Language Usage, Participation, Employment and Earnings

Evidence for Foreigners in West Germany with Multiple Sources of Selection

Alisher Aldashev, Johannes Gernandt, and Stephan L. Thomsen

ZEW

Zentrum für Europäische
Wirtschaftsforschung GmbH

Centre for European
Economic Research

Discussion Paper No. 08-090

Language Usage, Participation, Employment and Earnings

Evidence for Foreigners in West Germany with Multiple Sources of Selection

Alisher Aldashev, Johannes Gernandt,
and Stephan L. Thomsen

Non-technical summary

A common finding in the international literature on immigrants' language proficiency on labor market outcomes is a positive effect on earnings. An open question so far is, whether this is due to a direct effect of language proficiency, or whether there is an indirect effect only through the choice of better paid jobs. Put it in other words, are immigrants with better language proficiency paid better in a firm than immigrants with poorer language command or do they earn different wages because they work in different firms and do different jobs?

In this paper, we answer this question by analyzing the effect of language proficiency on earnings for foreigners in Germany using data from GSOEP. The empirical analysis, motivated by a theoretical model, considers the effect of language proficiency on four possible stages of self-selection in the labor market. The first two stages comprise the decision to participate in the labor market and the employment decision. Persons who are employed further select themselves into economic sectors and occupation, i.e. the second two stages of self-selection. Concerning economic sectors we differentiate between basic and high-tech industry/manufacturing and other sectors; concerning occupation between white- and blue-collar occupations.

The results show that foreigners with better language command have a significantly higher participation rate and employment chances. Moreover, better language proficiency leads to higher probability of working in white-collar occupations. Once, all these selection mechanisms are controlled for, we do not find any direct effect of language proficiency on earnings. Instead, without controlling for selection one observes a significant positive effect of language proficiency on earnings.

Das Wichtigste in Kürze

Die internationale Literatur zum Einfluss der Sprachfähigkeit von Immigranten auf die Arbeitsmarktintegration zeigt einen positiven Effekt auf das Einkommen, d.h. eine bessere Sprachfähigkeit führt im Durchschnitt zu einem höheren Einkommen. Eine offene Frage in diesem Zusammenhang ist, ob die Sprachfähigkeit direkt auf den Verdienst wirkt oder ob lediglich die Wahl der Beschäftigung beeinflusst wird. Anders ausgedrückt: Verdienen Immigranten mit besseren Sprachfähigkeiten bei ansonsten vergleichbaren Eigenschaften mehr Geld im gleichen Beruf, oder selektieren sie sich in besser bezahlte Beschäftigungsverhältnisse?

Diese Studie beantwortet diese Frage anhand der Untersuchung des Einflusses der Sprachfähigkeit von Ausländern auf ihren Arbeitsverdienst in Deutschland mit Daten des Sozioökonomischen Panels. In der empirischen Analyse werden hierzu vier Stufen der Selbstselektion berücksichtigt: Erstens, ob die Person überhaupt eine Beschäftigung sucht. Sollte dies der Fall sein, gibt es zweitens die Möglichkeit, dass die Person beschäftigt oder arbeitslos ist. Sollte die Person einer Beschäftigung nachgehen, so können als dritte und vierte Stufe die unterschiedlichen Wirtschaftsbereiche (z.B. Industrie) und die verschiedenen Arbeitsverhältnisse (z.B. Arbeiter und Angestellte) identifiziert werden.

Die Ergebnisse zeigen, dass Personen mit besseren Sprachfähigkeiten mit einer signifikant höheren Wahrscheinlichkeit am Arbeitsmarkt partizipieren und auch eher eine Beschäftigung finden. Des Weiteren erhöhen bessere Sprachfähigkeiten die Wahrscheinlichkeit, in einem Angestelltenverhältnis beschäftigt zu werden. Wenn man alle Selektionsstufen berücksichtigt, so zeigt die Sprachfähigkeit keinen statistisch signifikanten Einfluss auf das Einkommen, ein direkter Effekt ist also nicht nachweisbar. Dies ist ein wichtiges, neues Ergebnis, da der geschätzte Effekt ohne die Selektionskontrolle – wie in der Literatur beschrieben – hingegen signifikant positiv ist.

Language Usage, Participation, Employment and Earnings[*]

Evidence for Foreigners in West Germany with Multiple Sources of Selection

Alisher Aldashev[†], Johannes Gernandt[‡] and Stephan L. Thomsen[§]

[†,‡] *ZEW, Mannheim*
[§] *University of Magdeburg & ZEW, Mannheim*

This version: October 27, 2008

Abstract

Language proficiency may not only affect the earnings of the individual, but the probability to participate in the labor market or becoming employed as well. It may also affect selection of people into economic sector and occupation. In this paper, the effects of language proficiency on earnings are analyzed for foreigners in Germany with joint consideration of up to four types of selection. The results show that language proficiency significantly increases participation and employment probability and affects occupational choice. When selection into economic sector and occupation is regarded, we do not find an impact of language ability on earnings thereby implying an indirect effect.

Keywords: Foreigners, Participation, Employment, Language Ability, Multiple Selection
JEL Classification: J61, J24, J15, C35

[†]Alisher Aldashev is Research Fellow at the Centre for European Economic Research (ZEW), Mannheim, e-mail: aldashev@zew.de.

[‡]Johannes Gernandt is Research Fellow at the Centre for European Economic Research (ZEW), Mannheim, e-mail: gernandt@zew.de.

[§]Stephan L. Thomsen (corresponding author) is Assistant Professor of Labor Economics at Otto-von-Guericke-University Magdeburg and Research Professor at Centre for European Economic Research (ZEW) Mannheim. Address: PO Box 4120, D-39016 Magdeburg, e-mail: stephan.thomsen@ovgu.de, phone: +49 391 6718736, fax: +49 391 6711218.

1 Introduction

A core topic of the migration literature are earnings' differentials between immigrants and natives that have been explored in numerous studies since the seminal paper of Chiswick (1978). A well-established finding of many studies is that migrants' earnings lack behind those of the equally-qualified and experienced native population. In addition, catching up at a later stage depending on the duration of residence is documented in the literature leading to the so-called assimilation hypothesis. One explanation for the gap in earnings between foreigners and natives is language deficiency of the immigrants that may hamper productivity. From a policy perspective, language effect is interesting as language learning could be stimulated through language courses or similar activities. Consequently, the impact of immigrants' language ability on earnings has been studied by many authors. Whereas some early studies, see, e.g., Carliner (1981) or Reimers (1983; 1985), could not establish significant positive effects of language proficiency on earnings, the majority of later studies revised that finding presenting clear positive effects, see, e.g., McManus, Gould, and Welch (1983), Tainer (1988), Dustmann (1994), or Chiswick and Miller (1995).[1] Given these studies are not directly comparable due to differences with respect to the analyzed questions, the formulation of the estimation equations and sample selection criteria, makes this common positive effect of language on earnings even more remarkable.

