre- and post-migration human capital. If immigrants can utilize a substantial size of the pre-migration skills in the host country, they would not face a high initial disadvantage in host country’s labor market. As a result, augmenting human capital stock after migration would be more expensive. Borjas predicts that in such a case, immigrants would have a slower wage growth.

The covariance structure also shows that immigrants who have a higher propensity to migrate early, also have a higher rate of assimilation, thus, indicating a positive correlation. This can be explained by two scenarios: (1) individuals with a higher propensity migrate early and invest more in host-country specific human capital, or (2) individuals who expect a high wage growth after migration choose to migrate earlier. Although, we cannot separately identify the relative importance of the two cases, it is clear that timing of migration and wage assimilation are not independent.

6 Conclusion

This paper shows both theoretically and empirically that the commonly made exogeneity assumption of the length of stay in the host country is incorrect. This is the first paper in the literature to develop and estimate a joint model of timing of migration and economic assimilation. The joint model has two advantages: (1) it accounts for the selective timing of migration; and (2) it estimates the distributions of individual-specific rates of assimilation, immigrant quality, and the propensity to migrate early, and it estimates the correlation between these components.

Estimates from the joint model show that the unobserved propensity to migrate early and individual rates of assimilation are positively correlated. Such findings validate my assertion that the two processes must be estimated together. We also observe that individual rates of assimilation are higher among immigrants who are of comparatively lower quality and have lower skill-transferability. Hence, a catch-up effect is observed between low-quality and high-quality migrants. These findings address concerns about immigrant assimilation, especially in recent times when several countries have received an influx of forced migrants. The model predicts that although forced migrants face an initial disadvantage in the host-country labor-market, they rapidly invest in host-country-specific human capital and eventually reach their potential.

My estimates also suggest that the commonly estimated average rate of assimilation suffers from an upward bias if the timing of migration is not accounted for. Moreover, we observe that individual-specific rates of assimilation vary a great deal between immigrants. Clearly, differences in the labor market performance of immigrants are due both to differences in unobserved immigrant quality and to the tendency to acquire human capital post-migration. To date, immigration policies have emphasized that when immigrants are screened it is important to identify individual ability. However, variation in rates of assimilation between immigrants of similar quality indicates that host countries will be better served if they also selected immigrants who have a higher incentive and an ability to augment their existing stock of human capital after migration.

References

Antecol, H., Kuhn, P., and Trejo, S. J. (2006). Assimilation via prices or quantities? sources of immigrant earnings growth in australia, canada, and the united states. *Journal of human Resources*, 41(4):821–840. 1, 5.1

Bakker, D., Müller, A., Velupillai, V., Wichmann, S., Brown, C. H., Brown, P., Egorov, D., Mailhammer, R., Grant, A., and Holman, E. W. (2009). Adding typology to lexicostatistics: A combined approach to language classification. *Linguistic Typology*, 13(1):169–181. 17

Basilio, L., Bauer, T. K., and Kramer, A. (2014). Transferability of human capital and immigrant assimilation: An analysis for germany. 5.1

Bleakley, H. and Chin, A. (2004). Language skills and earnings: Evidence from childhood immigrants*. *Review of Economics and Statistics*, 86(2):481–496. 1, 2

Bleakley, H. and Chin, A. (2008). What holds back the second generation? the intergenerational transmission of language human capital among immigrants. *Journal of Human Resources*, 43(2):267–298. 1, 2

Bleakley, H. and Chin, A. (2010). Age at arrival, english proficiency, and social assimilation among us immigrants. *American economic journal. Applied economics*, 2(1):165. 1, 2

Borjas, G. J. (1985). Assimilation, changes in cohort quality, and the earnings of immigrants. *Journal of labor Economics*, pages 463–489. 3.1

Borjas, G. J. (1987). Self-selection and the earnings of immigrants. 1, 3.1, 3.3, 5.2

Borjas, G. J. (1988). *International Differences in the Labor Market Performance of Immigrants.* ERIC. 1, 5.1

Borjas, G. J. (1989). Immigrant and emigrant earnings: a longitudinal study. *Economic inquiry*, 27(1):21–37. 5.1

Borjas, G. J. (1994). Assimilation and changes in cohort quality revisited: what happened to immigrant earnings in the 1980s? Technical report, National Bureau of Economic Research. 1, 5.2

Borjas, G. J. (1998). The economic progress of immigrants. Technical report, National Bureau of Economic Research. 5.2

Borjas, G. J. (2013). The slowdown in the economic assimilation of immigrants: Aging and cohort effects revisited again. Technical report, National Bureau of Economic Research. 1, 3

Bratsberg, B. and Ragan Jr, J. F. (2002). The impact of host-country schooling on earnings: A study of male immigrants in the united states. *Journal of Human resources*, pages 63–105. 4.2

Chiswick, B. R. (1978). The effect of americanization on the earnings of foreign-born men. *The Journal of Political Economy*, pages 897–921. 1, 3.1

Chiswick, B. R. and Miller, W. (2011). The negative assimilation of immigrants: a special case. *Industrial & Labor Relations Review*, 64(3):502–525. 5.2

Cobb-Clark, D. A. (1993). Immigrant selectivity and wages: The evidence for women. *The American Economic Review*, pages 986–993. 2

Cobb-Clark, D. A., Broadway nee Hanel, B., and McVicar, D. (2012). Immigrant wage and employment assimilation: A comparison of methods. 1

Constant, A. and Massey, D. S. (2003). Self-selection, earnings, and out-migration: A longitudinal study of immigrants to germany. *Journal of population Economics*, 16(4):631–653. 5.1

Duleep, H. O. (2015). Chapter 3 - the adjustment of immigrants in the labor market. In Chiswick, B. R. and Miller, P. W., editors, *Handbook of the Economics of International Migration*, volume 1 of *Handbook of the Economics of International Migration*, pages 105 – 182. North-Holland. 4.2

Duleep, H. O. and Regets, M. C. (1999). Immigrants and human-capital investment. *American Economic Review*, pages 186–191. 1, 2, 5.2

Duleep, H. O. and Regets, M. C. (2002). The elusive concept of immigrant quality: Evidence from 1970-1990. 3.1, 3.3, 5.2

Dustmann, C. (1993). Earnings adjustment of temporary migrants. *Journal of Population Economics*, 6(2):153–168. 5.1

Ezra, M. and Kiros, G.-E. (2001). Rural out-migration in the drought prone areas of ethiopia: a multilevel analysis. *International Migration Review*, pages 749–771. 1

Fertig, M. and Schurer, S. (2007). Earnings assimilation of immigrants in germany: the importance of heterogeneity and attrition bias. 1, 1, 3, 3.1

Friedberg, R. M. (1992). The labor market assimilation of immigrants in the united states: the role of age at arrival. 3.1

Friedberg, R. M. (2000). You can't take it with you? immigrant assimilation and the portability of human capital. *Journal of Labor Economics*, 18(2):221–251. 4.2

Hancock, G. R. and Mueller, R. O. (2010). *The reviewers guide to quantitative methods in the social sciences*. Routledge. 9

Hare, D. (1999). pushversus pullfactors in migration outflows and returns: Determinants of migration status and spell duration among china's rural population. *The Journal of Development Studies*, 35(3):45–72. 1

Henry, S., Schoumaker, B., and Beauchemin, C. (2004). The impact of rainfall on the first out-migration: A multi-level event-history analysis in burkina faso. *Population and Environment*, 25(5):423–460. 1

Hu, W.-Y. (2000). Immigrant earnings assimilation: estimates from longitudinal data. *American Economic Review*, pages 368–372. 1

LaLonde, R. J. and Topel, R. H. (1991). Immigrants in the american labor market: Quality, assimilation, and distributional effects. *The American economic review*, pages 297–302. 15

LaLonde, R. J. and Topel, R. H. (1992). The assimilation of immigrants in the us labor market. In *Immigration and the workforce: Economic consequences for the United States and source areas*, pages 67–92. University of Chicago Press. 3.1

Lubotsky, D. (2007). Chutes or ladders? a longitudinal analysis of immigrant earnings. *Journal of Political Economy*, 115(5):820–867. 1

Reed, H. E., Andrzejewski, C. S., and White, M. J. (2010). Mens and womens migration in coastal ghana: An event history analysis. *Demographic research*, 22. 1

Sanromá, E., Ramos, R., and Simón, H. (2015). How relevant is the origin of human capital for immigrant wages? evidence from spain. *Journal of Applied Economics*, 18(1):149–172. 4.2

Schaafsma, J. and Sweetman, A. (2001). Immigrant earnings: age at immigration matters. *Canadian journal of Economics*, pages 1066–1099. 3.1

Wagner, G. G., Frick, J. R., and Schupp, J. (2007). The german socio-economic panel study (soep)-evolution, scope and enhancements. 4

Zimmermann, K. F. and Constant, A. (2012). *How labor migrants fare*. Springer Science & Business Media. 2

7 Tables

Table 1: Summary Statistics of the Key Variables

	Male	***Female***	***Both***
Length of Stay	18.274	17.397	17.865
	(9.078)	(9.011)	(9.057)
Total years of schooling	10.261	10.207	10.236
	(2.269)	(2.532)	(2.395)
Years of actual work experience	21.908	12.023	17.302
	(11.204)	(10.653)	(12.010)
Linguistic distance			
Zero	0.009	0.011	0.010
Low	0.010	0.014	0.012
Medium	0.016	0.013	0.014
High	0.964	0.960	0.962
Ethnic German	0.219	0.263	0.239
Parents education			
Basic Secondary and Lower Vocational	0.769	0.707	0.740
General Secondary & Upper Vocational	0.162	0.197	0.178
Higher education	0.068	0.940	0.080
Place of Upbringing			
City	0.344	0.378	0.360
Small city	0.235	0.243	0.239
Rural	0.420	0.377	0.400
Urban current residence	0.846	0.848	0.847
N	21,668	18,907	40,575

Notes: The summary statistics are given for the observations directly available from the survey and does not include pre-migration histories. The missing category in 'Highest education level of a parent' and 'Place of upbringing' are not shown. Standard deviations are given in parenthesis.

Table 2: Key Variables in Timing of Migration and Wage Assimilation Equations

	Timing of Migration	***Wage Assimilation***
Time-varying Macro Variables		
GDP per capita at origin	✓	
GDP per capita at host	✓	
Growth of GDP per capita at host in next 5 years	✓	
Life expectancy at home	✓	
Political Instability at home	✓	
Member of EU	✓	
Guestworker treaty	✓	
Unified Germany	✓	
Time-constant Macro Variables		
Origin country's continent	✓	
Contiguity	✓	
Time-varying Individual Variables		
Pre-migration schooling	✓	
Total years of schooling		✓
Length of Stay		✓
Actual years of work experience		✓
Type of current residence		✓
Time-constant Individual Variables		
Female	✓	✓
Ethnic German	✓	✓
Linguistic distance	✓	✓
Type of birthplace	✓	✓
Highest education level of a parent	✓	✓

Notes: Wage Assimilation equation also includes a second order polynomial of time trend and squared term of actual years of work experience. Political Violence, home country's continent, type of birth place, linguistic distance and highest education level of a parent are categorical variables. Member of EU, guestworker treaty, unified Germany, contiguity, ethnic German and type of current residence are indicator variables. Indicator for female is included when the joint-model is estimated for both male and female.

