

THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF ECONOMICS

LANGUAGE TRAINING, LANGUAGE PROFICIENCY AND THE EFFECT OF
LANGUAGE PROFICIENCY ON EARNINGS OF MIGRANTS IN GERMANY

SARAH VLAZNY
SPRING 2018

A thesis
submitted in partial fulfillment
of the requirements
for baccalaureate degrees in
Economics and Advertising/Public Relations
With honors in Economics

Reviewed and approved* by the following:

Amanda Mandzik
Assistant Teaching Professor of Economics
Thesis Supervisor

Russell Chuderewicz
Teaching Professor of Economics
Honors Adviser

* Signatures are on file in the Schreyer Honors College.

ABSTRACT

This paper uses a probit model to determine the impact of taking a language course on German language proficiency for migrants in Germany. It also models the income returns to language proficiency for fully-employed male migrants. The results show that those who have enrolled in a language course are significantly more likely to be proficient at writing in German. The results also show that certain ethnic identity factors, such as whether the migrant has visited Germans in their home and whether or not they have been discriminated against because of their ethnicity are significantly important in predicting language proficiency. The results imply that those who feel more assimilated in Germany and who have experienced less ethnic discrimination are more proficient in the German language. The model of income returns to language proficiency uses the language the migrants speaks with their friends and whether or not the migrant has visited a German citizen in their home as instrumental variables to correct for potential endogeneity issues. After this correction, the results show that language proficiency does have significantly positive effect on earnings. However, after adding additional employment variables such as the size of the company and the training required for the job, language proficiency is no longer significant. A likely explanation is that there is simultaneous causality between being proficient at the German language and having a high level of training or working at a large company. Another potential explanation is the absence of a variable to control for whether the migrant works in an urban or a rural area, which could be highly correlated with company size, training required, and earnings.

TABLE OF CONTENTS

LIST OF TABLES	iii
ACKNOWLEDGEMENTS	v
Chapter 1 Introduction	1
Chapter 2 German Migration	4
History of Migration in Germany	4
German Immigration Policy and Language	7
Chapter 3 Literature Review	9
Language Economics	9
The Effect of Language on Learning, Earnings, and Unemployment.....	12
Chapter 4 Data and Descriptive Statistics.....	18
Data	18
Descriptive Statistics	19
Chapter 5 Language Training and Language Proficiency	22
Methodology	22
Results	24
Discussion	27
Chapter 6 Language Proficiency and Earnings.....	29
Methodology	29
Results	32
Discussion	35
Chapter 7 Conclusion.....	37
Appendix A Descriptive Statistics	38
BIBLIOGRAPHY	39

LIST OF TABLES

Table 1. Language Training and Language Proficiency Results24

Table 2. Language Proficiency and Earnings32

ACKNOWLEDGEMENTS

I would like to thank Amanda Mandzik for her help in completing this paper. Professor Mandzik is one of the dedicated faculty that make the economics department great. When I had her as an instructor, I could immediately tell that she had a passion for economics as well as a genuine interest in seeing her students succeed. As a professor, she went above and beyond in providing help and support for students, and she was no different as a thesis supervisor. She took an enthusiastic interest in my subject topic, and worked hard to help me make my paper the best it could be. It is because of her dedication and guidance that I was able to put together a paper that I am proud of.

I would also like to thank my academic advisor, Lisa Brooks, who has been a tremendous resource to me throughout my college career. She is incredibly knowledgeable and helpful, and thanks to her guidance I feel that I was really able to get the most out of my Penn State experience.

Chapter 1

Introduction

Throughout history to the present day, the issue of immigration has remained an enduring issue for nations all around the world. Immigration concerns have been unquestionably influential in elections and political movements. In the United States, a self-proclaimed “nation of immigrants,” immigration has remained a divisive political issue since the country’s founding. The issue again took center-stage during the recent presidential campaign of Donald J. Trump, whose supporters rallied behind his anti-immigration rhetoric (Felter & Renwick, 2017).

Immigration has become an increasingly polarizing issue in Europe as well. In 2016, when the United Kingdom held a referendum vote to determine whether or not to remain in the European Union, immigration was at the forefront of the debate. In fact, according to Ipsos MORI, one week before the vote was held, immigration had surpassed the economy as the most important issue for referendum voters (“Immigration is now the top issue for voters in the EU referendum”, 2016).

Immigration fueled the 2017 French presidential election, when a run-off was needed after no candidate won a majority in the first round. The two candidates had starkly different immigration policies; Marine Le Pen called for “automatic” expulsion of illegal immigrants and a sharp decrease in legal immigration (“French election 2017: The final candidates”, 2017), whereas Emmanuel Macron defended a liberal immigration policy as a moral duty for France (Wolff, 2017). Le Pen lost in the run-off election, but still secured about a third of the French votes (Aisch, Bloch, Lai & Morenne, 2017). This anti-immigration sentiment was echoed more

strongly in Austria, when an anti-immigration party founded by ex-Nazis won the October 2016 elections (Eddy, 2017).

There are numerous other examples of the controversy stirred by the immigration issue around the world. In Chile, presidential candidates are utilizing anti-immigration sentiment to attract voters, and immigrants report being treated as political scapegoats (Esposito & Iturrieta, 2017). The Argentinian president issued a decree making it easier to deport immigrants in 2017, saying that Argentina had to put its citizens first and stop allowing immigrants to commit crimes (Mohor & Haynie, 2017). In South Africa, natives worry that immigrants are taking their jobs and amplifying problems with crime (Brock, 2017). In Australia, historically one of the most immigrant-friendly countries, anti-immigration sentiment is rising with half of Australians reportedly wanting lower levels of migration (Babones, 2018).

Much of the anti-immigration sentiment is economically motivated. Natives are concerned that immigrants take their jobs and are a strain on their country's resources. However, much of the economic literature on the subject fails to support these beliefs (Peri & Sparber, 2011). Though some studies find some adverse effects on natives (Borjas, Freeman, Katz, DiNardo & Abowd, 1997), other researchers assert that the models finding native displacement, particularly those of Borjas, are biased in their specifications (Peri & Sparber, 2011). Many studies have actually found little evidence that inflows of low-skilled immigrants are associated with higher unemployment rates of low-skilled native workers (Altonji & Card, 1991). Studies have also been conducted which examine the impact of highly educated immigrants and found that highly educated immigrants tend to work in different occupations than highly educated natives (Peri & Sparber, 2011). Moreover, some studies actually find that immigrants contribute more positively to the national economy than their native counterparts (Dustmann & Frattini,

2014). Although there are prevailing opinions that immigration always results in worse labor market outcomes for natives, in fact, much of the empirical evidence does not support that claim, and some studies actually find the opposite (Dustmann & Frattini, 2014). In other words, investment in programs which improve the labor market outcomes of immigrants has the potential to improve the welfare of natives as well, and does not automatically translate to worse labor outcomes for native workers. This is important to note in the context of this paper, especially when considering the policy implications of these findings.

There are countless issues to be considered when discussing the economics of immigration and immigration policy; however my goal is to limit the scope of this paper. Specifically, I will examine the relationship between the earnings of adult migrants in Germany and the immigrants' knowledge of the German language. Additionally, I hope to provide some insight as to the potential economic benefits of investment in language training programs. My study uses data from the German Socio-Economic Panel, which is a wide ranging longitudinal study of private households. The dataset will be described further in section Chapter 4.

Chapter 2

German Migration

History of Migration in Germany

A brief overview of the history of immigration in Germany will provide context for this analysis. Germany has been both a country of emigration and immigration at various times in its history. In the 19th century, many Germans immigrated to countries such as the United States and the United Kingdom. Reasons for emigration ranged from political differences, religious conflicts, and lack of opportunity. From the period of 1820 to 1920, it is estimated that six million emigrants left Germany ("Migration History in Germany | domid.org", 2017). This outflow changed somewhat in the early 1900s, as Germany entered the industrial era and began to become more economically prosperous. The first major group to immigrate to Germany during this time were Polish workers who worked in the mining sector ("Germany: Immigration in Transition", 2017). The coal and steel industries brought foreign workers as well, and the early 1900s was the turning point where there were more people entering Germany than leaving the country.