Despite the large number of studies on language effects on earnings, the context of employment or job type have been scarcely considered. McManus, Gould, and Welch (1983) argue that if economy consisted of one-person producing units and impersonal products, communication would have no effect on earnings. If, alternatively, products were personal, so that the value of the service may depend on a producer's ability to communicate and convey the product to a consumer, communication ability would affect earnings. More practically, Kossoudji (1988) argues that jobs are extremely heterogenous in their use of language. Although communication is necessary in most jobs, its contribution to productivity clearly varies by occupation. In a similar vein, Tainer (1988) argues that language proficiency is an important communication factor, and communication is a vital aspect of any job. For that reason, language deficiency could, on the one hand, prevent someone from getting a job. On the other hand, even if the person finds a job, it could be below the skill level depending on the degree of the deficiency and the communication skills of the co-workers. Nevertheless, characteristics of occupation are not regarded in the majority of studies and earnings are used to reflect the communication inability.

An early attempt to include occupation is given by Kossoudji (1988) who explicitly analyzes the effect of language on occupation-specific earnings and occupational mobility in the US. Language ability is measured in terms of English language deficiency. Estimating a mixed simultaneous

---

[1] Further studies establishing positive effects, amongst others, are Grenier (1984), Kossoudji (1988), Rivera-Batiz (1990), Chiswick (1991), Chiswick and Miller (1999), Dustmann (1999), Dustmann and van Soest (2001), Dustmann and van Soest (2002), Shields and Price (2002), Berman, Lang, and Siniver (2003), and Dustmann and Fabbri (2003). Chiswick and Miller (1995) provide a comparison for adult male immigrants (aged 25-64 years) in Australia, the United States, Canada, and Israel. For all countries they find significant positive effects of host-country-language fluency, with the largest effect for the United States.

model the results indicate that English language deficiency has no effect on earnings in either of the two highest productivity occupations (managers and professionals). The explanation provided by the author is that immigrant professionals may possess unique skills highly demanded in the labor market that employers are willing to ignore language deficiencies. With regards to other occupations English language deficiency alters access to jobs. Very recently, Chiswick and Miller (2007a) pick up the topic for the US again. They argue that occupation provides information about the channels through which earnings gains are achieved. Similar to education, language can be viewed as having both direct and indirect effects on earnings, where indirect effects operate via occupational attainment. In their empirical analysis, they include occupations as further regressors in the earnings equation, implicitly assuming that occupation is independent of language usage. Their results show that inclusion of occupation reduces the earnings gap, and the effect is strongest for the low-skilled. In that sense, the authors conclude that some of the earnings disadvantage of the immigrants could be attributed to their occupational choice.

Further reasons may complicate the evaluation of the impact of language proficiency on earnings. On the one hand, the endogeneity of language proficiency of immigrants may induce further issues of self-selection beyond occupational choice. Time of residence and age at migration may affect language proficiency (e.g., Bleakley and Chin, 2004) in a way that persons who stayed longer in the host country accumulate better language skills and, therefore, earn higher wages.[2] Moreover, language proficiency may depend on country of origin (e.g., Espenshade and Fu, 1997) or linguistic distance (e.g., Chiswick and Miller, 2007b; Beenstock, Chiswick, and Repetto, 2001). This indicates that we would expect language proficiency not only have an impact on earnings, but also on the selection processes in the labor market. Ignoring these impacts may lead to biased estimates. On the other hand, interpretation of the effects of language proficiency on earnings could be hampered due to measurement or misclassification errors in the language proficiency variable. Dustmann and van Soest (2001; 2002) analyze the bias due to time-varying and time-persistent misclassification errors in self-reported language indicators in German data.[3] The results show that potential measurement errors lead to downward biased estimates of coefficients of language proficiency on earnings in that sense that individuals over-report their language abilities. Controlling for this leads to larger effects of language proficiency on earnings and also to a higher standard deviation. Charette and Meng (1994) examine the nature and potential measurement error resulting from self-assessed measure of literacy using Canadian data including subjective and objective measures of literacy. Their results show that the measurement error in the self-assessed literacy variable does not have a serious effect on the estimates of the wage determination equation.

The study at hand contributes to the existing literature in a number of aspects. Based on a theoretical model that provides a link between language proficiency and the various selection processes in the labor market, individuals with better language skills are more likely to participate and to be employed even in the absence of a language premium in wages. In the empirical analysis, we

---

[2] It should be noted that Beenstock, Chiswick, and Paltiel (2005) reject this so-called Immigration Assimilation Hypothesis.

[3] A similar analysis is provided by Dustmann and Fabbri (2003) using data for the UK.

explicitly regard these different selection processes. Namely, we take account of self-selection into participation as well as of self-selection into employment. In addition, more proficient individuals could be expected to end up working in higher-paying firms or in higher-paid occupations. Therefore, we extend the model to estimate the effect of language proficiency on earnings regarding additional patterns of self-selection into economic sector and occupation, which is a major difference to the previous literature. In contrast to Chiswick and Miller (2007a) who regard occupation as a further regressor, we take account of the potential endogeneity due to language proficiency by including a further selection equation. Moreover, we extend the set-up of the analyses of Kossoudji (1988) and Chiswick and Miller (2007a) by considering selection processes prior to occupation as well. In addition, besides occupation we also take account of economic sector choice that has not been done before. To consider skill heterogeneity and capture the potential relationship between language proficiency and skill-level, we estimate skill-specific effects. Finally, by addressing these issues we add to the scarce literature on effects of occupation (so far available only for the US) with first results for Germany.[4]

The empirical analysis is based on data from GSOEP. To mitigate problems of measurement error in the language proficiency variable, we use information on language usage in the household as a proxy for individual language command, which supposedly suffers less from measurement error than the self-reported language proficiency information. To test the sensitivity of the results we also present the estimates using the self-assessed language proficiency variable. In that sensitivity analysis, we use language usage in the household and time of residence as instruments to capture measurement error. The estimates are robust irrespective to which variable is used to approximate language proficiency. The results show that language proficiency significantly increases participation and employment probability of foreigners in Germany. Moreover, earnings are clearly higher for persons speaking mainly or at least partly German in the household compared to people using mainly their native language. When additional selection into economic sector and type of occupation is considered, foreigners with better language proficiency appear more likely to be white-collar workers. With regards to earnings, once selection into occupation and economic sector is taken into account, no significant effect of language usage on earnings for blue-collar workers can be observed. For white-collar workers no conclusive evidence is found.