Table 3: Timing of Migration: Reduced Form Estimates for Men

	Weibull (I)	***Gompertz (II)***	***Cox (III)***
Pre-migration years of schooling	-0.098***	-0.078***	-0.088***
	(0.009)	(0.010)	(0.009)
Ethnic German	0.371***	0.345***	0.386***
	(0.056)	(0.061)	(0.056)
Linguistic Distance			
Lowest	-0.328	-0.455	-0.340
	(0.286)	(0.294)	(0.280)
Medium	-0.341	-0.345	-0.332
	(0.218)	(0.213)	(0.211)
Highest	-0.925***	-0.963***	-0.917***
	(0.182)	(0.172)	(0.176)
Parents education			
General Secondary & Upper Vocational	-0.108**	-0.118**	-0.126**
	(0.051)	(0.053)	(0.049)
Higher Education	0.069	0.043	0.038
	(0.063)	(0.064)	(0.061)
Place of Upbringing			
Small City	-0.051	-0.033	-0.041
	(0.052)	(0.053)	(0.050)
Rural	-0.071	-0.050	-0.048
	(0.046)	(0.048)	(0.045)
Continent of home country			
Asia	-0.769***	-0.781***	-0.777***
	(0.122)	(0.122)	(0.119)
America	0.313**	0.324**	0.289**
	(0.140)	(0.146)	(0.135)
Africa	-0.279***	-0.221***	-0.260***
	(0.070)	(0.073)	(0.069)
Political Violence			
Limited	0.326***	0.295***	0.314***
	(0.061)	(0.061)	(0.059)
Serious	-0.746***	-0.728***	-0.760***
	(0.069)	(0.068)	(0.068)
Warfare	0.243***	0.173***	0.218***
	(0.060)	(0.060)	(0.059)

Continued on next page ...

Table 3 continued:

	Weibull (I)	***Gompertz (II)***	***Cox (III)***
Unified Germany	0.302***	0.335***	0.340***
	(0.065)	(0.064)	(0.064)
Contiguity	0.258***	0.281***	0.260***
	(0.081)	(0.089)	(0.080)
Guestworker treaty	1.181***	1.239***	1.188***
	(0.064)	(0.066)	(0.063)
European Union	-0.260***	-0.245***	-0.262***
	(0.067)	(0.068)	(0.066)
Log of GDP per capita in origin country	-0.390***	-0.414***	-0.391***
	(0.045)	(0.045)	(0.044)
Log of GDP per capita in host country	1.110***	1.075***	1.120***
	(0.146)	(0.146)	(0.144)
Economic growth in the next 5 years in host country	-0.122***	-0.109***	-0.109***
	(0.029)	(0.028)	(0.028)
Life expectancy in origin country	0.013***	0.013***	0.014***
	(0.003)	(0.003)	(0.003)
Constant	-11.430***	-9.910***	
	(1.425)	(1.425)	
N	39,113	39,113	39,113

Notes: A constant and a second order polynomial of time trend was also included in all specifications. Missing Category in Parent's education is "Basic secondary & lower vocation"; in Place of upbringing is "Big or medium city" and in Linguistic Distance is "Zero Distance". Standard errors are given in parenthesis. Adjusted R2: 0.248 for specification (I). *** p<0.01, ** p<0.05, * p<0.1

Table 4: Timing of Migration: Reduced Form Estimates for the Full Sample

	Weibull (I)	*Gompertz* (II)	*Cox* (III)
Pre-migration years of schooling	-0.081***	-0.058***	-0.073***
	(0.007)	(0.007)	(0.006)
Ethnic German	0.326***	0.309***	0.346***
	(0.040)	(0.042)	(0.039)
Linguistic Distance			
Lowest	-0.534***	-0.592***	-0.523***
	(0.205)	(0.209)	(0.196)
Medium	-0.230	-0.225	-0.224
	(0.182)	(0.181)	(0.173)
Highest	-0.852***	-0.857***	-0.843***
	(0.155)	(0.152)	(0.147)
Parents' education			
General Secondary & Upper Vocational	-0.135***	-0.153***	-0.154***
	(0.037)	(0.038)	(0.036)
Higher Education	0.058	0.019	0.019
	(0.045)	(0.045)	(0.043)
Place of Upbringing			
Small City	0.004	0.014	0.012
	(0.038)	(0.038)	(0.036)
Rural	-0.023	-0.004	-0.004
	(0.034)	(0.035)	(0.033)
Continent of home country			
Asia	-0.756***	-0.751***	-0.766***
	(0.095)	(0.094)	(0.092)
America	0.076	0.088	0.058
	(0.107)	(0.111)	(0.102)
Africa	-0.320***	-0.282***	-0.310***
	(0.049)	(0.051)	(0.048)
Political Violence			
Limited	0.305***	0.279***	0.296***
	(0.046)	(0.046)	(0.045)
Serious	-0.792***	-0.782***	-0.815***
	(0.051)	(0.051)	(0.051)
Warfare	0.212***	0.146***	0.200***
	(0.042)	(0.042)	(0.041)

Continued on next page . . .

Table 4 continued:

	Weibull (I)	***Gompertz (II)***	***Cox (III)***
Unified Germany	0.348***	0.379***	0.386***
	(0.047)	(0.047)	(0.046)
Contiguity	0.331***	0.347***	0.334***
	(0.054)	(0.057)	(0.052)
Guestworker treaty	1.204***	1.241***	1.205***
	(0.048)	(0.049)	(0.046)
European Union	-0.342***	-0.335***	-0.350***
	(0.048)	(0.048)	(0.046)
Log of GDP per capita in origin country	-0.297***	-0.315***	-0.291***
	(0.032)	(0.032)	(0.031)
Log of GDP per capita in host country	1.544***	1.503***	1.553***
	(0.113)	(0.113)	(0.111)
Economic growth in the next 5 years in host country	-0.048**	-0.037*	-0.039*
	(0.021)	(0.021)	(0.021)
Life expectancy in origin country	0.007***	0.008***	0.007***
	(0.003)	(0.003)	(0.002)
Female	0.014	0.010	0.016
	(0.027)	(0.028)	(0.026)
Constant	-16.525***	-15.123***	...
	(1.115)	(1.107)	
N	75,235	75,235	75,235

Notes: A constant and was also included in all specifications. Missing Category in Parent's education is "Basic secondary & lower vocation"; in Place of upbringing is "Big or medium city" and in Linguistic Distance is "Zero Distance". Standard errors are given in parenthesis. Adjusted R2: 0.248 for specification (I). *** p<0.01, ** p<0.05, * p<0.1

Table 5: Probit Estimates for Selection in Employment

	Full Sample (I)	***Only Men (II)***
Average commuting distance to work	-0.024***	-0.032***
	(0.004)	(0.007)
Length of stay	-0.005***	-0.016***
	(0.001)	(0.001)
Total years of schooling	0.050***	0.049***
	(0.003)	(0.005)
Years of actual work experience	0.078***	0.056***
	(0.002)	(0.003)
Years of actual work experience, squared	-0.002***	-0.001***
	(0.000)	(0.000)
Linguistic Distance		
Lowest	-0.221***	-0.390***
	(0.081)	(0.127)
Medium	-0.073	0.078
	(0.078)	(0.122)
Highest	-0.100*	-0.162*
	(0.059)	(0.096)
German ethnicity	0.080***	0.018
	(0.016)	(0.025)
Parents' education		
General Secondary & Upper Vocational	0.011	-0.005
	(0.019)	(0.029)
Higher Education	-0.062**	-0.071*
	(0.027)	(0.041)
Place of upbringing		
Small city	0.164***	0.219***
	(0.017)	(0.026)
Rural	0.087***	0.108***
	(0.015)	(0.022)
Current urban residence	0.052***	0.064**
	(0.017)	(0.025)
Female	-0.446***	
	(0.014)	
N	48,071	23,245

Notes: A constant and a second order polynomial of time trend was included in both the specifications. Missing Category in Parent's education is "Basic secondary & lower vocation" and in Place of upbringing is "Big or medium city". Standard errors are given in parenthesis. Adjusted R2: 0.101 for specification (I) and 0.051 for specification (II). *** p<0.01, ** p<0.05, * p<0.1

Table 6: Wage Assimilation: Reduced Form Estimates for Men

	OLS	***Random Effects***	***Linear Mixed***
	(I)	***(II)***	***(III)***
Length of Stay	0.009***	0.008***	0.010***
	(0.000)	(0.001)	(0.001)
Total years of schooling	0.032***	0.033***	0.026***
	(0.002)	(0.004)	(0.004)
Years of actual work experience	0.020***	0.024***	0.022***
	(0.001)	(0.002)	(0.002)
Years of actual work experience, squared	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)
Linguistic Distance			
Lowest	0.073	0.093	0.145
	(0.049)	(0.116)	(0.126)
Medium	-0.207***	-0.133	-0.111
	(0.045)	(0.097)	(0.110)
Highest	-0.349***	-0.352***	-0.352***
	(0.035)	(0.079)	(0.089)
Ethnic German	-0.005	-0.015	-0.001
	(0.008)	(0.016)	(0.016)
Parents education			
General Secondary & Upper Vocational	0.017*	0.045**	0.047***
	(0.009)	(0.018)	(0.018)
Higher Education	0.075***	0.150***	0.144***
	(0.016)	(0.030)	(0.032)
Place of Upbringing			
Small City	0.009	-0.020	-0.028*
	(0.007)	(0.016)	(0.017)
Rural	-0.003	-0.015	-0.023
	(0.006)	(0.014)	(0.015)
Urban current residence	0.077***	0.080***	0.081***
	(0.007)	(0.017)	(0.019)
Constant	1.787***	1.736***	1.812***
	(0.043)	(0.098)	(0.103)
N	17,264	17,264	17,264

Notes: A second order polynomial of time trend was also included in all specifications. Missing Category in Parent's education is "Basic secondary & lower vocation" and in Place of upbringing is "Big or medium city". Standard errors are given in parenthesis. Adjusted R2: 0.166 for specification (I). *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 7: Wage Assimilation: Reduced Form Estimates for the Full Sample with Inverse Propensity Weighting