Between 1900 and 1945, many of the foreigners who immigrated to Germany were put in forced labor camps under the national-socialist regime ("Migration History in Germany | domid.org", 2017). The local population generally turned a blind eye to this mistreatment of immigrants. From 1941 to the end of 1944, the Nazi regime operated six death camps in Eastern Europe. Altogether, 1.6 million people were killed at Auschwitz alone. The victims included Jews, Polish Catholics, Gypsies, and Russian prisoners (Mintz, 2018).

In the years between 1945 and 1949, millions of those who had been previously expelled from Germany, as well as other refugees, flocked back to the country. Though this caused tension between the immigrants and local residents, tensions were eased somewhat by the post-war economic boom. By the end of the 1940s, the immigrant inflow had slowed and the demand for labor had grown so much that Germany had a labor shortage.

In fact, during the following decade unemployment shrank drastically-- from 11% in 1950 to under 1% in 1961 ("Migration History in Germany | domid.org", 2017). The labor shortages became so severe that the federal government started to actively recruit foreign workers to come to Germany temporarily. The program started as an agreement with Italy, but soon expanded to incorporate Greece, Spain, Turkey, Morocco, Portugal, Tunisia, and Yugoslavia. Immigrants from these countries were considered "guest workers," and were treated as temporary workers rather than permanent immigrants. This policy was advantageous for native Germans in that these "guest workers" took unskilled work that allowed natives to move into more skilled positions with higher wages.

Despite the economic advantages of the foreign workers, the economic crisis of 1966 made West Germans reconsider the policy of employing foreigners. This recession led to paranoia and panic which in turn resulted in a recruitment ban in 1973. This ban, which prevented the entry of "guest workers" from non-EEC countries (European Economic Community), ended up having the opposite effect than the original intentions of the ban. Instead of the guest workers eventually returning to their country of origin, the ban made them worry that they would not be able to return if they left. This made them decide to stay in Germany permanently and resulted in many of their family members coming to the country to join them.

In the 1990s, immigration spiked again due to the many political changes of the decade throughout Germany. Because of this, immigration became an important political issue as growing xenophobia and racism caused resistance and conflict with immigrants. In 1993, a new policy was imposed that essentially made it impossible for refugees to enter Germany by land. It did this by declaring that those trying to enter Germany from “safe” countries would be denied entry, and deemed “safe” all the countries that border Germany.

The next groundbreaking immigration policy came in 2000, when a new citizenship law allowed for dual German citizenship. The law allowed children born in Germany to non-German parents to hold a German passport in addition to the passport of their parents’ country of origin, the stipulation being that the child must choose which nationality to keep before they are considered a legal adult. Dual citizenship could only be kept if the child’s parents are citizens of the EU or from nations with which Germany has a special agreement. This dual citizenship reform had a profound effect on the demographics of Germany, as tens of thousands of children born to non-German citizens were no longer considered foreigners after 2000.

The next major change in immigration came in 2005, when Germany declared itself a country of immigration through a new immigration policy. The new law required formal integration as a legal duty, and required language courses for the first time. The first naturalization test was introduced in 2008, which required immigrants to receive a passing grade to be granted German citizenship. This test was another important step in the policy of integration because it required people to familiarize themselves with German history, laws, and culture. It also raised the German language level requirement.

Immigration to Germany has been high in recent years, and the country has seen an inflow of qualified professionals, students, and asylum seekers alike. In August 2015,

Germany suspended the Dublin Procedure for Syrians, which meant that Syrian refugees no longer had to be sent back to the first country in the European Union that they entered.

Chancellor Angela Merkel declared it a “national duty” to grant protection to these refugees, and by the end of 2015, 890,000 refugees had come to Germany (www.dw.com), 2017). As of August 2017, a record-breaking total of about 18.6 million people with foreign roots live in Germany. As a percentage of the total population, this means that over one in five people in Germany have a migrant background (www.dw.com), 2017). According to the German Federal Statistics Office, the vast majority come from other European countries and Turkey, yet the country is seeing the greatest increase in the number of migrants from the Middle East and Africa.

German Immigration Policy and Language

To further contextualize the study, it is important to conduct a brief overview of the role of German language requirements in the country’s immigration policy. German language skills are an important requirement for most migratory statuses. For anything above a residence permit, which is only issued for a limited time, proof of “adequate German-language skills” is required (“BAMF - Bundesamt für Migration und Flüchtlinge - Important information”, 2017).

After holding a resident permit for at least five years, migrants may apply for a settlement permit. A settlement permit does not have a time limit and does require that the immigrant have adequate German language skills. Naturalization also requires adequate German language skills, sometimes in the form of a certificate called the “DTZ- German test for immigrants.”

The DTZ includes both written and oral tests. The written aspect of the exam lasts for 100 minutes and includes both listening and reading exercises, and requires the migrant to write a short letter. The oral part of the examination is a 15-minute discussion with a proctor of the exam and another test taker. Based on the applicant's overall score, he will receive a language certificate that corresponds to the level of his language ability.

The most common way applicants prepare for the DTZ exam is through integration courses. The German government offers and often requires integration courses designed to teach immigrants German customs, culture, and language. Depending on migrant status (i.e. foreign national, asylum applicant, or ethnic German resettler), the migrant may be required to complete an integration course. The integration courses are usually about 660 hours and cost the migrant around 1.95 Euros per lesson which amounts to EUR 1,287 total. However, cost exemptions are possible for those that receive unemployment benefits or have a prohibiting personal or financial situation.

Chapter 3

Literature Review

Language Economics

When making comparisons to industrial organization or labor economics, the field of language economics has received comparatively little attention and study. However, there has been a significant increase in the plurality and prominence of papers studying the economics of language in some capacity. Language economics, which studies the relationships between linguistic and economic variables, includes many research studies wide-ranging in nature (Grin, 2008). Because there have been numerous studies that have in some way combined sociolinguistics and economic factors, I will limit my scope of “language economics” to include only the more rigorous economic-focused analyses, as has been standard in the field (Gazzola, Grin & Wickström, 2015).

Language economics examines how linguistic and economic variables influence each other. This may be done with either the economic or linguistic variables as the dependent variable, for example looking at how language knowledge affects trade patterns or how GDP affects government investment in foreign language learning. Language economics, especially in recent decades, has also come to mean exploration into various language policy issues (Grin 2008). In their bibliography, Gazzola, Grin & Wickstrom broke down the field into topics and subtopics, which I will briefly review in order to contextualize my research within the broader field.

The first section of the Gazzola, Grin & Wickstrom bibliography includes papers that are essentially overviews of language economics as a whole and cover all the subdivisions of

economic research. These papers include anything from general frameworks of how to access the economics of language to the recent trends in language economics study. Each of them discuss the economics of language broadly, in addition to including a more narrow discussion about specific components of language economics. These papers include everything from Mora's work on the economics of language in the U.S. labor market to Harris's paper on language economics in a global economy.

The second section includes research on the evolution and structure of language corpus. This research focuses on everything from applications of game theory in linguistics to optimization in language planning. One key concept relevant to discussion later in this paper is the idea of linguistic distance. Linguistic distance refers to the difference in vocabulary, grammar, written form, syntax, and many other linguistic characteristics from one language to another. For example, the linguistic distance between English and Chinese is much greater than the "distance" between English and Spanish. While it is easy to know intuitively that Spanish is "closer" than Chinese, the difficulty arises when trying to quantify just how much closer it is, and when trying to compare a language like Chinese to a language like Japanese to see which is closer to English. Chiswick & Miller outline a method for quantifying linguistic distance using test scores at the U.S. Department of State School of Language Studies (Chiswick & Miller, 2005). Though this study was done using native English-speakers as a benchmark, it is still a useful tool when considering linguistic distances of immigrant languages in Germany. It also suggests that the immigrant's mother tongue is an important factor to control for when evaluating language proficiency, which is something I will incorporate into my analysis.

The final and largest division of language economics refers to issues that arise because more than one language is present. This division incorporates many subtopics, using the influence of linguistic variables on economic variables and vice versa.

The influence of linguistic variables on economic variables refers to everything from the study of the economic value of linguistic diversity and bilingualism to the effect of linguistic attributes on Macro-level economic indicators. Examples of studies include the impact of language proficiency on immigration, how language may influence trade patterns, and language as a determinant of investment decisions.