The paper is organized as follows: Section 2 presents the estimation strategy starting with the outline of a theoretical model with flat wages within each firm and the econometric methodology afterwards. Details on the data and some selected descriptives are given in section 3. The empirical estimates of the effects of language usage on the selection processes and earnings are discussed in section 4. In addition, details of the sensitivity analysis using different measures of language

---

[4] There are number of studies exploring the effects of language usage on earnings in Germany. Dustmann (1994), e.g., analyzes the determinants of German language ability and the effects on earnings testing the assimilation hypothesis. However, he addresses potential self-selection bias for women only and neglects potential occurrence for men. In addition, Dustmann (1999) studies the incentives to invest in language capital conditional on the expected duration of stay, i.e. comparing temporary and permanent immigration. Finally, Dustmann and van Soest (2001; 2002) focus on the potential bias of estimates due to measurement or misclassification errors. All studies establish positive effects of language on earnings but self-selection with regards to the stages discussed here is not regarded.

proficiency are provided. The final section 5 concludes.

2 Estimation strategy

2.1 Theoretical background

Assuming language proficiency to be related to productivity and through this affecting earnings, participation and employment, a useful theoretical framework for analyzing the central question of the paper is given by a variant of the Burdett and Mortensen (1998) model (see also Manning, 2003, and Mortensen, 2003). This model provides argumentation for controlling for self-selection even if an employer pays the same wage for workers with different language abilities.

Consider an economy consisting of three types of individuals with productivities p_0, p_1, and p_2, such that $p_0 < p_1 < p_2$. Suppose that there is a guaranteed minimal income b (for example welfare assistance or minimum wage). The lowest possible wage a firm can set is thus b. Suppose that $p_0 < b$. This implies that no firm employs individuals with productivity p_0 and they do not participate. Assume the inflow of job offers to unemployed workers happens in continuous time according to a stationary Poisson process so that each worker receives only one offer at maximum during an infinitesimal short time interval. Employed workers may search on the job for higher paid vacancies. The arrival rates of job offers to employed and unemployed workers are assumed to be equal. Since the arrival rates are the same for out-of-job and on-the-job search, it is optimal for a worker to accept the first offer he receives and to continue searching on the job for a better offer (if returns to search are higher than the search costs). As a result in equilibrium there will still be a fraction of highly productive workers employed at low-wage firms.

For simplicity, suppose further that there is a flat wage policy in any firm, so that a firm that hires both p_1 and p_2 workers pays equal wages to them. A firm offering a wage lower than p_1 can hire both p_1 and p_2 individuals, but a firm with a wage above p_1 can hire only p_2 workers. Consider an arbitrary firm 1, which offers a wage equal to b and has a profit of $\pi_1 = (p_1 - b)L_1(p_1) + (p_2 - b)L_1(p_2)$, where $L_1(p_1)$ is the labor supply of p_1 workers to firm 1 and $L_1(p_2)$ is the labor supply of p_2 workers to firm 1. A firm 2, offering a wage w greater than p_1, has a profit of $\pi_2 = (p_2 - w)L_2(p_2)$, where $L_2(p_2)$ is the labor supply of p_2 workers to firm 2. Then, as in a Burdett and Mortensen (1998) model, there exists an equilibrium such that both firms are equally profitable. Firm 2 pays a higher wage, but at the same time has a larger workforce of p_2 workers as it would "steal away" some of the workers from firm 1 attracted by a higher wage. Ultimately, more p_2 workers would be concentrated in high-wage firms.

The model implies that even in the presence of a flat wage policy within a firm, the observed wages in the sample between p_1 and p_2 workers would be different due to differences in employment probabilities. When estimating the effect of p on wages one has to keep in mind that p_1 workers are more likely to be unemployed than p_2 workers. Moreover, participation rates differ with p

as p_0 individuals do not participate. This implies that the estimated effect of p on wages based on the sample of employed individuals is biased as the sample of employed wage-earners is self-selected and a distribution of p in a sample of employed individuals is not a correct estimate of the distribution of p in the population.

2.2 Econometric model and selection issues

Suppose that worker's productivity cannot be observed by a firm, but a firm knows that language proficiency is related to productivity as

$$p = \psi(H) + \upsilon, \tag{1}$$

where H is language proficiency, ψ is some arbitrary function and υ includes other factors affecting productivity including a random component. Language proficiency then signals the potential productivity.

The theoretical model is constructed in such a way to allow a flat-wage policy, i.e. a firm pays the same wage to workers with different productivities. In reality, this would mean that firms pay the same wage independent of language proficiency. The rationale is although persons with better language proficiency are more productive, on average paying different wages based on language spoken may be deemed overtly discriminatory. Nevertheless, sorting with respect to employment might be present.

Thereby, we would expect a higher participation probability and higher employment chances of foreigners with a better command of German language in case of an equal pay policy within firms. On the other hand, according to theory, there is a critical level of productivity for participation, b. Persons with productivity below this value do not participate. Therefore, the participating individuals with a good language command could have lower values of υ as higher values of H compensate for this to reach the critical level of productivity. Alternatively higher values of υ may compensate lower values of H. Moreover, one needs to keep in mind that the most productive workers are more likely to be employed in high-paid firms raising out another source of self-selection.

In the theoretical model above it was shown that participating and employed individuals may be non-random samples of the population. Thus, to estimate the earnings equation we have to control for self-selection modeling participation and employment decision simultaneously. The approach could be interpreted as a variant of the classical Heckman-Lee method, following Heckman (1976) and Lee (1976), that is applied when one source of self-selection is present. In our case, we have to regard two sources (participation and employment) in a first step, requiring certain adjustments that are discussed below. Here, the participation equation is given as:

$$I_1^* = Z_1\gamma_1 + \epsilon_1, \tag{2}$$

where Z_1 is a matrix of exogenous variables, γ_1 is a parameter vector, and ϵ_1 is a random component. I_1^* is latent, instead we observe $I_1 = 1$ (in case of participation) if $w^R \geqslant b$ and $I_1 = 0$

(in case of nonparticipation) otherwise where w^R denotes the reservation wage. The employment equation is specified analogously as:

$$I_2^* = Z_2\gamma_2 + \epsilon_2, \tag{3}$$

where Z_2 is a matrix of exogenous variables, γ_2 is a parameter vector, and ϵ_2 is a random component. I_2^* is latent, instead we observe $I_2 = 1$ (in case of employment) if $w \geqslant w^R$ and $I_2 = 0$ (for unemployed individuals) otherwise. Both I_1^* and I_2^* depend on the reservation wage. If Z_1 and Z_2 contain all variables which determine w^R, ϵ_1 and ϵ_2 are independent. If some of these variables are not observed (or not contained) in the data they will be included in the error term, which could result in the correlation between ϵ_1 and ϵ_2. Hence, it might be advisable to allow for this correlation and estimate equations 2 and 3 jointly.