	OLS	***Random Effects***	***Linear Mixed***
	(I)	***(II)***	***(III)***
Length of Stay	0.008***	0.009***	0.010***
	(0.000)	(0.001)	(0.001)
Total years of schooling	0.036***	0.031***	0.027***
	(0.001)	(0.003)	(0.003)
Years of actual work experience	0.014***	0.019***	0.018***
	(0.001)	(0.001)	(0.001)
Years of actual work experience, squared	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)
Linguistic Distance			
Lowest	-0.044	-0.032	0.013
	(0.034)	(0.078)	(0.082)
Medium	-0.174***	-0.114	-0.071
	(0.033)	(0.071)	(0.079)
Highest	-0.244***	-0.283***	-0.251***
	(0.022)	(0.052)	(0.057)
Ethnic German	-0.037***	-0.028**	-0.025**
	(0.006)	(0.011)	(0.012)
Parents education			
General Secondary & Upper Vocational	0.014*	0.041***	0.034**
	(0.007)	(0.013)	(0.014)
Higher Education	0.069***	0.137***	0.122***
	(0.013)	(0.021)	(0.023)
Place of Upbringing			
Small City	0.011*	-0.012	-0.014
	(0.006)	(0.012)	(0.013)
Rural	-0.012**	-0.008	-0.009
	(0.006)	(0.011)	(0.012)
Urban current residence	0.068***	0.080***	0.077***
	(0.006)	(0.014)	(0.014)
Female	-0.251***	-0.222***	-0.227***
	(0.005)	(0.010)	(0.011)
N	29,712	29,712	29,712

Notes: A constant and a second order polynomial of time trend was also included in all specifications. Missing Category in Parent's education is "Basic secondary & lower vocation" and in Place of upbringing is "Big or medium city". Standard errors are given in parenthesis. Adjusted R2: 0.248 for specification (I). *** p<0.01, ** p<0.05, * p<0.1

Table 8: Timing of Migration: Joint Model Estimates for Men

	Coefficient	***Std. Error***
Pre-migration years of schooling	-0.129	(0.011)
Ethnic German	0.381	(0.068)
Linguistic distance		
Low	-0.065	(0.261)
Medium	0.019	(0.210)
High	-0.647	(0.109)
Parents education		
General Secondary & Upper Vocational	-0.153	(0.065)
Higher Education	0.043	(0.086)
Place of Upbringing		
Small City	-0.066	(0.054)
Rural	-0.122	(0.054)
Continent of home country		
Asia	-0.311	(0.076)
America	0.285	(0.176)
Africa	-0.750	(0.140)
Political Violence		
Limited	0.421	(0.072)
Serious	-0.667	(0.087)
Warfare	0.385	(0.071)
Unified Germany	0.686	(0.050)
Contiguity	0.342	(0.089)
Guestworker treaty	1.343	(0.008)
European Union	-0.256	(0.056)
Log of GDP per capita in origin country	-0.453	(0.011)
Log of GDP per capita in host country	-0.032	(0.000)
Economic growth in the next 5 years in host country	-0.328	(0.000)
Life expectancy in origin country	0.028	(0.000)
N	39,113	

Notes: Missing Category in Parent's education is "Basic secondary & lower vocation" and in Place of upbringing is "Big or medium city"; in Linguistic distance is "Zero linguistic distance". The standard errors were computed by taking square root of the diagonal elements of the inverted Hessian. The Hessian was computed numerically.

Table 9: Wage Assimilation: Joint Estimates for Men

	Coefficient	***Std. Error***
Length of Stay	0.006	(0.000)
Total years of schooling	0.030	(0.000)
Years of actual work experience	0.025	(0.000)
Years of actual work experience, squared	0.000	(0.000)
Ethnic German	-0.012	(0.014)
Linguistic distance		
Low	0.261	(0.093)
Medium	0.028	(0.076)
High	-0.221	(0.064)
Parents' education		
General Secondary & Upper Vocational	0.033	(0.010)
Higher Education	0.134	(0.022)
Place of Upbringing		
Small City	-0.031	(0.014)
Rural	-0.022	(0.012)
Urban current residence	0.087	(0.014)
Time trend	0.014	(0.000)
Time trend, square	0.000	(0.000)
Constant	1.625	(0.066)
N	17,264	

Notes: Missing Category in Parent's education is "Basic secondary & lower vocation" and in Place of upbringing is "Big or medium city"; in Linguistic distance is "Zero linguistic distance". The standard errors were computed by taking square root of the diagonal elements of the inverted Hessian. The Hessian was computed numerically.

Table 10: Variance-Covariance Structure of Immigrant Quality, Propensity of Early-Migration and Individual Variation from Average Rate of Assimilation: Joint Model Estimates for Men

	Parameter	***Std. Error***		***Parameter***	***Std. Error***
σ_a	0.431	(0.006)	ρ_{ab}	-0.776	(0.000)
σ_b	0.021	(0.000)	ρ_{ac}	0.074	(0.000)
σ_c	0.490	(0.001)	ρ_{bc}	0.185	(0.000)

Notes: Individual variation from the average rate of assimilation is denoted by 'b', propensity to migrate early by 'c' and immigrant quality by 'a'. The standard errors were computed by taking square root of the diagonal elements of the inverted Hessian. The Hessian was computed numerically.

Table 11: Timing of Migration: Joint Model Estimates for Full Sample

	Coefficient	***Std. Error***
Female	-0.006	(0.041)
Pre-migration years of schooling	-0.106	(0.008)
Ethnic German	0.500	(0.053)
Linguistic distance		
Low	-0.786	(0.242)
Medium	-0.557	(0.320)
High	-1.357	(0.158)
Parents education		
General Secondary & Upper Vocational	-0.190	(0.046)
Higher Education	0.075	(0.059)
Place of Upbringing		
Small City	-0.021	(0.051)
Rural	-0.081	(0.040)
Continent of home country		
Asia	-0.398	(0.082)
America	0.101	(0.163)
Africa	-0.734	(0.172)
Political Violence		
Limited	0.376	(0.045)
Serious	-0.596	(0.053)
Warfare	0.550	(0.048)
Unified Germany	0.876	(0.040)
Contiguity	0.344	(0.070)
Guest-worker treaty	1.508	(0.054)
European Union	-0.275	(0.052)
Log of GDP per capita in origin country	-0.428	(0.041)
Log of GDP per capita in host country	-0.001	(0.033)
Economic growth in the next 5 years in host country	-0.299	(0.018)
Life expectancy in origin country	0.027	(0.001)
N	75,235	

Notes: Missing Category in Parent's education is "Basic secondary & lower vocation" and in Place of upbringing is "Big or medium city"; in Linguistic distance is "Zero linguistic distance". The standard errors were computed by taking square root of the diagonal elements of the inverted Hessian. The Hessian was computed numerically.

Table 12: **Wage Assimilation: Joint Estimates for Full Sample**

	Coefficient	***Std. Error***
Length of Stay	0.007	(0.000)
Female	-0.221	(0.009)
Total years of schooling	0.028	(0.001)
Years of actual work experience	0.019	(0.000)
Years of actual work experience, squared	0.000	(0.000)
Linguistic distance		
Low	-0.027	(0.047)
Medium	-0.100	(0.057)
High	-0.289	(0.046)
Ethnic German	-0.030	(0.010)
Parents' education		
General Secondary & Upper Vocational	0.033	(0.011)
Higher Education	0.114	(0.016)
Place of Upbringing		
Small City	-0.008	(0.011)
Rural	-0.006	(0.010)
Urban current residence	0.082	(0.010)
Time trend	0.012	(0.000)
Time trend square	0.000	(0.000)
Constant	1.750	(0.047)
N	29,712	

Notes: Missing Category in Parent's education is "Basic secondary & lower vocation" and in Place of upbringing is "Big or medium city"; in Linguistic distance is "Zero linguistic distance". The standard errors were computed by taking square root of the diagonal elements of the inverted Hessian. The Hessian was computed numerically.

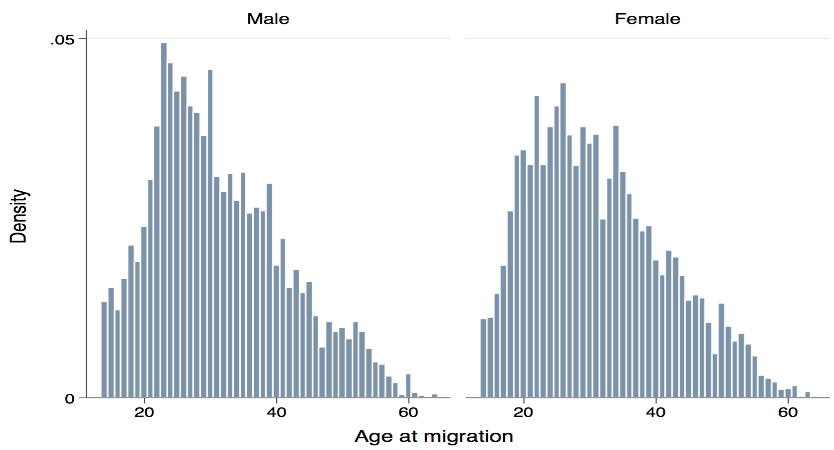
Table 13: Variance-Covariance Structure of Immigrant Quality, Propensity of Early-Migration and Individual Variation from Average Rate of Assimilation: Joint Model Estimates for Full Sample

	Parameter	***Std. Error***		***Parameter***	***Std. Error***
σ_a	0.446	(0.005)	ρ_{ab}	-0.787	(0.007)
σ_b	0.021	(0.000)	ρ_{ac}	0.027	(0.015)
σ_c	0.530	(0.005)	ρ_{bc}	0.142	(0.005)

Notes: Individual variation from the average rate of assimilation is denoted by 'b', propensity to migrate early by 'c' and immigrant quality by 'a'. The standard errors were computed by taking square root of the diagonal elements of the inverted Hessian. The Hessian was computed numerically.