Arguably, the most important and widely researched subsection of the study of linguistic variables on economic variables concerns language, earnings, and the labor market. Due to the relevance of this research to this paper, critical studies in this field will be discussed extensively in the following section. More broadly, past studies have looked at language skills and the connection to indigenous labor, migrant labor, and human capital and production. Additional topics that have been explored but that are not as directly relevant to the topic of discussion are the relationship between language and nationalism, and a cost benefit analysis of language policies.

In the other division of research that focuses on the opposite relationship of economic variables on linguistic variables, research has been conducted concerning the economics of minority languages, the effectiveness of alternative language policy, and the economic variables considered when predicting language use and learning decisions (Gazzola, Grin & Wickstrom, 2015). There have also been a few papers written about the economics of translation, which explores everything from the demand and supply of translation as well as the costs and inefficiencies resulting from translation needs (Gazzola, Grin & Wickstrom, 2015).

The Effect of Language on Learning, Earnings, and Unemployment

The most widely researched area of language economics concerns the effect of language skills on the labor market. Many of these studies concern Hispanic immigrants in the United States, such as Chiswick's study concerning the determinants of English language skills among immigrants and the effects of language skills on earnings (Chiswick, 1991). Another popular area of research analyzes data on bilingualism in French-speaking and non-French-speaking areas of Canada, such as Carliner's paper examining the wage premium for being bilingual (Carliner, 1981).

For the purpose of this literature review, I will focus primarily on the research that relates the most directly to the focus of my paper. Aldashev et al.'s study of language proficiency, earnings, and labor market participation is particularly relevant because it uses the same German SOEP data in the analysis (Aldashev, Gernandt & Thomsen, 2008). The authors found that language proficiency affects the likelihood of an individual participating in the labor market and finding employment, in addition to affecting the earnings of the individual. They concluded that language proficiency increases participation and employment probability significantly, as well as affects the selection of people into economic sector and occupation. However, they found no impact of language ability on earnings when making the distinction between economic sectors and occupation.

Interestingly, this study uses information on language usage in the household instead of individual language command data as a means of correcting for problems in measurement error. This was due to a study in which Dustmann and van Soest find that individuals over-report their language abilities, resulting in downward biased estimates of language proficiency on earnings (Dustmann, van Soest, 2001). However, upon performing a

sensitivity analysis using self-reported language information, they found robust results no matter which variable they used.

Another interesting aspect of this study is that it controls for language skills, education, socio-demographics and separates by economic sector and occupation, however does not incorporate variables concerning language training and language use, which will be a key component of my analysis.

In another relevant study, Dustmann and Fabbri examine the determinants of immigrant language fluency and its effects on earnings and employment (Dustmann & Fabbri, 2003). Using data from the UK, they conclude that language proficiency has a positive effect on employment outcomes. Of particular interest to my study is their discussion of language proficiency determinants and potential endogeneity in language proficiency and employment outcomes. To address the endogenous choice of language acquisition, they use a matching estimator; to address the bias due to measurement error they used an instrumental variable estimator. After correcting for endogeneity and measurement error bias, they found that the bias of these two problems actually points in opposite directions, with a slightly greater downward bias (Dustmann & Fabbri, 2003). This suggests that though these corrections may be valuable, they may not be necessary additions to a simple OLS analysis if we consider those results to be modest estimations. Another important contribution from this study concerns the variables used as determinants of language ability. In addition to basic demographic and educational variables, the study controls for the immigrant's country of origin, which will be important for my analysis.

Chiswick et al. likewise address the problem of endogeneity and language hypothesizing that language fluency is a function of three fundamental variables: language exposure, second language acquisition efficiency, and the economic benefits of language fluency

(Chiswick & Miller, 1995). This theoretical equation is complemented by empirical elements such as the expected wage increment for language fluency, age of the individual, and the individual's duration in the host country. This theoretical model can serve as a guiding principle when choosing language proficiency determinants, as Chiswick et al. found robust empirical results when analyzing census data from Australia, Canada, Israel, and the United States (Chiswick & Miller, 1995). The study therefore concludes that language skills respond to incentives, and that they have an important effect on the labor market. However, the study did find that language fluency is endogenous to earnings, suggesting that they are determined jointly (Chiswick & Miller, 1995).

Chiswick et al. also provide an important explanation for why language skills can be treated as a form of human capital. Language skills satisfy the basic requirements for human capital in that they are productive in the labor market or in consumption, they require a sacrifice of time or resources to be created, and they are embodied in the individual (Chiswick & Miller, 1995). It is important to establish language as human capital so as to allow the use of Mincer's model in guiding analyses of immigrant economic performance, as I will do in my analysis and as many others have done in this field of research.

Constant and Zimmermann analyze another important dynamic to consider in my analysis of German language proficiency and labor market outcomes: ethnic identity. They define ethnic identity using variables related to language, culture, societal interaction, history of migration, and ethnic self-identification. They use the same German SOEP data that I will use in my analysis to determine the level of the immigrant's cultural and ethnic integration, identifying four levels of identification with the host state: assimilation, integration, separation, and marginalization. Interestingly, they find ethnic identity to be significant in labor market

participation, but not with regards to earnings (Constant & Zimmerman, 2009). Their research is also applicable to my analysis because I will be considering some of the same characteristics and variables, just in the context of language proficiency rather than ethnic integration.

Though no study has been conducted which estimates the employment effects of language training in Germany, Fitzenberger and Speckesser were able to estimate the effects of one important type of public sector sponsored training in Germany on employment. Using administrative data, rather than survey data, they calculated the average treatment effect of the specific professional skills and techniques (SPST) course. The SPST course is just one of the courses offered by the Federal Employment Office (Bundesagentur für Arbeit); German language courses for immigrants are also offered by the office but not included in this study. The study finds a negative lock-in effect for a short period after the beginning of the program, due to the migrant taking time from their job search to participate in the program, but a significant positive effect on employment rates of at least 10% after one year (Fitzenberger & Speckesser, 2006). Though limited in scope, the study does paint a positive picture of public sector sponsored training, which may be important in any discussion of the policy implications of my analysis.

Dustmann also uses German SOEP data to analyze the determinants of language skills among immigrants and the effect of language skills on their earnings. He employs a probit analysis of German speaking and writing fluency for males and females using education, employment, training, and demographic variables. After analyzing the effect of language proficiency on earnings, he concludes that language has an important impact on migrant earnings (Dustmann, 1994). Specifically, he finds that writing proficiency is more important for labor market outcomes than speaking proficiency.

Dustmann and Van Soest provide a significant contribution regarding bias issues in estimating the effect of language proficiency on earnings using ordinary least squares regressions. They address three main potential problems with these regressions. The first problem is the fact that language proficiency could be affected by the same unobserved individual characteristics as earnings, which would lead to an overestimation of language's effect on earnings (Dustmann & van Soest, 2002). The second is potential measurement error in self-reported language measures, which would lead to a downward bias in language effects. This is because in self-reported measures of language, there is often unsystematic measurement errors that are independent over time (Dustmann & van Soest, 2002). The third potential bias is measurement error that is typical in data with variables with subjective standards, which, assuming the error is independent of unknown wage variables, would also lead to a negatively biased estimate (Dustmann & van Soest, 2002). Dustmann and van Soest address these biases in the German SOEP data using fixed effects models of language proficiency, IV estimations, and additional variables as instruments to explore this endogeneity. They find that the downward bias tends to be much greater than upward bias, and conclude that language may be much more important as a determination of earnings than OLS regressions suggest (Dustmann and Van Soest, 2002). They also suggest that their findings may be generalized to other datasets and other countries. Because they too worked with German SOEP data, it is even more likely that the OLS regressions in my study will suffer from measurement error bias in the same way.