Finally, the wage offer equation is of a standard Becker-Mincer type, $w = X\beta + u$, where w is the log wage, X is a matrix of exogenous variables, β is a parameter vector, and u is an error component, which is normally distributed with mean zero. Wages are observed if both $I_1 = 1$ and $I_2 = 1$. Hence, expected observed wage is given by:

$$E(w|I_1 = 1, I_2 = 1) = X\hat{\beta} + E(u|I_1 = 1, I_2 = 1). \tag{4}$$

Define the covariance between the error terms of the participation and the earnings equation as $\sigma_{u1} = \text{cov}(u, \epsilon_1)$, and analogously between employment and earnings $\sigma_{u2} = \text{cov}(u, \epsilon_2)$. Moreover, let $\text{var}(u) = \sigma_u^2$. In order to estimate the selection model, variances of the error terms have to be standardized as $\text{var}(\epsilon_1) = \text{var}(\epsilon_2) = 1$ and $\text{cov}(\epsilon_1, \epsilon_2) = \rho_1$. Following Mohanty (2001) (see also Maddala, 1983), $E(u|I_1 = 1, I_2 = 1) = \sigma_{u1}\lambda_1 + \sigma_{u2}\lambda_2$, where $\lambda_1 = \phi(Z_1\gamma_1)\Phi(A)/F(Z_1\gamma_1, Z_2\gamma_2; \rho_1)$ and $\lambda_2 = \phi(Z_2\gamma_2)\Phi(B)/F(Z_1\gamma_1, Z_2\gamma_2; \rho_1)$, $A = (Z_2\gamma_2 - \rho_1 \cdot Z_1\gamma_1)/\sqrt{(1 - \rho_1^2)}$, $B = (Z_1\gamma_1 - \rho_1 \cdot Z_2\gamma_2)/\sqrt{(1 - \rho_1^2)}$. ϕ is the univariate standard normal density function, Φ is the univariate standard normal distribution function, and F is the bivariate standard normal distribution function.

It is worth noting that the λ's are the inverse Mill's ratios adjusted for the bivariate case. In fact, if participation and employment decisions are unrelated, i.e. if ρ_1 is not statistically significant, then $F(Z_1\gamma_1, Z_2\gamma_2; \rho_1) = \Phi(A) \cdot \Phi(B)$ (conditional probability of independent events) and hence $\lambda_1 = \phi(Z_1\gamma_1)/\Phi(Z_1\gamma_1)$ and $\lambda_2 = \phi(Z_2\gamma_2)/\Phi(Z_2\gamma_2)$, which are the inverse Mill's ratios in a standard two-stage Heckit model (see Heckman, 1979).

The conditional wage in equation 4 can be rewritten as:

$$E(w|X) = X\hat{\beta} + \lambda_1\sigma_{u1} + \lambda_2\sigma_{u2}. \tag{5}$$

To estimate equation 5, in a first step equations 2 and 3 have to be estimated jointly. Estimates obtained at the first stage $(\gamma_1, \gamma_2, \rho_1)$ are used to construct λ_1 and λ_2 as defined above. At the second stage, wage is regressed on X, λ_1 and λ_2 by OLS which produces the parameter estimates of β, σ_{u1}, and σ_{u2}.

The theoretical model attributes a higher probability of better jobs to workers with a higher productivity. Consequently, even having controlled for self-selection into participation and employment,

earnings may be affected by worker's occupation and economic sector choice (in the subsample of employed individuals). In the empirical analysis, we take account of self-selection into economic sector modeled as the probability of working in basic or high tech industry and manufacturing and of self-selection into occupation modeled as the probability of being a qualified/highly-qualified white-collar worker. Analogously to the basic model, both choices have to be considered as joint decisions in a first step. Assuming joint normality of the errors, estimation is carried out using a bivariate probit model. However, since the participation-employment decision is a double-hurdle, i.e. the employment decision is observed only for participating individuals and is a precondition for occupational and economic sector choice, this has to be regarded in the estimation procedure. A detailed explanation is given in appendix A.

3 Data and selected descriptives

3.1 Data

The empirical analysis is based on data from eight waves of the German Socio-Economic Panel (GSOEP) for the years 1996 to 2005 excluding 2002 and 2004 due to missing information on the variable of interest (language usage at home). GSOEP is a wide-ranging representative longitudinal study of private households carried out since 1984 in Germany. It provides information about all household members covering Germans, foreigners and recent immigrants to Germany. In 2005, there were almost 12,000 households and more than 21,000 persons sampled in GSOEP. An advantage of GSOEP is that it contains an over-sample of migrants to provide a sufficient data base for statistical analyses of the immigrant minorities.

For homogeneity reasons, we impose some restrictions on the sample. First, we use only data on male foreigners in West Germany who have a personal migration experience (first generation immigrants). Foreigners who arrived before 1948 are not considered. Second, we only regard persons from so-called *guest worker* countries. We define persons as guest workers if their country of origin is one of the traditional German recruiting countries from the mid-1950s to 1970s.[5] The countries comprised by our sample are: Turkey (48 percent), Former Yugoslavia (11 percent), Greece (11 percent), Italy (23 percent), Spain (7 percent) and Portugal (1 percent). These persons are a large group of all foreigners in Germany (about 80 percent). Moreover, guest workers seem to be more homogeneous a group by definition. On the one hand, we exclude persons who came as refugees, e.g., from Africa or the Middle East, and, on the other hand, we exclude persons who came from highly advanced countries like the U.S., Western and Northern Europe, or Austria. Third, the analysis takes account of people aged 25 to 55 years to avoid bias due to education or early retirement decisions (those in education are explicitly discarded from our data). In addition, we exclude individuals who do not report the language usage in the household from the sample

---

[5] The goal of the guest worker program was to solve labor supply shortages in West Germany in the industrial sector that emerged during the strong economic development of that time.

(less than 2 percent). Finally, information on wages is symmetrically two percent trimmed to exclude extreme values.

Variables for language proficiency may be prone to measurement error due to self-assessment of the respondents in many surveys. For example, Dustmann and van Soest (2001; 2002) show that reliability of the self-assessed language proficiency variable in GSOEP may be limited in terms of inter-personal and intra-personal comparability.[6] In addition, in about half of the observations self-assessed language proficiency is missing in our sample. Language usage at home of the respondents contains of three categories: 1) speaking mainly German, 2) speaking partly German and partly the mother tongue, or 3) speaking mainly the language of the home country. As it could be expected that reporting the type of the language used in the household is easier than assessing language proficiency in terms of written or oral skills, we suppose the variable to be less prone to measurement errors.