8 Figures

Figure 2: Histogram of Age at Migration by Gender

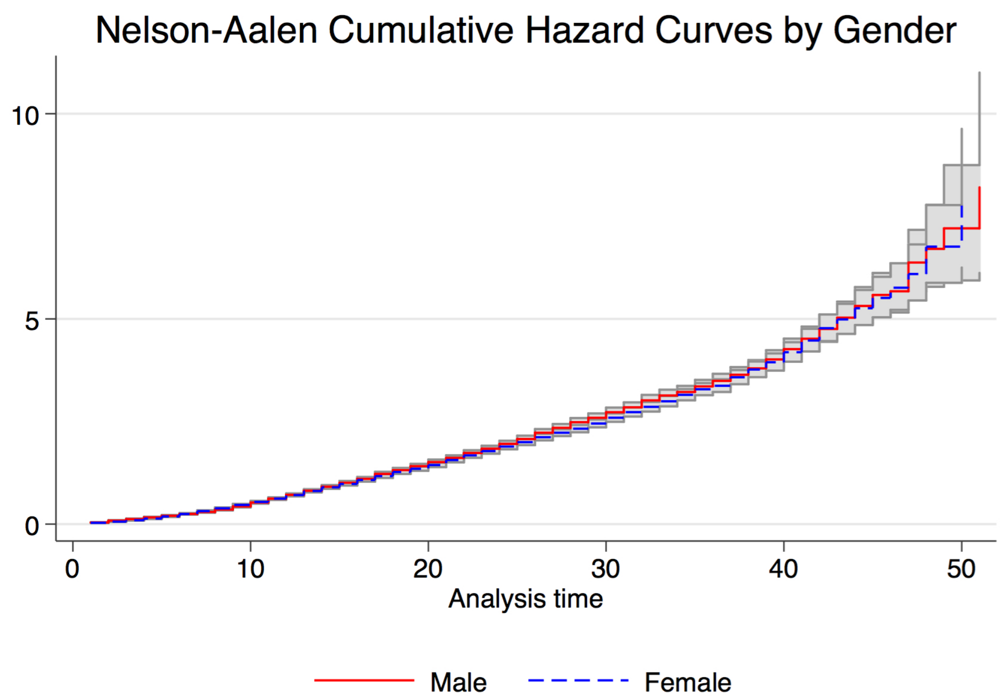


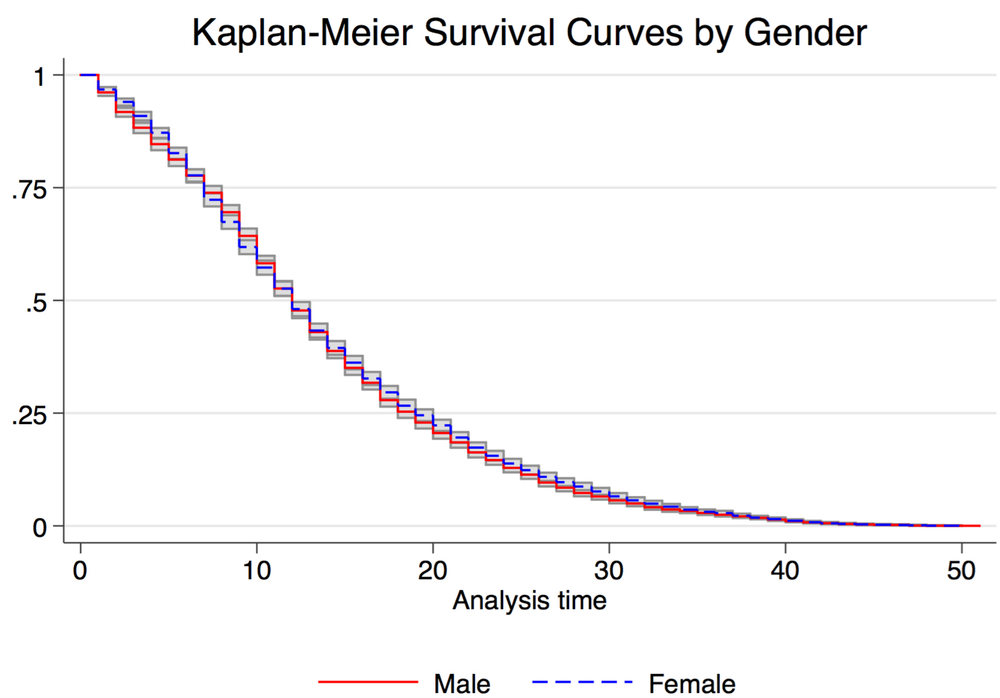
Notes: The earliest age at migration is 14 and the highest is 64.

Figure 3: Age-Earnings Profile by Gender

Notes: The estimates are calculated from ordinary least square regression of the log hourly wage on a quadratic polynomial of age and its interaction with the indicator for female and with robust standard errors.

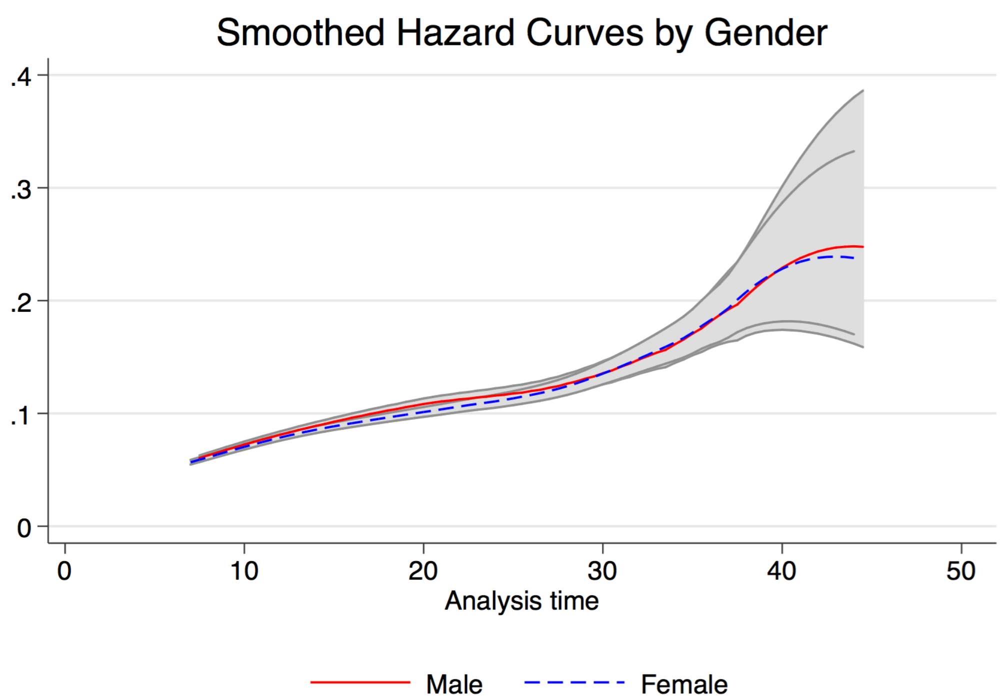
Figure 4: Non Parametric Estimates





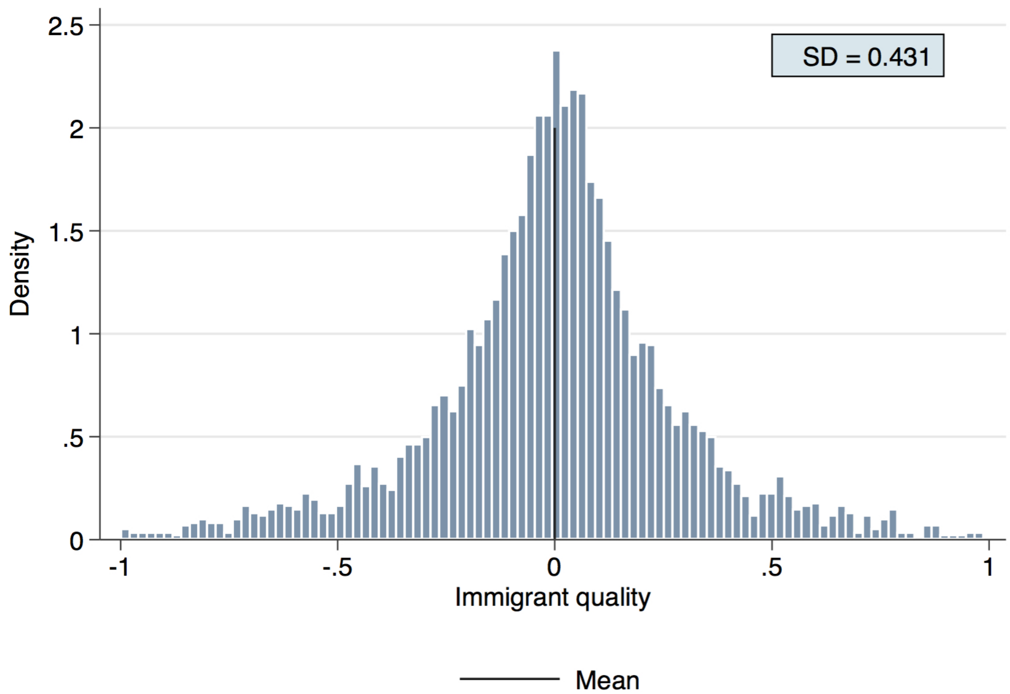
Notes: The analysis time begins at age 14 and ends at age 64. Survival function is the opposite of cumulative hazard function. Specifically, $S(t) = exp \int_0^t \lambda(x)dx$ where $\Lambda(t) = \int_0^t \lambda(x)dx$ is the cumulative hazard function. Cumulative hazard is the sum of the risks you face going from duration 0 to t. The smoothed hazard function is computed from taking the derivative of the cumulative hazard function.

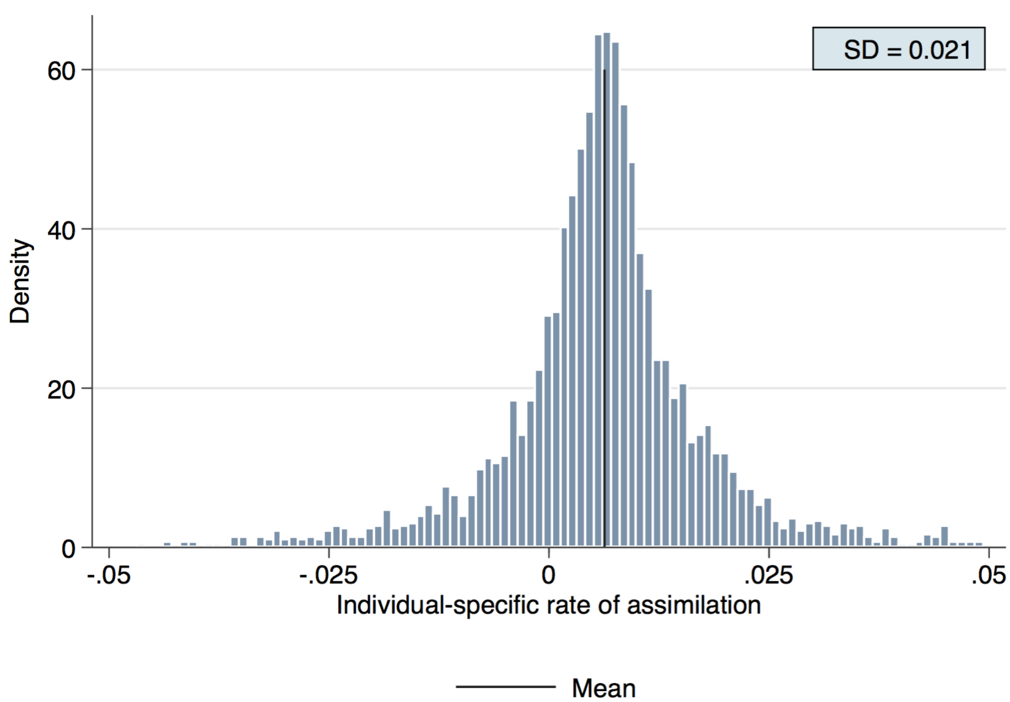
Figure 5: Non Parametric Estimates



Notes: The analysis time begins at age 14 and ends at age 64. Survival function is the opposite of cumulative hazard function. Specifically, $S(t) = exp \int_0^t \lambda(x)dx$ where $\Lambda(t) = \int_0^t \lambda(x)dx$ is the cumulative hazard function. Cumulative hazard is the sum of the risks you face going from duration 0 to t. The smoothed hazard function is computed from taking the derivative of the cumulative hazard function.