Surprisingly, relatively little research has been done regarding the effect of language training on language skills in the context of labor market outcomes. One study by Hayfron uses Norwegian survey data to estimate the impact of language training on the language proficiency of Third World immigrant men in Norway. The study distinguishes between literacy

and fluency, and estimates a simple probit model using demographic variables and dummy variables indicating whether the individual participated in the language program and whether they completed the language program (Hayfron, 2001). The paper found a significant effect of language training on language proficiency, but unexpectedly found no effect of language proficiency on earnings. Hayfron suggests that this finding may be insignificant because immigrants need language proficiency to enter the Norwegian labor market in the first place, but that once they have found employment, their wages are not impacted by their language skills. However, the study does find that immigrants in earlier waves have higher wages than more recent immigrants, which could be due in part to higher language proficiency (Hayfron, 2001).

Overall, this paper seeks to fill several substantial gaps in related literature. First, it seeks to incorporate language-training variables as determinants of language proficiency, which surprisingly few papers have done. Second, it seeks to update and expand upon the analysis by Dustmann and Aldashev et al. by controlling for more demographic and educational variables. Third, following the work of Constant and Zimmermann, it will recognize the importance of ethnic identity as a determinant of language outcomes, and control for more cultural integration variables. Finally, it seeks to provide the basis for future research regarding language training policy and the economic implications of language training.

Chapter 4

Data and Descriptive Statistics

Data

This analysis uses data obtained from the German Institute for Economic Research (DIW). Specifically, the dataset is the 2015 IAB-SOEP Migration Sample, a joint project of the Institute for Employment Research (IAB) and the Socio-Economic Panel (SOEP). The sample of about 1,500 participants was drawn from the IAB's "Integrated Employment Biographies." This database consists of all employees, unemployed persons, and participants in government initiatives to promote unemployment in Germany. The target population of the survey is recent immigrants to Germany and serves as a continuation of the migration sample that was updated in 2013. The 2013 update expands on previous German migrant data by substantially increasing sample size and including new groups of questions providing more information on earnings, migration, and biographical data.

The following sampling method was used for collecting the data. First, the population was restricted to people who first appeared in the "Integrated Employment Biographies" (IEB) database after 1994. Then, they clustered available address information into groups of 2,500 people to create geographically distinct primary sampling units. From there, they limited the population members to those with a migration background and randomly selected 80 addresses from each of the primary sampling units. They also employed a disproportional sampling scheme and a simulated random walk procedure to find members of the population that live close together and to ensure that chosen migrant groups were represented. The over-sampled populations that were the focus of the sample include migrants from Poland, Romania, the CIS

countries, Turkey, Italy, Spain, Greece and predominantly Muslim countries. These groups were selected at a rate higher than their fraction of the migrant population to make the migrant sample more diverse. After a short screening to ensure that chosen individuals were in fact members of the target population and they had consented to participate, in-person interviews were conducted using the IAB-SOEP Migration questionnaire. The questionnaire is available in five languages and is intended to provide a complete picture of the migrants' life from social, economic, and cultural perspectives.

The 2015 IAB-SOEP Migration Sample includes about 4,500 observations, consisting of previously sampled individuals who had migrated to Germany between 1995 and 2011 adding a new sample of migrants arriving in Germany after 2010. After I limited the sample for this analysis to observations which had information about two of the most important variables: income and whether or not they participated in language training, the sample still includes 1,400 observations.

Descriptive Statistics

My analysis will consist of two parts. The first will be a probit model used to determine the effects of language training on German language proficiency, and the second will be a regression of earnings on education and experience variables in addition to language variables.

The variables most relevant to the probit model are Language Proficiency and Language Course. *Langprof* is a generated variable with the value of 1 if the migrant is proficient in German and a value of 0 if the migrant is not proficient in German. This dummy variable is

derived from *writegerman*, which measures self-reported writing ability. I counted German language proficiency as equal to 1 if the reported writing ability was “good” or “very good” and 0 if it was reported as “fair” “poor” or “not at all.” I made this distinction because studies have shown that writing abilities are most important for employment, especially for migrants, and people tend to over report their abilities (Dustmann, 1994; Aldashev, Gernandt & Thomsen, 2008). Therefore, “fair” would not be considered proficient.

Langcoursebi is a variable generated from *langcourse* equal to 1 if the migrant had taken a language course at all, 0 if not. The original *langcourse* variable differentiated between those who had taken a German language course in their home country, those who had taken one in Germany, those who had taken one in both Germany and their home country, and those who had not taken one at all. For the purposes of this analysis, I will focus only on whether they have taken a language course, and ignore the location of the language course.

Other variables that will be included in my language proficiency model will be basic demographic variables such as age (*age*), relationship status (*relationship*), years since migration (*yearsaway*), years of education (*schoolyears*) and whether or not the migrant is literate in their native language (*literateorigin*). I expect all of these variables to be positive and significant except for age and relationship status, which I expect to be negative and significant (Dustmann & Fabbri, 2003; Hayfron 2001).

The variables most relevant to my earnings regression are *larnings*, *schoolyears*, *yearslabormkt*, *sqyearslabormarket*. This group of variables is based on the Mincer earnings function, which has been used as a basis for much of the analysis regarding the effect of language on earnings (Chiswick & Miller, 1999, 2002, 2009; Dustmann & Fabbri, 1994, 2003; Hayfron, 2001). The *larnings* variable is the natural log of *grossincome*. *Grossincome* is

defined as the migrant's income for the previous month before the deduction of taxes and social security, or estimated monthly income before taxes if the migrant is self-employed. It does not include extra income such as vacation or back-pay, but does include overtime pay. The average gross income is 2603.92 euros with a standard deviation of 1476.40. *Schoolyears* is the self-reported number of years the migrant was in school, with an average of 11 years and a standard deviation of 2 years. *Yearslabormkt* is a variable generated from the migrant's age in 2015, when the survey was conducted, and the self-reported age the migrant started working. The average number of years in the labor market is 16.5 with a standard deviation of 9.56. It is expected that the first three of the Mincer variables will have a positive impact on earnings. The final variable, *sqyearslabormarket*, is a quadratic variable that should exhibit diminishing returns from the number of years in the labor market over time.

A second iteration of the earnings model will also include additional employment variables, including company size and training required. *Companysize* refers to the estimated number of workers at the company of the migrant's current job, and incorporates not only the local unit of the company but the company as a whole. Company size is an ordinal categorical variable with values one to seven, with a value of 1 corresponding to <5 workers and 7 corresponding to >2,000 workers. *Trainingreq* refers to the training required for the migrant's current job. Training required is also an ordinal categorical variable, with 1 referring to no training required and 4 referring to a completed university education.

A table with all of the relevant variables, their possible values, and their sample characteristics can be found in Appendix A.

Chapter 5

Language Training and Language Proficiency

Methodology

To measure the effects of participation in a language course on language proficiency, both a linear probability model (LPM) and a probit model are shown. While the LPM lends itself to more straightforward interpretation, the shortcoming of the LPM is that it is possible to get an expected probability less than 0 or greater than 1, both of which are not possible. This issue is corrected with a probit model, which always results in a predicted probability between 0 and 1. Though the coefficients obtained from a probit model do not lend themselves to straightforward interpretation, by calculating the marginal effect for the “average” person in the data we can determine how much more likely those who have taken a language course are likely to be proficient in German.

Two variations of the model were conducted, beginning with a model using the fundamental demographic variables and basic language variables included in the language proficiency literature, (i.e. the models of Dustmann & Fabbri, 1994 and Hayfron, 2001). The model is:

$$langprof_{ij} = \alpha_0 + \alpha_1 langcourseb_{ij} + \alpha_2 L_{ij} + \alpha_3 Z_{ij} + v_{ij}$$

where L is a vector of language variables, such as literacy in the migrant’s origin country and language spoken at work, Z is a vector of demographic variables such as age and school years, α are estimated parameters, and v is the error term.

The next variation of the model included ethnic identity variables which are not common in many language proficiency models, with the exception of an ethnic concentration variable from Dustmann and Fabbri 2003. The model is as follows:

$$langprof_{ij} = \alpha_0 + \alpha_1 langcoursebi_{ij} + \alpha_2 L_{ij} + \alpha_3 Z_{ij} + \alpha_4 E_{ij} + v_{ij}$$

where E represents a vector of ethnic identity variables, such as whether the migrant has visited a German citizen in their home in the past 12 months, and whether they trust Germans more than foreigners.