Raw descriptive statistics reveal a strong relationship between self-assessed language proficiency and language usage in the household, suggesting that language spoken in the household could be a good proxy for language proficiency. For example, 51 percent of people who speak mainly German at home report to have *very good* speaking ability (in German) and more than 90 percent report at least *good* speaking ability. More than 60 percent of those who speak partly German at home report to have at least *good* speaking ability and over 30 percent report *satisfactory* speaking ability. Persons speaking mainly mother tongue at home mostly report *satisfactory* (about 45%) and *poor* (about 30%) speaking ability. Although these descriptive statistics tend to support the use of language usage as a proxy for self-assessed language proficiency, differences in interpretation may remain. For that reason, we check the robustness of the results by estimating the effects of self-assessed language proficiency in section 4.4.

Even with the information on labor market states of non-participation, employment and unemployment of GSOEP at hand, modeling the two-stage self-selection process requires some further treatment of the variables. Respondents are asked two separate questions, whether they are registered unemployed and whether they are non-participants. However, non-participation is not necessarily understood by respondents as being out of the labor market in an economic sense (some people mix up non-participation and registered unemployment). For that reason, we define people as non-participants if they responded *not in the labor market* and *not registered as unemployed* simultaneously. It has to be noted that this group may still include some active participants who are not registered at the labor office.[7] A further complication could arise from the fact that respondents do not necessarily understand employment and unemployment as exclusive labor market states. For example, a person having a low-paid job[8] is eligible for receiving additional subsistence allowance. Officially registered unemployed are allowed to hold a minor job or work part-time and earn up to a certain threshold. In the empirical application, unemployed people who are registered at the labor office but earn more than 1,000 Euro per month are counted as

---

[6] The language spoken in the household is also not free from inter-personal variation.

[7] Potential reasons could be, for example, expiration of unemployment benefits eligibility or benefit sanctions.

[8] with earnings below the subsistence level

employed. The outcome variable (real gross hourly wage) is obtained for all employees by dividing the gross earnings in the month prior to the interview by the reported working hours of the last week that are extrapolated to monthly hours. Wages are deflated using the consumption price index based on the year 2000 to get real consumption wages of comparable purchasing power.[9]

In GSOEP, foreigners could leave the sample for two reasons. First, there is some common panel mortality, i.e. persons decide not to participate in subsequent interviews or they change their place of residence and interviewers lose track of them. Second, foreigners could be naturalized. As there is considerable return migration from Germany this may cause non-random panel attrition. Dustmann and van Soest (2002) have analyzed the extent with respect to language ability and earnings. Their results show no significant estimates neither on language ability nor on earnings. Therefore, we could assume panel mortality to be random. In contrast, naturalization could be expected to be non-random. Performing distribution tests of equality of language usage for nationality changers and non-changers showed no significant differences. Overall about 5 percent from our sample of foreigners changed nationality (76 persons out of 1,282). Hence, we refrain from controlling for selection into German citizenship explicitly in the analysis. Persons who were naturalized are not considered in the estimations after the date of naturalization.

3.2   Selected descriptives

Before presenting the empirical application and the estimation results, it is useful to take a closer look at the data available. Table 1 provides means of selected variables used in the empirical model with a distinction between non-participants (first column), persons who participate but are not employed (second column) and employed persons (third column). Instead of discussing the single figures in the table, we will concentrate on findings that are of primary interest for our analysis. Although we use language usage in the household in the empirical analysis, we additionally include self-assessed language proficiency for comparison. First, compared to the other groups employed individuals are more likely to speak mainly German at home. Whereas about 32 percent of the employed persons speak German in the household, the shares in the group of non-participants are 22 percent and in the group of the unemployed 14 percent only. Vice versa, non-participants (38 percent) and unemployed persons (44 percent) speak mainly their mother tongue at home compared to the employed (22 percent). With respect to self-assessed language proficiency the picture is similar. As about 63 percent of the employed persons report a good or very good proficiency this group tends to have a clearly better command of German language than unemployed persons (42 percent) or non-participants (54 percent). All in all, employed persons report a better language ability than the other groups.

A major determinant of productivity is skill level. We consider it in the analysis in three different categories. Low-skilled are people who lack professional training. This group could be expected

---

[9] It should be noted that the reported gross earnings in the month prior to the interview have not been adjusted for end-of-year bonuses, overtime-payments, holiday allowances etc.

Tab. 1: Means of selected variables

	Part.=0	Part. =1	
		Emp.=0	Emp.=1
Language usage in the household			
Mainly German	0.22	0.14	0.32
Mainly mother tongue	0.38	0.44	0.22
Partly German	0.40	0.42	0.46
Self-assessed language proficiency			
Very good	0.22	0.09	0.21
Good	0.32	0.33	0.42
Satisfactory	0.32	0.38	0.28
Poor	0.12	0.19	0.10
Not at all	0.02	0.01	0.00
Level of education			
Low skilled	0.60	0.66	0.51
Medium skilled	0.35	0.28	0.41
High skilled	0.05	0.06	0.08
Socio-demographic variables			
Age	38.79	41.41	39.18
Time of residence	22.89	22.78	23.55
Married	0.75	0.83	0.86
Child	0.80	0.75	0.78
Economic sector			
Manufacturing	–	–	0.65
Transportation	–	–	0.05
Construction	–	–	0.13
Trading services	–	–	0.14
Social services and health	–	–	0.03
Firm size			
fewer than 20 employees	–	–	0.18
20-199 employees	–	–	0.28
200-1999 employees	–	–	0.32
2000 and more employees	–	–	0.20
Observations	255	290	2,285

Part.=Participation; Emp.=Employment; – not available

to be most strongly disadvantaged in the labor market, in particular in a high-developed country with a regulated labor market as Germany. Medium-skilled comprise people who completed a professional training (not necessarily in the German apprenticeship system, but comparable to it). Finally, high-skilled persons are those who graduated from advanced technical college (*Fachhochschule*) or university. The shares of the low-skilled are particularly large in the group of non-participants (60 percent) and the group of unemployed persons (66 percent). In contrast, the share is lower in the group of employed persons (51 percent) but still considerable. With regard to the socio-demographic characteristics further heterogeneity could be obtained. The non-participants are, on average, about half a year younger than the employed; unemployed persons, in contrast, are more than two years older than the employed. The time of residence is comparable between groups but employed persons have lived in Germany longer than the other groups. Finally, taking a look at the economic sectors and firm sizes of the employed persons shows that the majority works in manufacturing, trading services and construction and more than half are employed at firms with at least 200 employees.

The last point we want to present descriptively is the correlation between language ability and

Tab. 2: Distribution of language ability across skill groups

	low skilled	med. skilled	high skilled
	Language usage in the household		
Mainly German	0.24	0.34	0.43
Partly German	0.32	0.20	0.18
Mainly mother tongue	0.45	0.46	0.39
	Self-assessed language proficiency		
Very Good	0.14	0.25	0.31
Good	0.38	0.44	0.31
Satisfactory	0.34	0.22	0.30
Poor	0.13	0.08	0.07
Not at all	0.00	0.01	0.01

level of education (Table 2). In general, language ability increases with the skill level. Low-skilled persons show the lowest language proficiency irrespective if it is measured by language usage or self-reported ability. High-skilled workers have higher variance in self-reported language ability compared to medium-skilled. While more high-skilled workers report very good language ability this is also true for satisfactory language ability, which could imply that high-skilled are too heterogeneous a group.