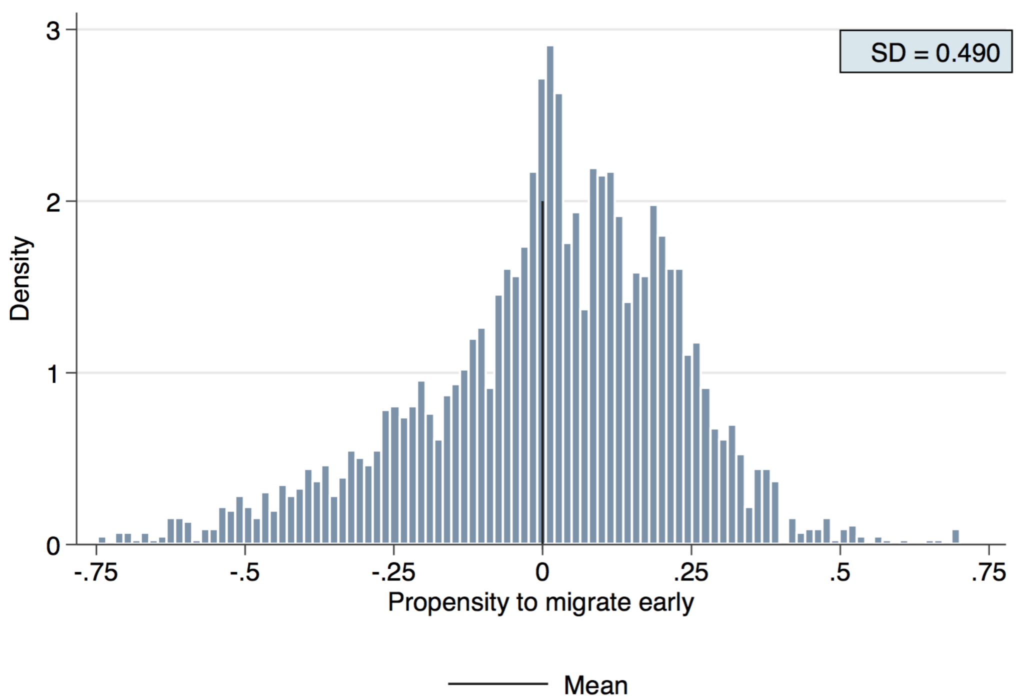
Figure 6: Distribution of Immigrant Quality and Individual-specific Rate of Assimilation: Only Males Specification





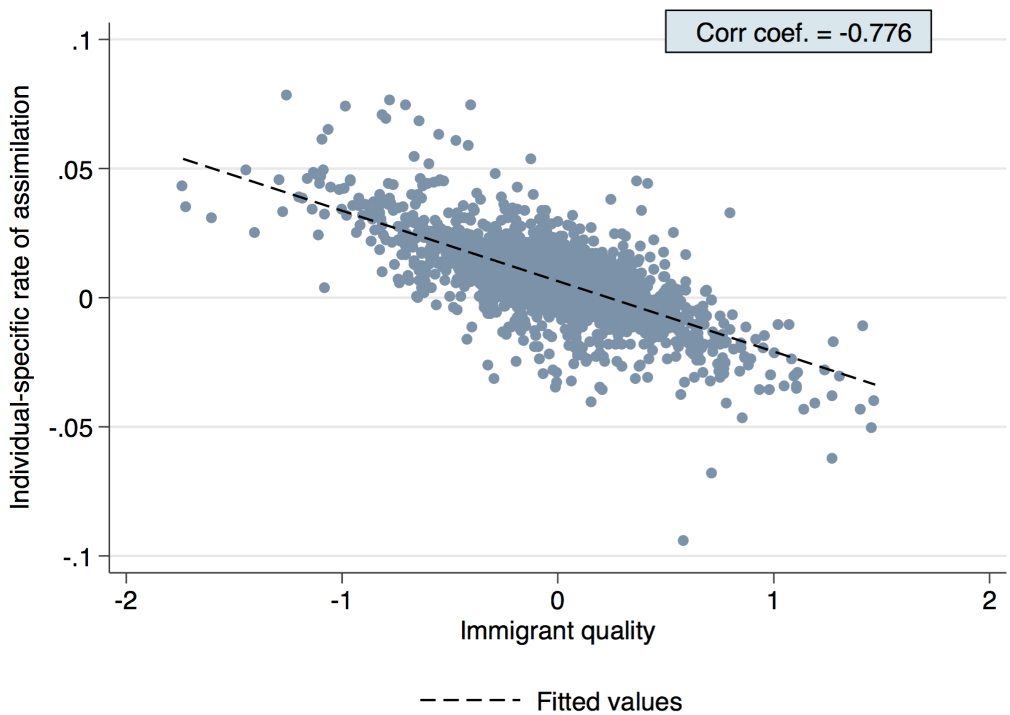
Notes: The distribution was estimated by calculating Best Linear Unbiased Predictions of immigrant quality 'a', individual-specific rate of assimilation '$b + \delta$'

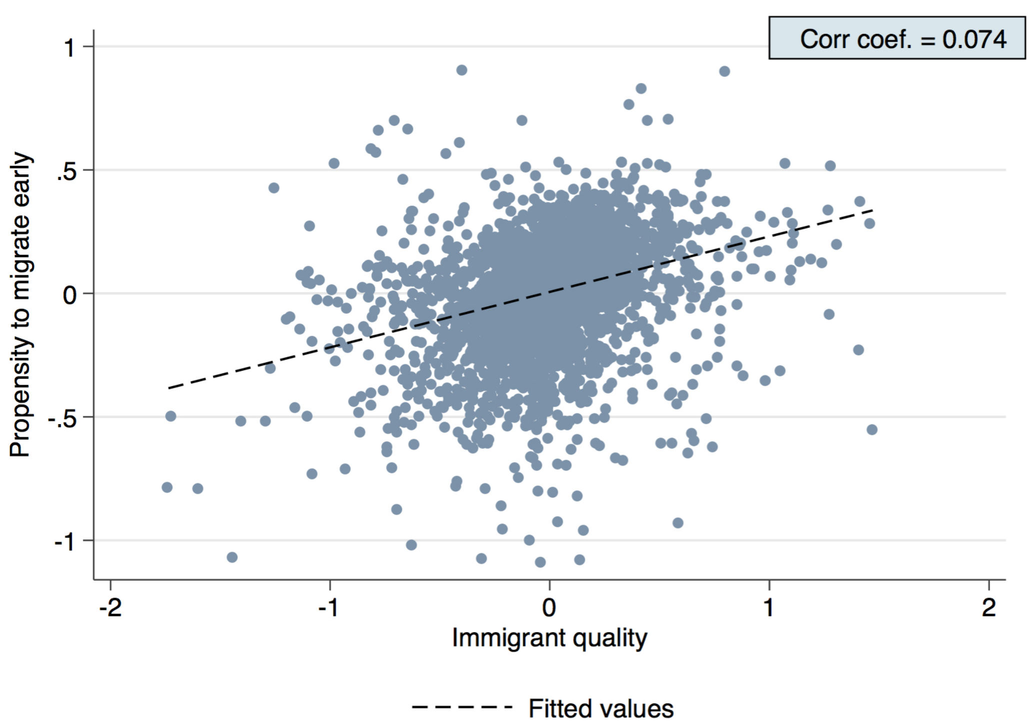
Figure 7: Distribution of Propensity to Migrate Early: Only Males Specification



Notes: The distribution was estimated by calculating Best Linear Unbiased Predictions of propensity to migrate early 'c'.

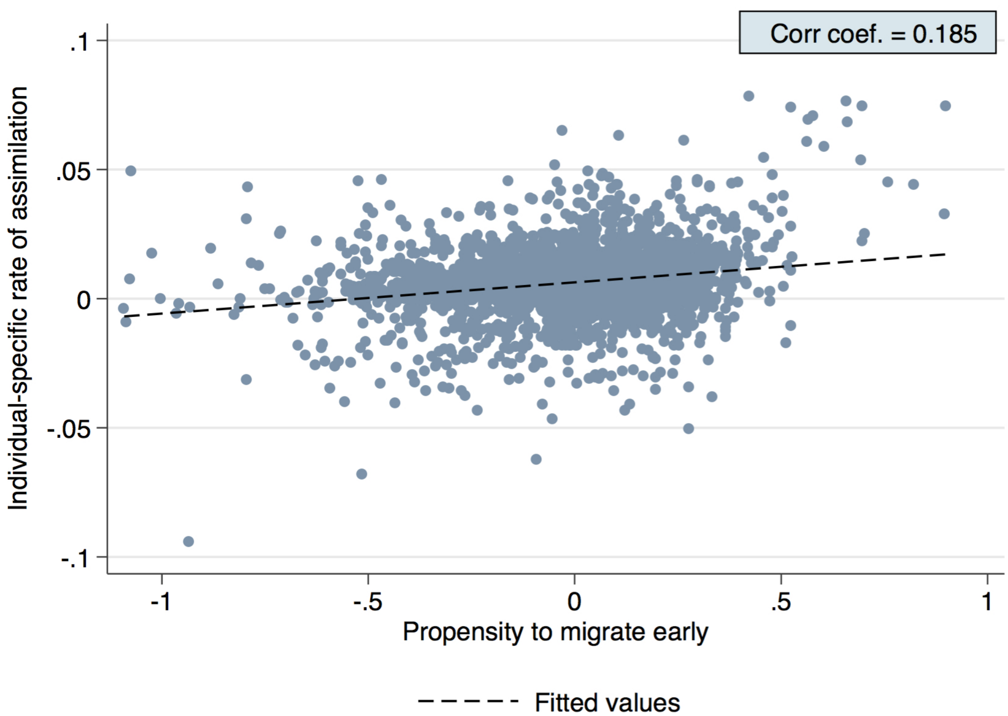
Figure 8: Correlation between Immigrant Quality, Individual-specific Rate of Assimilation, and Propensity to Migrate Early: Only Males Specification