For the language proficiency analysis, the sample size includes all adult migrants with non-missing values for the relevant variables. This differs from the sample size that will be used in the earnings analysis, where the sample size will be limited to adult migrant males working full-time. The reason for this difference is the expectation that a more heterogeneous sample size is less likely to have an impact on language proficiency than it would on earnings, and the more observations included in the sample the more robust the results. To verify this assertion, the model was run again with the smaller sample size, and no significantly different results were found.

Results

Table 1. Language Training and Language Proficiency Results

Language Proficiency	LPM 1	PROBIT 1	LPM 2	PROBIT 2
age	-0.0050** (0.0020)	-0.0147** (0.0058)	-0.0055** (0.0021)	-0.0154** (0.0061)
sex	0.0997** (0.0333)	0.27867** (0.0964)	.0754** (0.0354)	.2176** (0.1019)
yearsaway	0.0080*** (0.0022)	0.0233*** (0.0064)	0.0072** (0.0024)	0.0211** (0.0067)
schoolyears	0.0046 (0.0090)	0.0135 (0.0264)	0.0021 (0.0095)	0.0053 (0.0279)
relationship	0.0692 (0.0489)	0.1947 (0.1413)	0.0616 (0.0516)	0.1731 (0.1487)
origincountry	0.0003 (0.0004)	0.0008 (0.0012)	0.0003 (0.0004)	0.0009 (0.0012)
prevoccup	0.0376** (0.0112)	0.1135** (0.0333)	0.0324** (0.0121)	0.0983** (0.0356)
literateorigin	0.1484** (0.0661)	0.5351** (0.2229)	0.1559** (0.0716)	0.5528** (0.2410)
langcoursebi	0.1240*** (0.0338)	0.3607*** (0.0992)	0.1267*** (0.0356)	0.3735*** (0.1040)
langfamily	0.0235 (0.0161)	0.0606 (0.0460)	0.0219 (0.0168)	0.0553 (0.0481)
langfriends	-0.0343** (0.0132)	-0.0948** (0.0380)	-0.0300** (0.0139)	-0.0857** (0.0399)
langwork	-0.0353** (0.0103)	-0.1021** (0.0301)	-0.0385*** (0.0110)	-0.1130*** (0.0323)
visitedgermans			-0.1294** (0.0530)	-0.3984** (0.1596)
germansvisited			0.0719 (0.0598)	0.2236 (0.1804)
trustgermans			0.0003 (0.0163)	0.0014 (0.0471)
ethnicdiscrimination			-0.0590** (0.0272)	-0.1709** (0.0784)

Constant	-0.0194 (0.1600)	-1.5922** (0.4846)	0.3187 (0.2010)	-0.6116 (0.6010)
Fit measure (Adjusted R ² / pseudo R ²)	0.0998	0.0915	0.105	0.1008
n	819	819	745	745

* indicates significance at $p < 0.10$, ** indicates significance at $p < 0.05$, *** indicates significance at $p < 0.01$

The first iteration of the language proficiency model, including basic demographic and language variables, yields important results. As expected, demographic variables such as age, sex, years since migration, and previous occupation were found to be significant in both the linear probability model and the probit model. Younger migrants were found to be more likely to be proficient in German, as were males. Both the linear probability model and the probit model also found that the number of years since migration (*yearsaway*) is significant. Specifically, the likelihood of being proficient in German goes up by almost 1% for each year since the migrant left their birth country.

The model also found the migrant's previous occupation to be positive and significant. Though previous occupation is a categorical variable that does not lend itself to straightforward interpretation, generally it can be stated that migrants who had previously held occupations that require higher levels of skill are more likely to be proficient in German.

Whether or not the migrant was literate in the language of their country of origin (*literateorigin*) was also found to be significant. A migrant reporting that their written ability in their native language was "very good" improved their probability of being proficient in German by 14.84% in the linear probability model and 17.5% in the probit model, after taking the marginal effect, or "average effect" of each person in the data.

The language training variable, *languagecoursebi*, was found to be significant at the $p < .01$ level in both the linear probability model and the probit model. Attending a German

language course was found to have increased the likelihood of being proficient in German by 12.4% in the linear probability model and by 13.6% in the probit model. Additionally, the language the migrant speaks with their friends and the language the migrant speaks at work were both found to be significant at the $p < .05$ significance level. Though both *langfriends* and *langwork* are ordinal categorical variables that are not clearly interpretable, the more German the migrant spoke in both these settings, the more likely they were to be proficient in German.

The second iteration of the model includes ethnic identity variables as well as demographic and language variables. The first model's estimations were not significantly affected by the addition of the ethnic identity variables, but adding these variables did yield some interesting findings. First, the adjusted R^2 rose with the additional variables, which suggests that this model may explain more of the variation in language proficiency. Second, two ethnic identity variables were found to be significant at the $p < .05$ significance level. The first, *visitedgermans*, implies that migrants who had reported that they had visited people of German origin in their home in the last 12 months were about 13% more likely to be proficient in German according to the linear probability model, and about 15% more likely to be proficient in German according to the probit model.

The second variable, *ethnicdiscrimination*, is an ordinal categorical variable that describes how frequently the migrant reported being discriminated against because of their ethnic origin in the last 24 months, and it does not lend itself to straightforward interpretation. However, the results imply that the less the migrant reported being discriminated against, the more likely they were to be proficient in German.

Discussion

These findings are fairly consistent with the literature, and the significance of ethnic identity variables provides important insight as to possible determinants of migrant language proficiency. Most of the demographic variable estimates were similar to what was expected with the exception of relationship status (*relationship*), which was not significant. This could be explained by the expanded definition of the variable, which included all manners of “permanent relationships” as compared to previous studies, which have only accounted for traditional marriage.

Additionally, language training was found to have a significantly positive effect on migrant’s earnings. Though the magnitude of this effect-- the migrant being 13.6% more likely to be proficient in German after attending the course-- may seem rather small, it is important to remember the standard of what is considered “proficient.” For this analysis, a migrant was only considered “proficient” if they are “very good” or “good” at writing in Germany, which many would consider a higher standard than being “fair” at speaking German. This paper is surprisingly one of the few that has measured the effect of language training on language proficiency, and this finding suggests that language training is indeed an important tool for language learning.

Perhaps the most interesting finding of this analysis was the significance of two ethnic identity variables. This was the first study to include ethnic identity variables in models of language proficiency, with the exception of Dustmann and Fabbri, who included an ethnic concentration variable in their model. They found that higher ethnic concentration (migrants of similar backgrounds living in more isolated communities) results in lower language proficiency (Dustmann and Fabbri, 2003). This analysis found that migrants who had recently visited a

German's home or reported less discrimination based on their ethnicity were more likely to be proficient in German. This suggests that migrants who feel more assimilated with natives of their host country are more likely to be proficient in the language of their host country.

Additionally, adding the ethnic identity variables took some of the significance away from the number of years since the migrant left their birth country. This may suggest that in past models of language proficiency determinants, the "years since migration" variables may have incorporated many unobserved characteristics, including cultural assimilation factors and ethnic identity variables. Clearly, there is a lot of insight that could be gained from further study of these ethnic identity factors. A deeper understanding how cultural assimilation and ethnic discrimination factors play a role in language proficiency, and ultimately economic outcomes, would have powerful implications for immigration reform, education, and public policy.

Chapter 6

Language Proficiency and Earnings

Methodology

To measure the effects of language proficiency on earnings, I conducted various analysis using *lnearnings* as the dependent variable. The analysis was based on the Mincer earnings function, which explains the logarithm of earnings as sum of years of education and the square of "years of potential experience" (Thurow & Mincer, 1976). As discussed, this has served as the basis for many econometric analyses of language variables and earnings. These analyses are based on the common hypothesis that language proficiency is a form of human capital and should therefore increase productivity (Chiswick & Miller, 1995).

However, in this case, the use of a simple OLS estimation may be biased due to an association between errors and regressors. In the simple OLS model:

$$\ln earnings_{ij} = \alpha_0 + \alpha_1 langprof_{ij} + \alpha_2 X_{ij} + \varepsilon_{ij}$$

where X is a vector of independent variables such as years of schooling, work experience, and origin country and ε is the error term. The error term includes unobserved omitted variables such as work ethic and ability that would affect both a migrant's language proficiency and his earnings. This results in endogeneity in the model, and potentially biased parameter estimates. To correct this, a method is used that is meant to generate only exogenous variation attributed to *langprof*. In this case, a two-stage least squares process using instrumental variables is employed to address potential bias.