4 Empirical results

4.1 Participation and employment

To answer the question what impact language ability has on foreigners' earnings, we will start our discussion with the effects of language proficiency on the decisions on participation and employment. As shown in the set-up of the empirical model, individual's decisions on participation and employment may be correlated. Hence, in a first step we estimate both decisions jointly using full-information maximum likelihood. For identification purposes not to solely rely on distributional assumptions, we choose marital status (*married*) and children (*children*) as the exclusion restrictions that enter the participation equation, but not the employment equation. Marital status has a statistically significant effect, thereby, justifying the choice of exclusion variables. The correlation between the errors was statistically insignificant and, therefore, we estimate both equations separately.

The variables of language usage are included in both equations. Effects are estimated with dummies for speaking mainly German language and for speaking partly German language at home relative to speaking mother tongue (the reference group). Moreover, to improve explanatory power, a number of socio-economic variables are added into the models. In particular, person's age and time of residence (i.e. the years the individual lives in Germany) as well as the squares are considered in both equations. We do not include immigration cohort effects in the model as they are a perfect linear combination of age, time of residence and year dummies. Therefore, we refrain

from interpreting effects of age or time of residence as we cannot identify them separately from cohort effects. It should be noted that the effect of language proficiency is unaffected by the exclusion of cohort effects. Those are absorbed in the parameter estimates for the constant term. As productivity is closely related to qualification, we estimate the effect of medium- or high-qualification in reference to low-qualification.

Another issue, which one has to deal with, are differences in requirements of language proficiency across jobs. It may be possible that low-skilled workers mainly supplying manual labor are less likely to need a good language command in order to perform their tasks, unlike their high-skilled counterparts who need the language of the host country on a daily basis for communication and decision-making. For that reason we interact language usage with skill dummies. As a result the effect of language usage on earnings for each skill group is a sum of the coefficients of language usage and the respective skill dummy. The effects of language usage on participation and employment across skill groups are given in Table 3.[10] The estimates of the language variables clearly point towards a positive effect of usage of the host country's language on both the decision to participate in the labor market and the employment chances. Although the coefficients in the table cannot be interpreted as marginal effects, it becomes obvious from the relative scale that speaking mainly German has an even stronger effect than using it only partly. Surprisingly, no significant effect is found for high skilled workers. However, given that the share of non-participants and unemployed among the high skilled is only about 5%, the overall number of high skilled workers is relatively low in the sample, and given possible heterogeneity of this group one should be cautious in interpreting insignificance as non-existence of the effect. The scale of the point estimates supports this argument.[11]

Tab. 3: Language usage effects on participation and employment across skill groups

	participation[a]		
	low skilled	med. skilled	high skilled
mainly German	0.3511*	0.6010#	0.7771*
partly German	0.3248#	0.3277**	0.5769
	employment[b]		
	low skilled	med. skilled	high skilled
mainly German	0.5653#	0.9404#	0.1461
partly German	0.3193**	0.4711#	-0.0576

- significant at 0.1% level, *** - at 1% level, ** - at 5% level, * - at 10% level
[a] The complete estimates are given in Table B.1 (first column) in the appendix.
[b] The complete estimates are given in Table B.1 (second column) in the appendix.

With regard to other variables (see Table B.1 in the appendix), most findings are in line with the expectations. Considering the level of education, medium-skilled people experience a higher

---

[10] The complete estimates are given in Table B.1 in the appendix.

[11] At this point, one should bear in mind that with respect to language ability high-skilled workers are more heterogeneous a group (see Table 2) than the other skill groups.

propensity to participate than the low-skilled. Maybe due to the small number of foreigners who are high-skilled and possible heterogeneity of this group, estimations do not provide evidence that those differ in their behavior from the low-skilled in terms of participation.

4.2   Impact of language usage on earnings

Having established selection patterns into participation and employment conditional on language usage, we want to analyze the effects on earnings in the next step. As a starting point we estimate the effects of language usage on earnings without controlling for self-selection. These results are summarized in the upper panel of Table 4.[12] Without controlling for selection, foreigners speaking mainly German earn about 5 percent more than foreigners speaking their mother tongue at home. For high skilled workers the gap is as large as 39 percent. Speaking partly German has a significant effect for high skilled and is smaller in magnitude (about 17 percent compared to speaking mother tongue).

Tab. 4: Language usage effects on earnings across skill groups

	Earnings (no selection control)[a]		
	low skilled	med. skilled	high skilled
mainly German	0.0531*	0.0709**	0.3857#
partly German	-0.0226	0.0525*	0.1706**
	Earnings (employment, participation selection)[b]		
	low skilled	med. skilled	high skilled
mainly German	0.0750	0.0855	0.3618#
partly German	-0.0171	0.0617*	0.1431*

   # - significant at 0.1% level, *** - at 1% level, ** - at 5% level, * - at 10% level

[a] The complete estimates are given in Table B.2 (model 1) in the appendix.

[b] The complete estimates are given in Table B.2 (model 2) in the appendix.

To consider self-selection of participation and employment, we calculate both inverse Mill's ratios based on the results from Table B.1. As shown in the set-up of the econometric model (section 2.2) both terms are plugged into the earnings equation as additional regressors. The corresponding estimates of the effects of language usage are summarized in the lower panel of Table 4.[13] Controlling for selection into participation and employment increases the earnings gap between foreigners speaking mainly German and foreigners speaking mainly their mother tongue by about 2 percentage points for the low skilled and by about 1.5 percentage points for the medium-skilled. In contrast, the effect for the high-skilled is reduced by about 2.5 percentage points.

These results so far point out at a wage premium for language proficiency even controlling for selection into participation and employment. This result is not new and has already been discussed

---

[12] Complete results are given in model 1 of Table B.2 in the appendix.

[13] Model 2 of Table B.2 in the appendix contains the complete estimation results.

in the literature. For example, Dustmann (1994) notes a 7.3 percent wage increase for males who report to have good or very good writing abilities in German. Chiswick and Miller (1999) report higher wages by about 8 percent for migrant males who are proficient in both speaking and reading English using the 1989 Legalized Population Survey (LPS) for the United States. For Great Britain Shields and Price (2002) establish that language fluency increases the mean occupational wage by about 16.5 percent. Chiswick, Lee, and Miller (2005) find out that immigrants who are proficient in English have 19 percent higher earnings than those with limited English language skills using the Longitudinal Survey of Immigrants to Australia 1993-1995. For Israel Berman, Lang, and Siniver (2003) predict a 23 percent earnings' increase for immigrants from the former Soviet Union who fluently speak Hebrew in 1994. Nevertheless, the results presented here clarify that neglecting self-selection of employed immigrants leads to biased estimates of the earnings effects.