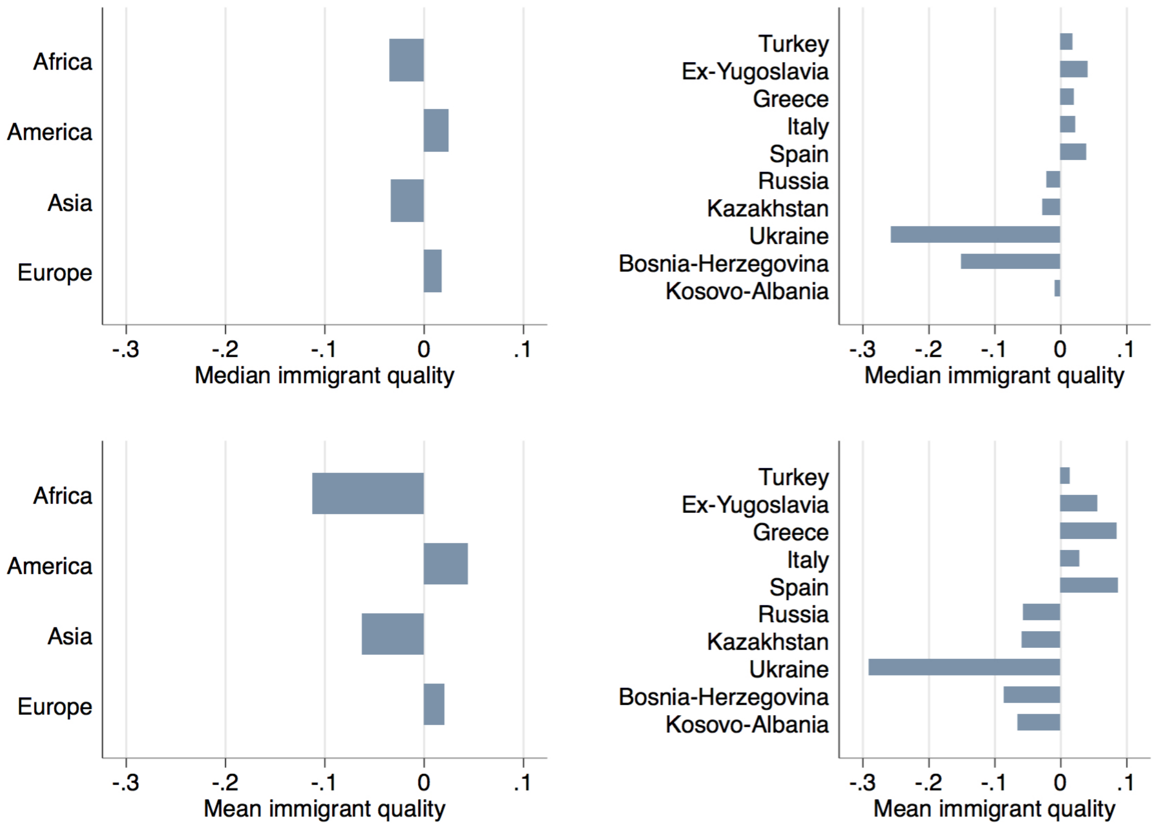
Notes: Scatter plots of Best Linear Unbiased Predictions of immigrant quality 'a', individual-specific rate of assimilation '$b+\delta$' and, propensity to migrate early 'c'.

Figure 9: Correlation between Immigrant Quality, Individual-specific Rate of Assimilation, and Propensity to Migrate Early: Only Males Specification



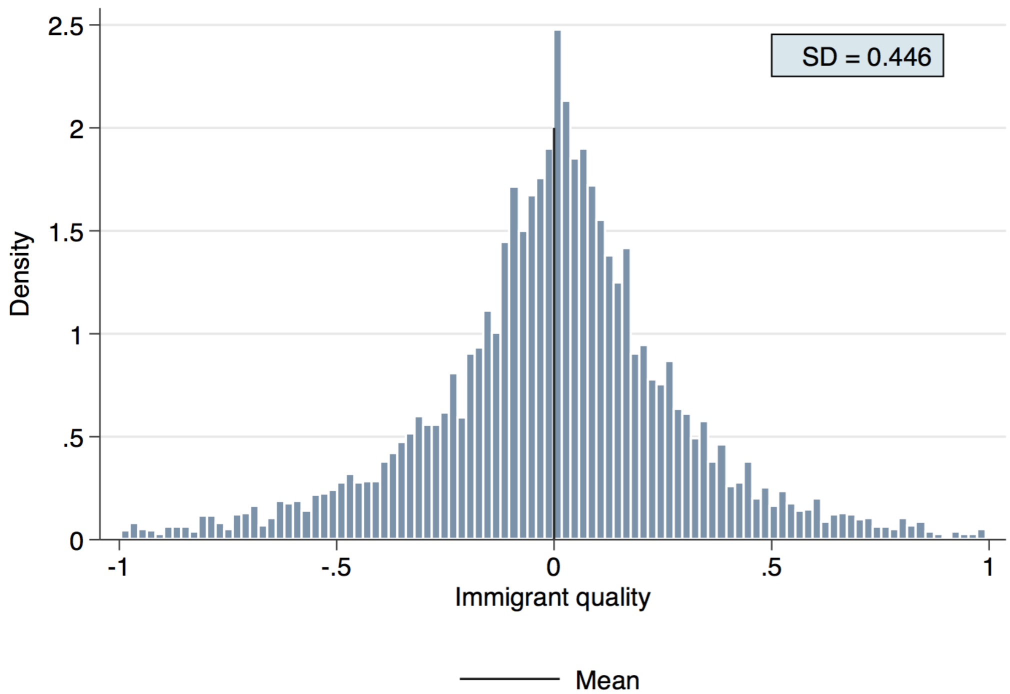
Notes: Scatter plots of Best Linear Unbiased Predictions of immigrant quality 'a', individual-specific rate of assimilation '$b + \delta$' and, propensity to migrate early 'c'.

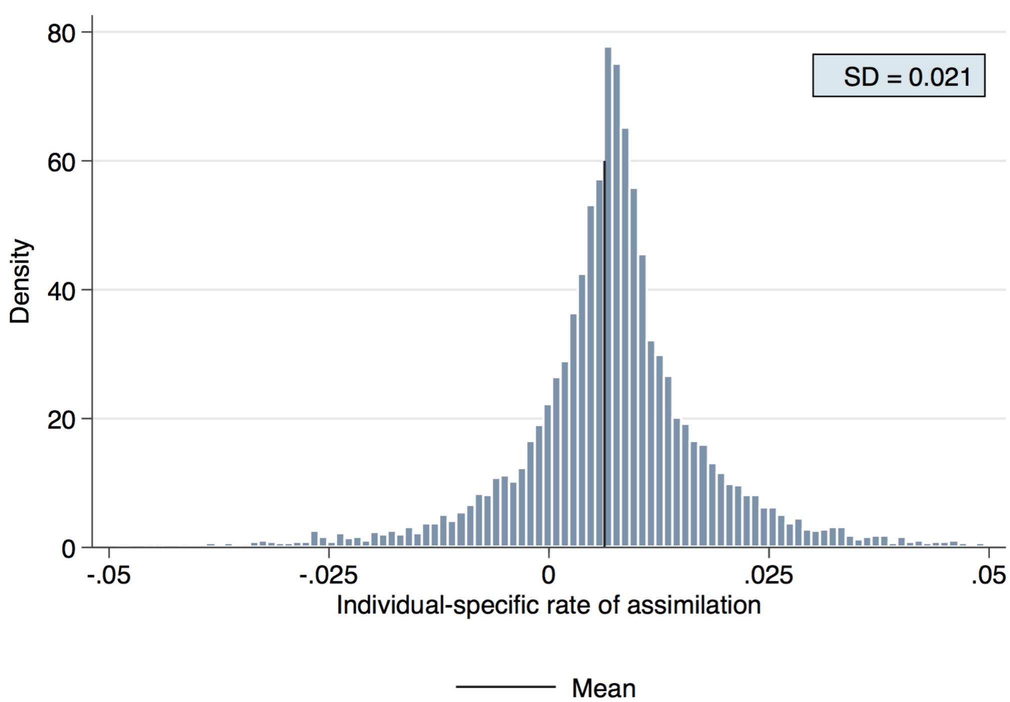
Figure 10: Mean and Median of Immigrant Quality by Continent, Guest-worker Treaty and Refugee Status: Only Males Specification



Notes: Turkey, Ex-Yugoslavia, Greece, Italy and Spain had a Guest-worker treaty with Germany in 1950's. Majority of immigrants from former Soviet states like Russia, Kazakhstan and Ukraine are refugees. Similarly, most immigrants from Bosnia-Herzegovina and Kosovo-Albania fled to Germany after war outbreaks in the origin countries.

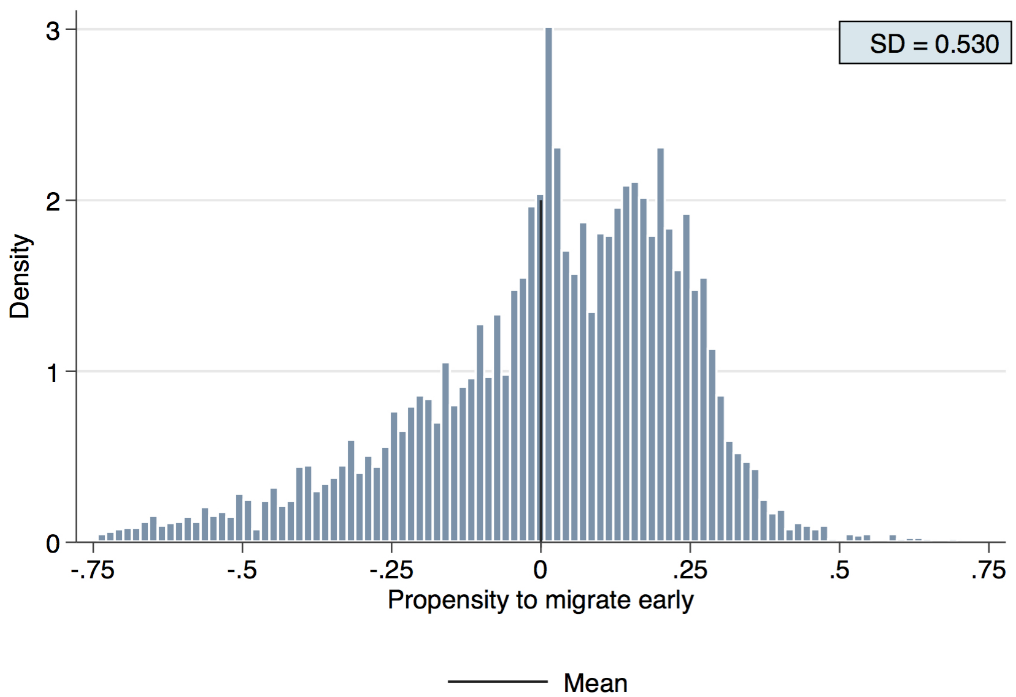
Figure 11: Distribution of Immigrant Quality and Individual-specific Rate of Assimilation: Full Sample Specification