The challenge for this approach is to find an instrumental variable that is correlated with language proficiency, but does not directly affect earnings. In this analysis, I utilize two instrumental variables: the language the migrant speaks with his friends and whether or not he has visited a German citizen in his home in the past 12 months. These are suitable instruments because they affect language proficiency-- as those who speak more German with their friends are more likely to be proficient in German, as are those who visit Germans in their homes-- but do not directly affect the migrant's earnings. Additionally, out of possible instruments that made sense from a theoretical standpoint, *langfriends* and *visitedgermans* were the most highly correlated with language proficiency.

In the first stage, the model generates an estimated probability of language proficiency based on the stated instruments and set of exogenous variables X:

$$\widehat{langprof}_{ij} = \widehat{\alpha}_0 + \widehat{\alpha}_1 langfriends_{ij} + \widehat{\alpha}_2 visitedgermans_{ij} + \widehat{\alpha}_3 X_{ij} + \varepsilon_{ij}$$

In the second stage, the model estimates earnings as a function of the estimated language proficiency values and a set of exogenous variables X:

$$lnearnings_{ij} = \alpha_0 + \alpha_1 \widehat{langprof}_{ij} + \alpha_2 X_{ij} + \varepsilon_{ij}$$

Now, the extent to which earnings changes as a result of changes in language proficiency should be isolated from other omitted variables.

First, a simple OLS model was implemented without the use of instrumental variables to provide a basis for comparison. Then, two iterations of this two-stage least squares model were

executed. The first includes only basic Mincer and demographic variables, in addition to the estimated language proficiency. The second model includes additional employment variables, many of which were used in a previous model of the effects of language proficiency on earnings (Aldashev 2008).

As mentioned above, this analysis was limited to only full-time employed males. The purpose of this was to make the sample more homogenous and to limit the influence of issues like partial employment, gender-pay inequality, and child-rearing obligations. Accordingly, I anticipate the variables included in the model will better explain variation in earnings, rather than other unobservable characteristics.

Results

Table 2. Language Proficiency and Earnings

Lnearnings	(Base, no IV)	(1)	(2)
age	0.0440*** (0.0069)	0.0333** (0.0106)	0.0053 (0.0092)
yearsaway	0.0065** (0.0032)	-0.0039 (0.0073)	0.0024 (0.0053)
schoolyears	0.0341** (0.1131)	0.0415** (0.0146)	0.0049 (0.0124)
relationship	-0.1060 (0.0683)	-0.1659* (0.0983)	-0.0598 (0.1016)
origincountry	-0.0001 (0.0005)	-0.0009 (0.0007)	-0.0012** (0.0006)
yearslabormkt	-0.0228** (0.0098)	0.0017 (0.0189)	0.0193 (0.0125)
sqyearslabormkt	-0.0001** (0.0002)	-0.0008** (0.0003)	-0.0006** (0.0002)
lang prof	0.0644 (0.0494)	0.7715* (0.4324)	0.3300 (0.3863)
prevoccup			0.0208 (0.0163)
companysize			0.0393** (0.0137)
avghoursworked			0.0018 (0.0023)
trainingreq			0.1880*** (0.0357)
_cons	6.4447*** (0.2262)	6.3795*** (0.2833)	6.7043*** (0.3050)
n	370	367	289

* indicates significance at $p < 0.10$, ** indicates significance at $p < 0.05$, *** indicates significance at $p < 0.01$

As discussed above, this earnings regression was executed using a two-stage least squares method to correct for potential endogeneity associated with language proficiency. A simple OLS regression without the use of instrumental variables is included to serve as a baseline. The instruments used for Models 1 and 2 were the language the migrant speaks with their friends and whether or not they have visited a German citizen in their home in the past 12 months. Model 1 includes only basic Mincer variables and the language proficiency variable estimated in the first step of the two-stage least squares model.

In the first model, many of the Mincer demographic variables we would expect to be significant are indeed significant. Age is significant at the $p < .05$ level, and implies that a migrant earns 3% more for every year older he is. The number of years of school is also positive and significant at the $p < .05$ level; a migrant earns about 4% more for every year he was in school. The *relationship* variable is significant at $p < 0.10$, meaning that if the migrant reports being in a “permanent” relationship (including marriage, same-sex partnership, or a living partner) he earns 16.6% more than his single counterpart.

The square of the number of years in the labor market is significant and negative. This is logical because as a quadratic variable, it is expected that diminishing returns will cause labor market experience to contribute less and less to earnings over time. The Mincer findings illustrate a similar pattern.

The language proficiency variable, estimated during the first step of our two-stage least squares regression, is significant at the $p < .10$ level. According to the results, German language proficiency increases earnings by 77%. However, there is significant variation in this finding with a standard error of 43%.

The second iteration of the two-stage least squares method incorporates basic Mincer variables, estimated language proficiency, and additional employment variables, many of which were used in Aldashev 2008.

When these variables are added to the model, age, school years, and relationship are no longer significant. The migrant's country of origin is significant at the $p < .05$ level, but *origincountry* is a categorical variable that serves as a control but is not accommodating to a clear-cut interpretation. The square of the number of years in the labor market is negative and significant in this model as well, but with a .06% implied earnings increase with a standard deviation of .02%, this finding may not be very meaningful either.

Furthermore, after the addition of these employment variables, language proficiency is no longer significant. The parameter estimate for language proficiency is 33% with a standard deviation of 39%.

The employment variables that are significant in the second iteration of the earnings model are company size and training required. Company size, significant at $p < .05$, shows that for each level increase in company size, a migrant's earnings increased by 4% with a standard error of 1%.

Training required is significant at the $p < .01$ level, and the model suggests that for each increase in the level of training required, migrants' wages increase by 19% with a standard error of 4%.

Discussion

The baseline simple OLS model and the first iteration of the two-stage least squares method are consistent with previous research in that the Mincer variables behave in the ways we would expect. Additionally, language proficiency has a significantly positive effect on earnings, which makes sense according to human capital theory. The second iteration of the two-stage least squares method is unusual in that no significance was found for any of the Mincer variables except the square of number of years in the labor market. Additionally, after adding the additional employment variables, language proficiency was no longer found to be significant.

Failing to find significance for language proficiency in earnings regressions is common in the literature, and many papers have offered explanations for this. Hayfron theorizes that immigrants need language proficiency to enter the labor market but that once they are employed, their wages are not necessarily determined by their language skills. A version of this theory could apply here, where German language proficiency may be a prerequisite for jobs requiring a certain level of skill, but that once an immigrant is qualified for a position at that skill level, their language skills may not greatly affect earnings. Aldashev et al. found that the impact of language proficiency affected employment probabilities and occupational choice, but once those factors were controlled for, the impact on earnings was not significant.

It is also important to consider these findings in the context of Dustmann and van Soest's work evaluating bias in estimating language effects on earnings. As stated previously, they found that negative bias tends to be much greater than positive bias, and that language may be much more important as a determination of earnings than OLS regressions suggest. They concluded that parameter estimates of language effects tend to be under-estimated, even when employing instrumental variables to correct for potential endogeneity (Dustmann & van Soest, 2002).

Keeping this in mind, it is likely that the significance of language proficiency on earnings is under-estimated in both iterations of the two-stage least squares regression.

Second, there could be simultaneous causality between language proficiency, training required, and firm size. Those who speak better German are more likely to seek higher training in Germany, just as those with more training are more likely to become better at German.

Likewise, those who work at large companies in Germany have a greater need to learn German, just as those who know German well are more likely to be hired at a large company.

Another potential explanation could be the lack of a variable to control for the type of town where the migrant is employed. Whether the migrant is working in an urban, suburban, or rural area would have a significant impact on their earnings, in addition to being highly correlated with both the size of their company and the level of training required.