4.3   Considering self-selection into economic sector and occupation

However, selection into occupation and industrial sector was not regarded explicitly in earlier studies. Put it in other words, are immigrants with better language proficiency paid better in a firm than immigrants with poorer language command or do they earn different wages because they work in different firms and do different jobs? Theory predicts that workers with higher productivity enter higher paid jobs. Therefore, even having controlled for self-selection into participation and employment, earnings may be affected by worker's occupation and economic sector choice. Further, we distinguish between blue- and white-collar workers and basic/high-tech industry and other economic sectors. Table 5 presents descriptive statistics of the language ability variables and skills for these groups. As it becomes obvious, there is a substantial difference in language ability (independently whether measured in terms of language usage or self-assessed language proficiency) between blue-collar and white-collar workers. White-collar workers are more likely to speak mainly German in the household, be more fluent in German language and be high-skilled. In contrast, comparing basic and high-tech industry with other industry establishes minor differences only.

Due to these differences, it is reasonable to estimate the separate earnings equations for each of the four groups: 1. white-collar in basic and high-tech industry, 2. blue-collar in basic and high-tech industry, 3. white-collar in other industry, and 4. blue-collar in other industry. Each of these groups is non-random and appropriate correction for self-selection is required (see section A in the appendix for a description). Estimation is carried out using bivariate probit models on choice of occupation type and economic sector. Complete estimates are given in Table B.4 in the appendix. The estimate of the correlation coefficient (ρ) is highly significant, hence, both choices have to be estimated jointly to avoid selection bias. The estimates of language usage across skill groups in both equations are summarized in Table 6. The results show that foreigners speaking mainly German at home have a clearly higher probability to be white-collar workers; speaking partly German is relevant for the probability of being a white-collar worker for medium-skilled persons.

Having constructed appropriate inverse Mill's ratios we estimated earnings equations for each

14

Tab. 5: Descriptives statistics by occupation and economic Sector

| | Occupation | | Economic sector | |
	white collar	blue collar	basic & high-tech industry	other industry
	Language usage in the household			
Mainly German	0.67	0.29	0.29	0.36
Partly German	0.03	0.24	0.24	0.21
Mainly mother tongue	0.29	0.47	0.48	0.44
	Self-reported German language proficiency			
Very Good	0.58	0.17	0.17	0.25
Good	0.41	0.42	0.44	0.40
Satisfactory	0.01	0.30	0.31	0.24
Poor	0.00	0.11	0.08	0.11
Not at all	0.00	0.00	0.01	0.00
	Level of education			
Low skilled	0.21	0.54	0.54	0.49
Medium skilled	0.37	0.41	0.41	0.41
High skilled	0.42	0.05	0.06	0.10

Tab. 6: Language usage effects on occupation and economic sector choice across skill groups[a]

| | Occupation choice | | |
	low skilled	med. skilled	high skilled
mainly German	0.7715**	0.7266**	1.9069#
partly German	0.2800	0.5411**	0.8115
	Economic sector choice		
	low skilled	med. skilled	high skilled
mainly German	-0.2693	-0.0543**	-0.8583*
partly German	-0.1739	0.1383	-0.4057

- significant at 0.1% level, *** - at 1% level, ** - at 5% level, * - at 10% level

[a] The complete estimates are given in Table B.4 in the appendix.

group restricting the β-coefficients to be equal across groups (Table B.2, model 3) and for two groups excluding white-collar workers who are a relatively small group (Table B.2, models 4 and 5). The effects of language usage and skills are summarized in Table 7. Consideration of all groups points out at a slightly significant effect of language usage on earnings especially for high-skilled workers. Excluding white-collar workers (156 observations out of 2,021) reduces the magnitude of the effect and its significance. In fact, none of the coefficients for the language effect on earnings are significant. This implies that language usage affects earnings of white- and blue-collar workers differently and occupational heterogeneity of workers should not be neglected. For blue-collar workers language proficiency does not seem to affect wages directly. It influences wages indirectly by affecting participation and employment probability as well as occupational choice (white- or blue-collar occupation). For white-collar workers the results tend to indicate that a wage premium for language proficiency exists, but given the small number of observations for white-collar workers we can neither prove it nor disprove.

Although these results are not directly comparable due to differences in the selection criteria of the sample and the methods used to estimate the effects of occupation on earnings, it may be useful

Tab. 7: Language usage effects on earnings across skill groups (additionally controlled for occupation and economic sector)

	low skilled	med. skilled	high skilled
	All groups		
mainly German	0.0985*	0.1116*	0.2690**
partly German	-0.0111	0.0854*	0.1412*
	Excluding white-collar workers		
mainly German	0.0855	0.0905	0.0961
partly German	-0.0116	0.0716	0.0932

- significant at 0.1% level, *** - at 1% level, ** - at 5% level, * - at 10% level

to discuss the findings in light of the results of Kossoudji (1988) and Chiswick and Miller (2007a). In contrast to our results, Kossoudji (1988) finds significant negative effect of language deficiency on earnings of craft and service personnel (blue-collar occupations). However, the results are not directly comparable as she analyzes Hispanics and East Asians in the US labor market in the 1970s which is different to German. On the other hand, selection into participation and employment is not considered in Kossoudji (1988). Chiswick and Miller (2007a) control for occupational choice by including dummy variables for different occupations. Their findings suggest that about one half of the typical earnings differential reported in the literature with respect to language ability can be attributed to occupational attainment, i.e. workers with better language command enter higher paid occupations. This is consistent with our finding that controlling for occupation reduces the observed wage premium for language proficiency. However, unlike Chiswick and Miller (2007a) we find that for blue-collar workers in Germany the effect of language proficiency on earnings is to full extent attributed to higher participation and employment rates and better occupational attainment of persons with better command of German language.

4.4 Language usage vs. language proficiency

So far we have used the information on the language usage in the household as a proxy for actual language proficiency. However, there may be concerns about the validity of language usage in the household to approximate the effects of language proficiency. The problem of approximation of language proficiency by self-assessed variables is common in many studies. For that reason, different language variables are used (as single variables or combinations). For Germany, e.g., Dustmann (1994), uses speaking and writing fluency, Dustmann and van Soest (2001; 2002) use information on speaking fluency compressed by a dummy instead of a five-scale variable. Other examples are McManus, Gould, and Welch (1983) who construct a combined indicator derived from answers to questions on language usage in the household and neighborhood. In a similar fashion, Chiswick and Miller (1995) and Chiswick and Miller (2007a) use combined information on language usage in the household and self-assessed speaking fluency. In constrast, Rivera-Batiz (1990) has access to data on objective test scores of literacy (see above). We have redone our analysis using information on self-assessed language proficiency to provide a sensitivity analysis

of the results discussed so far. Unfortunately, as noted above self-assessed language proficiency is missing in about half of the observations in our sample.