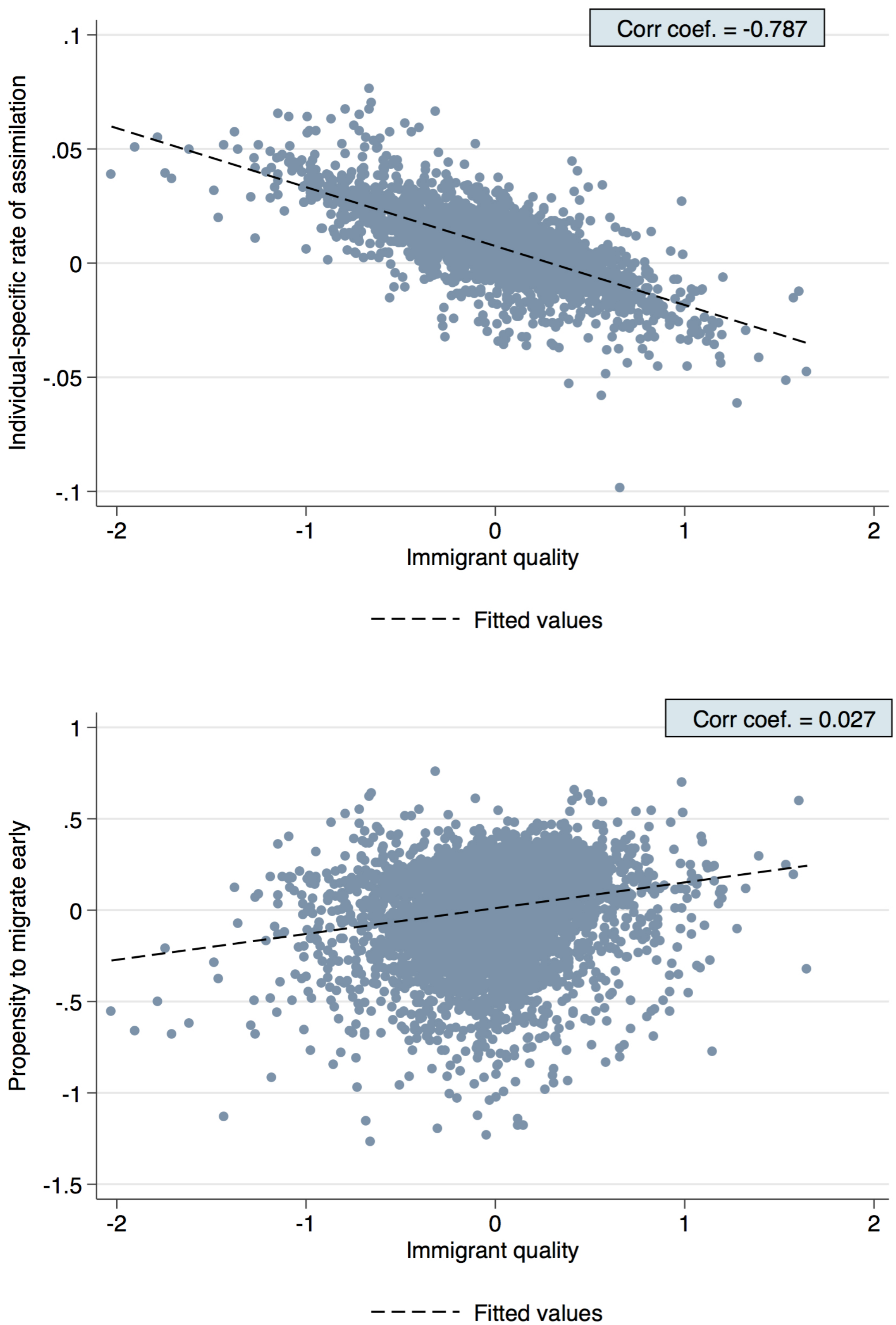
Notes: The distribution was estimated by calculating Best Linear Unbiased Predictions of immigrant quality 'a', individual-specific rate of assimilation '$b + \delta$'.

Figure 12: Distribution of Propensity to Migrate Early: Full Sample Specification



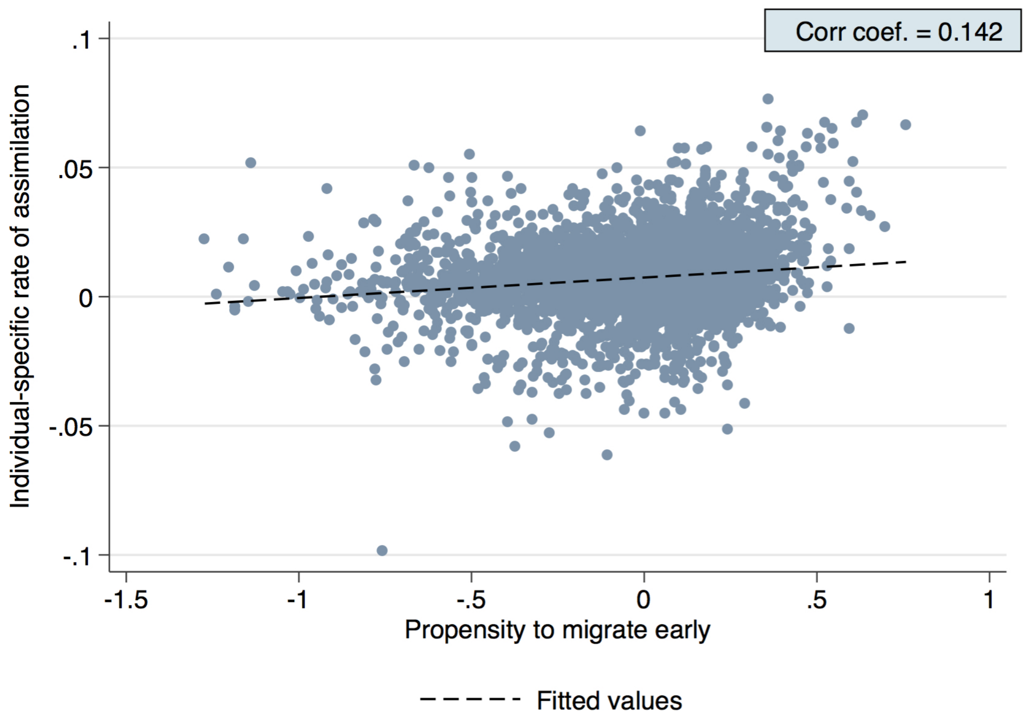
Notes: The distribution was estimated by calculating Best Linear Unbiased Predictions of propensity to migrate early'c'.

Figure 13: Correlation between Immigrant Quality, Individual-specific Rate of Assimilation, and Propensity of Early Migration: Full Sample Specification



Notes: Scatter plots of Best Linear Unbiased Predictions of immigrant quality 'a', individual-specific rate of assimilation '$b + \delta$' and, propensity to migrate early 'c'.

Figure 14: Correlation between Immigrant Quality, Individual-specific Rate of Assimilation, and Propensity of Early Migration: Full Sample Specification



Notes: Scatter plots of Best Linear Unbiased Predictions of immigrant quality 'a', individual-specific rate of assimilation '$b + \delta$' and, propensity to migrate early 'c'.

9 Data Appendix

Construction of Linguistic Distance I construct the Lavenshtein linguistic distance using the Automated Similarity Judgement Program (ASJP) provided by German Max Planck Institute for Evolutionary Anthropology[16]. The ASJP program uses a list of 40 words (similar to Swadesh (1955) list) 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, the ASJPcode and then the normalized divided Lavenshtein distance (LDND)[17] between each word pair of the two languages is calculated. ASJPcode uses only the symbols from QWERTY keyboard and has 7 vowel symbols and 34 consonant symbols.

Levenshtein distance (LD) is the number of consecutive additions, deletions or substitutions required to change one word into the other. Further, dividing each LD by its theoretical maximum yields the normalized LD (LDN). To correct for chance resemblances due to overlap in phoneme inventories or shared phonotactic preferences in the two languages, LDN is then divided by the average LDN of N(N-1)/2 pairings of words with different meanings to produce the final linguistic distance measure of normalized divided Levenshtein distance (LDND). The benefits of this measure is that it can be calculated between any pair of languages, is continuous and provides variation between languages even if they belong to same language families. For comparison, I present both ASJP linguistic distance and the linguistic proximity measure based on language family (LP(Tree)) in Table 14 . Linguistic proximity can only take 4 values: 0, 0.25, 0.5 and 0.75.

Table 14: Languages with Highest and Lowest Linguistic Distance from Standard German in SOEP

Highest			Lowest		
Language	LP (Tree)	LD (ASJP)	Language	LD (Tree)	LD (ASJP)
Korean	0	1.0468	Luxembourgish	0.75	0.4083
Palestinian Arabic	0	1.0332	Dutch	0.75	0.4883
Malay	0	1.03	Afrikaans	0.75	0.595
Arabic Gulf Spoken	0	1.024	Norwegian Bokmaal	0.5	0.6438
Maltese	0	1.0227	Swedish	0.5	0.6979

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

[17]For a detailed description of the LDND measure, refer Bakker et al. (2009)

Variable	*Description*
Log of 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.
Length of stay	Number of years since immigration, or the length of stay in the host country, is calculated as year of survey minus year of immigration.
Work experience	Total length of full-time employment, in years and months. This variable is part of the generated variables for public use; see documentation of generated variables in SOEP (2014b).
Years of schooling	Number of years of education or training. This variable is part of the generated variables for public use; see documentation of generated variables in SOEP (2014b).
Parents' education	The 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 parents 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 respondents in the unknown category is about 5 percent.
Current urban residence	Equals 1 if the induvidual's residence in an urban region as provided in HBRUTTO file.

Variable	***Description***
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 first category is chosen as a base category. The share of respondents in the unknown category is about 6 percent.
Ethnic German	A dummy variable indicating if an immigrant is of German descent from Eastern Europe.
Linguistic distance	The ASJP lingusitic distance is classified into 4 categories (LD1 - LD4): LD1 equals 1 if the linguistic distance is zero; LD2 equals 1 if linguistic distance is between 0.25 and 0.5; LD3 equals 1 if linguistic distance is between 0.5 and 0.75; and LD4 equals 1 if linguistic distance is between 0.75 and 1.
Average commuting distance to work	The average distance (in kilometers) between home and workplace varies by state and year . The variable is constructed using individual reports on commuting distance from home to work available in PL file, which is then averaged at the state-year level. The distance is top coded at 200 km. The information is available for selected years and the values for missing years are taken from the neighboring year: 1984-87 from 1985, 1988-89 from 1990, 1991-92 from 1993 (and 1990 for East Germany), 1994 and 1996 from 1995, 1997 and 1999 from 1998, and 2000 from 2001. After 2000, the question on commuting distance is asked every year. Individuals who have workplace and home in the same building are assigned a zero distance. Individuals whose location of work varies or answered ‘difficult to say are assigned a missing value for the distance.
Unified Germany	This dummy variable equals 1 if the survey year is higher than or equal to 1990.
Guest-worker Treaty	This is a time varying dummy variable that equals 1 if the country has or had in past been in a Guest-worker treaty with West Germany.
Contiguity	This dummy variable equals 1 if the country of origin and Germany share a border.
European Union	This dummy variable equals 1 if the country of origin is a a member of the European Union in the survey year.