Overall, this analysis shows that bias and endogeneity when estimating the economic returns to language proficiency continues to be ongoing issues. However, the use of instrumental variables yields interesting results, especially when the instruments are strong from both a theoretical and empirical standpoint. Because of the challenge of finding instruments that impact language proficiency but do not directly impact earnings, their use is not particularly widespread in the literature. With the exception of Hayfron, who uses the ethnicity of the migrant's wife and the migrant's mother tongue as instruments, and Dustmann and Van Soest, who use the migrant's parents' education, the use of this technique has been limited.

No previous studies have used cultural assimilation variables such as the language the migrant speaks with their friends and whether they have visited a German citizen in their home in the past 12 months. Based on these results, cultural assimilation variables could be valuable in correcting bias when finding the economic returns to language proficiency.

Chapter 7

Conclusion

The discussion surrounding policy, education, and the economics of immigration are of vital importance in the current political climate. Continued study of the economic impact of immigration on natives, as well as the issues that immigrants face, is necessary. This analysis suggests that language training is an important determinant of language proficiency, and that cultural assimilation variables and ethnic identity factors play an integral role in language learning as well. This paper also addresses some of the bias and endogeneity problems faced when modeling the effect of language training on earnings. It does this by employing instrumental variables that impact a migrant's language proficiency but do not directly affect migrant's earnings. These variables, again measures of cultural assimilation, correct for some of the measurement error and suggest that language proficiency may have a significant effect on earnings.

Overall, this study suggests that there are significant implications when migrants feel like active members of their community within safe and accepting surroundings of their host country. Much research is still necessary to further quantify the economic impact of cultural assimilation and ethnic identity. However, this paper suggests that the more prevalent concern of "how are migrants affecting my life?" ignores the equally important question of "how am I affecting the lives of the migrants around me?" By recognizing the issues facing migrants and further exploring methods of efficiently assimilating them to be productive members of society, the debate surrounding immigration can become much more balanced and informed.

Appendix A

Descriptive Statistics

Variable name	Variable description	Possible Values	Descriptive Statistics(n=819)	Descriptive Statistics(n=411)
sex	sex	1 if male, 2 if female	1.4460 (0.4973)	1 (0)
origincountry	coded values for the migrant's country of birth, for example [4] Greece	Over 30 different countries and their country codes	47.0746 (42.3041)	45.2586 (42.0689)
ethniodiscrimination	if the migrant thought they had been discriminated against because of their ethnic origin in the past 24	1 if often, 2 if seldom, 3 if never	2.5937 (0.6312)	2.6609 (0.5821)
langfamily	self-reported language used with family members	1 if mainly German, 2 if mainly language country of origin, 3 if mainly different language, 4 if equal proportion different language	2.4773 (1.0548)	2.4610 (1.0276)
langfriends	self-reported language used with friends	1 if mainly German, 2 if mainly language country of origin, 3 if mainly different language, 4 if equal proportion different language	2.4490 (1.2341)	2.4380 (1.2229)
langwork	self-reported language used at workplace	1 if mainly German, 2 if mainly language country of origin, 3 if mainly different language, 4 if equal proportion different language	2.0902 (1.5243)	1.6463 (1.0579)
prevoccup	the migrants last occupational status before moving to Germany	1 if blue-collar worker, no managerial function; 2 if blue collar worker, with managerial function, 3 if self-employed or working for a self-employed relative; 4 if white-collar worker, no managerial function; 5 if white-collar worker with managerial; 6 if civil servant, no managerial function; 7 if civil servant, no managerial function	3.2405 (1.5340)	3.0157 (1.5756)
avghrswork	self-reported average hours worked per week, including overtime	0 to 900	29.3754 (21.9455)	44.1190 (10.9690)
schoolyears	number of years spent at school	1 to 17	10.9112 (2.0464)	10.9947 (2.0013)
relationship	if the migrant is in a "permanent" relationship- including marriage, same-sex partnership, living partner, et.	1 if yes 2 if no	1.1728 (0.3782)	1.1144 (0.3186)
visitedgermans	whether the migrant has visited people of German origin in the past year	1 if yes, 2 if no	1.2591 (0.4384)	1.2506 (0.4339)
germansvisited	whether people of German origin have visited the migrant in the past year	1 if yes, 2 if no	1.2055 (0.4043)	1.1764 (0.3817)
trainingreq	training required for current job	1 if no training required, 2 if professional training completed, 3 if technical college, 4 if university education completed	1.8378 (1.1214)	1.9317 (1.1555)
companysize	estimated number of workers at company of current job, not just the local unit of the company but the whole company	1 if less than 5, 2 if from 5 up to 10 people, 3 if from 11 up to but less than 20 people, 4 if from 20 up to but less than 100 people, 5 if from 100 up to but less than 200 people, 6 if from 200 but less than 2,000 people, 7 if 2,000 people or more	4.6871 (2.0171)	4.9702 (1.9055)
grossincome	gross income, income before deduction of taxes and social security; estimated monthly income before taxes if self-employed; not including extra income such as vacation or back-pay; including overtime pay	0 to 10500	1510.529 (1560.701)	2603.917 (1476.402)
age		18 to 63	35.6977 (9.2303)	36.0535 (8.5390)
langprof	1 if reported to write German good or very good, 0 if fair, poor, or not at all	0 if no, 1 if yes	0.4011 (0.49033)	0.3366 (0.4731)
yearsaway		0 to 53	7.7261 (7.8684)	8.3512 (8.3419)
langcoursebi	marked 1 if had taken language course at all, 0 if not	0 if no, 1 if yes	0.6025 (0.4896)	0.5366 (0.4993)
lnearnings			7.3135 (0.8198)	7.7494 (0.4642)
yearslabormkt		0 to 47	16.1368 (9.8243)	16.4767 (9.5638)
sqyearslabormkt		0 to 2209	356.8205 (396.2609)	362.7224 (388.7578)
literateorigin		1 if very good, 0 if good, fairly, poorly, or none at all	0.9042 (0.2944)	0.9098 (0.2869)

BIBLIOGRAPHY

(www.dw.com), D. (2017). *Germany sees record number of people with immigrant background* / News / DW / 01.08.2017. DW.COM. Retrieved 9 November 2017, from <http://www.dw.com/en/germany-sees-record-number-of-people-with-immigrant-background/a-39917868>

Aisch, G., Bloch, M., Lai, K., & Morenne, B. (2017). *How France Voted*. *Nytimes.com*. Retrieved 7 November 2017, from <https://www.nytimes.com/interactive/2017/05/07/world/europe/france-election-results-maps.html>

Aldashev, A., Gernandt, J., & Thomsen, S. (2008). Language Usage, Participation, Employment and Earnings - Evidence for Foreigners in West Germany with Multiple Sources of Selection. *SSRN Electronic Journal*. <http://dx.doi.org/10.2139/ssrn.1307646>

Altonji, J., & Card, D. (1991). The Effects of Immigration on the Labor Market Outcomes of Less-skilled Natives. *National Bureau Of Economic Research*.

Babones, S. (2018). *Australia's Economic Miracle Depends On Immigration. So Why Curb It?*. *Forbes.com*. Retrieved 12 February 2018, from <https://www.forbes.com/sites/salvatorebabones/2017/10/26/australias-economic-miracle-depends-on-immigration-so-why-curb-it/#2a7e61d06141>

BAMF - Bundesamt für Migration und Flüchtlinge - Important information. (2017). *Bamf.de*. Retrieved 9 November 2017, from <http://www.bamf.de/EN/Willkommen/Aufenthalt/WichtigeInformationen/wichtigeinformationen-node.html>

Borjas, G., Freeman, R., Katz, L., DiNardo, J., & Abowd, J. (1997). How Much Do Immigration and Trade Affect Labor Market Outcomes?. *Brookings Papers On Economic Activity*, 1997(1), 1. <http://dx.doi.org/10.2307/2534701>

Brock, J. (2017). *In South Africa, immigration feeds corrupt officials and race hate*. *Reuters*. Retrieved 7 November 2017, from <http://www.reuters.com/investigates/special-report/safrica-migrants-corruption/>

Carlner, G. (1981). Wage Differences by Language Group and the Market for Language Skills in Canada. *The Journal Of Human Resources*, 16(3), 384. <http://dx.doi.org/10.2307/145627>

Chiswick, B. (1991). Speaking, Reading, and Earnings among Low-Skilled Immigrants. *Journal Of Labor Economics*, 9(2), 149-170. <http://dx.doi.org/10.1086/298263>

Chiswick, B., & Miller, P. (1995). The Endogeneity between Language and Earnings: International Analyses. *Journal Of Labor Economics*, 13(2), 246-288.