Variable	***Description***
Country of origin	Country of origin is defined as Germany if a person is born in Germany or immigrated before 1949. Other 130+ countries of origin are re-coded according to the UN country classification in order 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 of these variables are provided for public use as part of the biography and life history data; see documentation of biography variables in SOEP (2014a).
GDP per capital	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. 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, 2015), we construct a time-series of PPP-adjusted GDP per capita in constant 2011 international dollars.

Variable	***Description***
GPD per capita growth in next 5 years	Based on the above mentioned growth series, the variable is calculated at the average of the growth rate in the next five years.
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 Piece (2015). This dataset assigns an integer score between 0 and 10 to each major episode of the 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 (MEPVCAT): 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 involve a limited use of destruction technologies; population dislocations are in the tens of thousands people; 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.

10 Technical Appendix

10.1 Joint Likelihood Estimation

The likelihood given by Equation 14 can be written as:

$$L(\theta) = \prod_{i=1}^{n} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ \prod_{s=1}^{S} f(W_{is}|T_i, a_i, b_i; \theta_w) \right\} \times f(T_i|c_i; \theta_t) f(a_i|b_i, c_i; \theta_{a|b,c}) f(b_i|c_i; \theta_{b|c}) f(c_i; \theta_c) \mathrm{d}a_i \mathrm{d}b_i \mathrm{d}c_i \tag{19}$$

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. Lets, focus on these expression to eliminate a_i:

$$\begin{aligned}
&\int_{-\infty}^{\infty} f(W_{is}|a_i, b_i; \theta_w) f(a_i|b_i, c_i; \theta_{a|b,c}) \mathrm{d}a_i \\
&= \int_{-\infty}^{\infty} \left\{ \prod_{s=1}^{S} (2\pi\sigma_\epsilon^2)^{-1/2} \exp\{ \frac{-(W_{is} - \beta_0 - \beta_X X_{is} - (\delta + b_i) LOS_{is} - \phi(s) - a_i)^2}{2\sigma_\epsilon^2} \} \right\} \\
&\times (2\pi {\sigma_{a|b,c}}^2)^{-1/2} \exp \frac{-(a_i - \Sigma_{12}\Sigma_{22}^{-1} \begin{pmatrix} b_i \\ c_i \end{pmatrix})^2}{2{\sigma_{a|b,c}}^2} \mathrm{d}a_i \\
&= ((2\pi\sigma_\epsilon^2)^{-1/2})^S \times (2\pi{\sigma_{a|b,c}}^2)^{-1/2} \int_{-\infty}^{\infty} \exp \frac{-\sum_{s=1}^{S} (D_s - a_i)^2}{2\sigma_\epsilon^2} \times \exp \frac{-(a_i - \mu_{a|b,c})^2}{2{\sigma_{a|b,c}}^2} \mathrm{d}a_i \\
&= C \int_{-\infty}^{\infty} \exp \Big(\frac{-\sum_{s=1}^{S} ({D_s}^2 + {a_i}^2 - 2a_i D_s)}{2\sigma_\epsilon^2} \Big) \times \exp \Big(\frac{-({a_i}^2 + {\mu_{a|b,c}}^2 - 2a_i\mu_{a|b,c})}{2{\sigma_{a|b,c}}^2} \Big) \mathrm{d}a_i \\
&= C \times \exp\{ -\frac{\sum_{s=1}^{S} {D_s}^2}{2\sigma_\epsilon^2} \} \times \exp\{ -\frac{{\mu_{a|b,c}}^2}{2{\sigma_{a|b,c}}^2} \} \int_{-\infty}^{\infty} \exp \Big(\frac{-\sum_{s=1}^{S} ({a_i}^2 - 2a_i D_s)}{2\sigma_\epsilon^2} - \frac{({a_i}^2 - 2a_i\mu_{a|b,c})}{2{\sigma_{a|b,c}}^2} \Big) \mathrm{d}a_i
\end{aligned} \tag{20}$$

$$
\begin{aligned}
&= C_1 \int\limits_{-\infty}^{\infty} \exp\Big(-\frac{{a_i}^2}{2}\{\frac{S}{\sigma^2_\epsilon} + \frac{1}{{\sigma_{a|b,c}}^2}\} + a_i\{\frac{\sum\limits_{s=1}^{S} D_s}{\sigma^2_\epsilon} + \frac{\mu_{a|b,c}}{{\sigma_{a|b,c}}^2}\}\Big)\mathrm{d}a_i \\
&= C_1 \int\limits_{-\infty}^{\infty} \exp\Big(-\frac{{a_i}^2}{2}\{F\} + a_i\{E\}\Big)\mathrm{d}a_i \\
&= C_1 \int\limits_{-\infty}^{\infty} \exp\Big(\frac{-F}{2}({a_i}^2 - \frac{2a_iE}{F} + \frac{E^2}{F^2} - \frac{E^2}{F^2})\Big)\mathrm{d}a_i \\
&= C_1 \int\limits_{-\infty}^{\infty} \exp\Big(\frac{-F}{2}{(a_i - \frac{E}{F})}^2 + \frac{E^2}{2F}\Big)\mathrm{d}a_i \\
&= C_1 \times \exp\frac{E^2}{2F} \int\limits_{-\infty}^{\infty} 2\pi{\frac{1}{F}}^{1/2} \Big\{ 2\pi{\frac{1}{F}}^{-1/2} \exp(\frac{-1}{2\frac{1}{F}}{(a_i - \frac{E}{F})}^2) \Big\}\mathrm{d}a_i \\
&= C_1 \times \exp\frac{E^2}{2F} \times (2\pi\frac{1}{F})^{1/2} \int\limits_{-\infty}^{\infty} \Big\{ (2\pi\frac{1}{F})^{-1/2} \exp(\frac{-1}{2\frac{1}{F}}{(a_i - \frac{E}{F})}^2) \Big\}\mathrm{d}a_i \\
&= C_1 \times \exp\frac{E^2}{2F} \times (2\pi\frac{1}{F})^{1/2} \qquad (21)
\end{aligned}
$$

Thus, resulting in the final likelihood:

$$
L(\theta) = \prod_{i=1}^{n} \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} \Big\{ C_1 \times \exp\frac{E^2}{2F} \times (2\pi\frac{1}{F})^{1/2} \Big\} f(b_i, c_i; \theta_{bc}) f(T_i|c_i; \theta_t)\mathrm{d}b_i\mathrm{d}c_i \qquad (22)
$$

where

$$
\begin{aligned}
C1 &= C \times \exp\{-\frac{\sum\limits_{s=1}^{S} {D_s}^2}{2\sigma^2_\epsilon}\} \times \exp\{-\frac{{\mu_{a|b,c}}^2}{2\sigma^2_{a|b,c}}\} \\
C &= ((2\pi\sigma^2_\epsilon)^{-1/2})^S \times (2\pi\sigma^2_{a|b,c})^{-1/2} \\
E &= \{\frac{\sum\limits_{s=1}^{S} D_s}{\sigma^2_\epsilon} + \frac{\mu_{a|b,c}}{\sigma^2_{a|b,c}}\} \\
F &= \{\frac{S}{\sigma^2_\epsilon} + \frac{1}{\sigma^2_{a|b,c}}\} \\
D_s &= (W_{is} - \beta_0 - \beta_X X_{is} - (\delta + b_i)LOS_{is} - \phi(s))
\end{aligned}
$$

$$
\begin{aligned}
\mu_{a|b,c} &= \Sigma_{12}\Sigma_{22}^{-1}\begin{pmatrix} b_i \\ c_i \end{pmatrix} \\
&= \begin{pmatrix} \sigma_{ab} & \sigma_{ac} \end{pmatrix} \begin{pmatrix} \sigma_b^2 & \sigma_{bc} \\ \sigma_{bc} & \sigma_c^2 \end{pmatrix} \begin{pmatrix} b_i \\ c_i \end{pmatrix} \\
&= \begin{pmatrix} \sigma_{ab}\sigma_b^2 + \sigma_{ac}\sigma_{bc} & \sigma_{ab}\sigma_{bc} + \sigma_{ac}\sigma_c^2 \end{pmatrix} \begin{pmatrix} b_i \\ c_i \end{pmatrix} \\
&= (\sigma_{ab}\sigma_b^2 + \sigma_{ac}\sigma_{bc})b_i + (\sigma_{ab}\sigma_{bc} + \sigma_{ac}\sigma_c^2)c_i \\
\Sigma_{12} &= \begin{pmatrix} Cov(a_i, b_i) & Cov(a_i, c_i) \end{pmatrix} \\
&= \begin{pmatrix} \sigma_{ab} & \sigma_{ac} \end{pmatrix} \\
\Sigma_{22} &= \begin{pmatrix} Var(b_i) & Cov(b_i, c_i) \\ Cov(b_i, c_i) & Var(c_i) \end{pmatrix} \\
&= \begin{pmatrix} \sigma_b^2 & \sigma_{bc} \\ \sigma_{bc} & \sigma_c^2 \end{pmatrix} \\
\sigma_{a|b,c}^2 &= \sigma_a^2 - \begin{pmatrix} \sigma_{ab}\sigma_b^2 + \sigma_{ac}\sigma_{bc} & \sigma_{ab}\sigma_{bc} + \sigma_{ac}\sigma_c^2 \end{pmatrix} \begin{pmatrix} \sigma_{ab} \\ \sigma_{ac} \end{pmatrix} \\
&= \sigma_a^2 - \left(\sigma_{ab}^2\sigma_b^2 + \sigma_{ab}\sigma_{bc}\sigma_{ac} + \sigma_{ab}\sigma_{bc}\sigma_{ac} + \sigma_{ac}^2\sigma_c^2\right) \\
&= \sigma_a^2 - (\sigma_{ab}^2\sigma_b^2 + 2\sigma_{ab}\sigma_{bc}\sigma_{ac} + \sigma_{ac}^2\sigma_c^2) \\
\sigma_{b|c}^2 &= \sigma_b^2 - \frac{\sigma_{bc}^2}{\sigma_c^2} \\
\mu_{b|c} &= \mu_b + \Sigma_{12}\Sigma_{22}^{-1}(c_i - \mu_c) \\
&= 0 + \frac{\sigma_{bc}}{\sigma_c^2}(c_i - 0) \\
&= \frac{\sigma_{bc}}{\sigma_c^2}c_i
\end{aligned}
$$