Chiswick, B., & Miller, P. (1999). Language skills and earnings among legalized aliens. *Journal Of Population Economics*, 12(1), 63-89. <http://dx.doi.org/10.1007/s001480050091>

Chiswick, B., & Miller, P. (2002). Immigrant earnings: Language skills, linguistic concentrations and the business cycle. *Journal Of Population Economics*, 15(1), 31-57. <http://dx.doi.org/10.1007/pl00003838>

Chiswick, B., & Miller, P. (2005). Linguistic Distance: A Quantitative Measure of the Distance Between English and Other Languages. *Journal Of Multilingual And Multicultural Development*, 26(1), 1-11. <http://dx.doi.org/10.1080/14790710508668395>

Chiswick, B., & Miller, P. (2009). Occupational language requirements and the value of English in the US labor market. *Journal Of Population Economics*, 23(1), 353-372. <http://dx.doi.org/10.1007/s00148-008-0230-7>

Constant, A., & Zimmerman, K. (2009). *Work and Money: Payoffs by Ethnic Identity and Gender*. Berlin: German Institute for Economic Research. Retrieved from <http://www.diw.de/documents/publikationen/73/100081/dp908.pdf>

Dustmann, C. (1994). Speaking Fluency, Writing Fluency and Earnings of Migrants. *Journal Of Population Economics*, 7(2), 133-156.

Dustmann, C., & Fabbri, F. (2003). Language Proficiency and Labour Market Performance of Immigrants in the UK. *The Economic Journal*, Vol. 113(No. 489), pp. 695-717.

Dustmann, C., & Frattini, T. (2014). The Fiscal Effects of Immigration to the UK. *The Economic Journal*, 124(580), F593-F643. <http://dx.doi.org/10.1111/ecoj.12181>

Dustmann, C., & van Soest, A. (2002). LANGUAGE AND THE EARNINGS OF IMMIGRANTS. *Industrial And Labor Relations Review*, 55(3), 473-392.

Eddy, M. (2017). *Austria Shifts Right as Refashioned Conservatives Win*. *Nytimes.com*. Retrieved 7 November 2017, from <https://www.nytimes.com/2017/10/15/world/europe/austria-election-sebastian-kurz.html>

Esposito, A., & Iturrieta, F. (2017). *Chile's presidential hopefuls bet on anti-immigrant sentiment*. *U.S.*. Retrieved 7 November 2017, from <https://www.reuters.com/article/us-chile-immigration/chiles-presidential-hopefuls-bet-on-anti-immigrant-sentiment-idUSKBN1591CF>

Felter, C., & Renwick, D. (2017). *The U.S. Immigration Debate*. *Council on Foreign Relations*. Retrieved 7 November 2017, from <https://www.cfr.org/backgrounder/us-immigration-debate-0>

Fitzenberger, B., & Speckesser, S. (2006). Employment effects of the provision of specific professional skills and techniques in Germany. *Empirical Economics*, 32(2-3), 529-573. <http://dx.doi.org/10.1007/s00181-006-0088-z>

French election 2017: The final candidates. (2017). *BBC News*. Retrieved 7 November 2017, from <http://www.bbc.com/news/world-europe-38220690>

Gazzola, M., Grin, F., & Wickström, B. (2015). A Concise Bibliography of Language Economics. *Center For Economic Studies & Ifo Institute, Working Paper No. 5530*.

Germany: Immigration in Transition. (2017). *migrationpolicy.org*. Retrieved 9 November 2017, from <https://www.migrationpolicy.org/article/germany-immigration-transition>

Grin, F. (2008). The Economics of Language Education. *Encyclopedia Of Language And Education*, 83-92. <http://dx.doi.org/10.1007/978-0-387-30424-3>

Hayfron, J. (2001). Language training, language proficiency and earnings of immigrants in Norway. *Applied Economics*, 33(15), 1971-1979. <http://dx.doi.org/10.1080/00036840010018630>

Immigration is now the top issue for voters in the EU referendum. (2016). *Ipsos MORI*. Retrieved 7 November 2017, from <https://www.ipsos.com/ipsos-mori/en-uk/immigration-now-top-issue-voters-eu-referendum>

Migration History in Germany | domid.org. (2017). *Domid.org*. Retrieved 9 November 2017, from <http://www.domid.org/en/migration-history-germany>

Mintz, S. (2018). *Immigration Policy in World War II | Gilder Lehrman Institute of American History.* *Gilderlehrman.org*. Retrieved 12 February 2018, from <https://www.gilderlehrman.org/content/immigration-policy-world-war-ii>

Mohor, D., & Haynie, D. (2017). *Borderline Confusion Across South America.* *US News World Report*. Retrieved 12 February 2018, from <https://www.usnews.com/news/best-countries/articles/2017-07-10/south-americas-progressive-immigration-laws-begin-to-fray>

Peri, G., & Sparber, C. (2011). Assessing inherent model bias: An application to native displacement in response to immigration. *Journal Of Urban Economics*, 69(1), 82-91.
<http://dx.doi.org/10.1016/j.jue.2010.08.005>

Peri, G., & Sparber, C. (2011). Assessing inherent model bias: An application to native displacement in response to immigration. *Journal Of Urban Economics*, 69(1), 82-91.
<http://dx.doi.org/10.1016/j.jue.2010.08.005>

PERI, G., & SPARBER, C. (2011). Highly Educated Immigrants and Native Occupational Choice. *Industrial Relations: A Journal Of Economy And Society*, 50(3), 385-411.
<http://dx.doi.org/10.1111/j.1468-232x.2011.00643.x>

Thurow, L., & Mincer, J. (1976). Schooling, Experience, and Earnings. *Industrial And Labor Relations Review*, 29(3), 454. <http://dx.doi.org/10.2307/2521600>

Wolff, S. (2017). *Immigration, a consensual issue in the French presidential campaign?*. *EUROPP*. Retrieved 7 November 2017, from
<http://blogs.lse.ac.uk/europpblog/2017/04/15/immigration-a-consensual-issue-in-the-french-presidential-campaign/>

ACADEMIC VITA

Academic Vita of Sarah Vlazny

sarahvlazny7@gmail.com

Education

Major(s) and Minor(s): Economics and Public Relations with minors in Spanish and International and Global Studies

Honors: Honors in Economics

Thesis Title: Language Training, Language Proficiency, and the Effect of Language Proficiency on Earnings for Migrants in Germany

Thesis Supervisor: Amanda Mandzik

Work Experience

Date: October 2016- present

Title: Student Intern

Description: Designs graphics and edits digital content for online teaching modules, interviews industry executives and communications professors, writes blog posts, case studies, and other communications materials.

Institution/Company: The Arthur W. Page Center for Integrity in Communication

Supervisor Name: Denise Bortree

Date: June 2017- August 2017

Title: Summer Consultant

Description: Processed and analyzed large datasets using Stata and Excel, created visual exhibits and wrote summaries to communicate findings. Collaborated with team members to complete a case study, presented and defended case study results in front of over 50 full-time employees.

Institution/Company: Bates White Economic Consulting

Supervisor Name: Connor Kincheloe

Date: January 2017- August 2016

Title: PR Assistant

Description: Pitched stories and press releases to over 100 media outlets on behalf of clients, conducted market research for client deliverables.

Institution/Company: Bates White Economic Consulting

Supervisor Name: Connor Kincheloe

Grants Received: Department of Economics Undergraduate Scholarship, two Schreyer Travel Grants, Schreyer Internship Grant

Awards: Evan Pugh Scholar Senior Award, Provost Award and scholarship, Paterno Fellow Scholarship

Community Service Involvement: Tutor for State College Area School District, Volunteer Experience at Connell Academy in Costa Rica

International Education: University of Virginia Hispanic Studies Program in Valencia, Spain; Internship in Melbourne, Australia

Language Proficiency: Spanish