Denote the latent language proficiency by H. The self-assessed language proficiency is reported in 5 categories: *not at all*, *poor*, *satisfactory*, *good*, and *very good*. The probability that an outcome i is observed is given by:

$$Pr(O = i) = Pr(\kappa_{i-1} < H \leqslant \kappa_i), \tag{6}$$

where κ are the cutoff points and O is the observed self-assessed language proficiency. Equation 6 could be estimated by ordered probit method.[14] As a set of explanatory variables we choose age, time of residence (and the respective squares), skill and year dummies, and language usage in the household. The estimation results are presented in Table 8. Not surprisingly, persons who speak mainly German in the household also have better command of German language. Speaking partly German also increases language proficiency. However, the effect is smaller in magnitude.

Tab. 8: Ordered probit estimates on proficiency in spoken German language

Age	-0.0195
Age (squared)	-0.0004
Time of residence	0.1223#
Time of residence (squared)	-0.0015#
Level of education	
Medium skilled	0.1880**
High skilled	0.3897**
Language usage	
Mainly German	1.8088#
Partly German	0.8159#
Year dummies	YES
κ_1	-2.1356
κ_2	-0.2985
κ_3	1.1160
κ_4	2.7256
pseudo R^2	0.2338
Observations	1,314

*** - significant at 1% level, ** - at 5% level, * - at 10% level

Using the fitted values $\hat{H}$ as a regressor instead of language usage in the household we re-estimate equations 2, 3, A.1, and extended earnings equation 5. Given the parametric form $H = Q\eta + \zeta$ we also impute the missing values. Namely, O is observed in only about half of observations, and hence, $\hat{\eta}$ is obtained from the regression in eq. 6 using non-missing observations in O (half of the sample). However, once $\hat{\eta}$ is estimated and given that Q is observed for the whole sample, $\hat{H} = Q\hat{\eta}$ can be calculated for missing observations as well.

The results for earnings are given in the appendix in Table B.3, the corresponding estimates for

---

[14] Bleakley and Chin (2004) uses IV method to correct for measurement error in the self-assessed language proficiency variable. However, they fail to take into account the ordinal nature of the self-assessed proficiency variable.

the selection models in Tables B.1 and B.4. Qualitatively the results are similar to the estimates using language usage in the household. Self-assessed language proficiency significantly increases participation probability and employment chances especially for the medium- and the high-skilled. Besides, foreigners with better language proficiency are more likely to be white-collar workers. Without controlling for selection language proficiency has a significant positive effect on earnings especially for medium and high skilled persons. Taking selection into participation and employment into account slightly increases the magnitudes of the effect. However, once additional selection into economic sector and occupation is considered, the effect of language proficiency on earnings becomes statistically insignificant for blue-collar workers. In that sense, it could be concluded that independently of the language variables used, language ability affects the selection processes in the labor market but a direct effect of language on earnings (as a kind of wage premium) could not be established.

5   Conclusion

There is quite comprehensive international evidence showing that foreigners speaking the language of the host country well are better off in terms of earnings than those with poor command. One of the explanations is a relationship between productivity and language proficiency, and more productive workers are paid a wage premium. However, this explanation is not easy to reconcile with the empirical findings due to self-selection of workers.

Put it in other words, are immigrants with better language proficiency paid better in a firm than immigrants with poorer language command or do they earn different wages because they work in different firms and do different jobs? To shed more light on this issue we analyze earnings of foreigners in Germany considering selection of workers into participation, employment, economic sector, and occupation.

We begin our analysis by building a theoretical model which assumes no wage premium to productivity (and hence to language proficiency). We show that in the absence of wage premium, language proficiency heterogeneity would generate differences in participation and employment levels. Moreover, more proficient individuals end up working in higher-paid firms. In the empirical model, we show that without controlling for self-selection or controlling only for selection into participation and employment the estimates support the existence of wage premium for language proficiency. However, once selection into economic sector and occupation is controlled for, no significant effects of language proficiency on earnings of blue-collar workers are observed. For white-collar workers no conclusive evidence is found due to small number of observations.

Language proficiency, however, plays an important role at different selection stages. For example, it significantly increases participation probability and employment chances, especially for the low- and medium-skilled. Moreover, foreigners with better language command are more likely to be white-collar workers. This implies that language proficiency only indirectly affects foreigners' earn-

ings in Germany through higher participation rate and employment chances and higher likelihood of working in white-collar occupations. A wage premium per se or a direct effect could not be established. Using two different measures of language proficiency (language usage in the household and self-assessed language proficiency) does not change the results qualitatively.

References

BEENSTOCK, M., B. R. CHISWICK, AND A. PALTIEL (2005): "Endogenous Assimilation and Immigrant Adjustment in Longitudinal Data," Discussion Paper 1840, IZA.

BEENSTOCK, M., B. R. CHISWICK, AND G. L. REPETTO (2001): "The Effect of Linguistic Distance and Country of Origin on Immigrant Language Skills: Application to Israel," *International Migration*, 39(3), 33–60.

BERMAN, E., K. LANG, AND E. SINIVER (2003): "Language-Skill Complementarity: Returns to Immigrant Language Acquisition," *Labour Economics*, 10, 265–290.

BLEAKLEY, H., AND A. CHIN (2004): "Language Skills and Earnings: Evidence from Childhood Immigrants," *Review of Economics and Statistics*, 86(2), 481–496.

BURDETT, K., AND D. MORTENSEN (1998): "Wage Differentials, Employer Size and Unemployment," *International Economic Review*, 39, 257–273.

CARLINER, G. (1981): "Wage Differences by Language Group and the Market for Language Skills in Canada," *Journal of Human Resources*, 16(3), 384–399.

CHARETTE, M., AND R. MENG (1994): "Explaining Language Proficiency: Objective versus Self-Assessed Measures of Literacy," *Economics Letters*, 44, 313–321.

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

——— (1991): "Speaking, Reading, and Earnings among Low-skilled Immigrants," *Journal of Labor Economics*, 9(2), 149–170.

CHISWICK, B. R., Y. L. LEE, AND P. W. MILLER (2005): "Immigrant Earnings: A Longitudinal Analysis," *Review of Income and Wealth*, 51(4), 485–503.

CHISWICK, B. R., AND P. W. MILLER (1995): "The Endogeneity between Language and Earnings: International Analyses," *Journal of Labor Economics*, 13(2), 246–288.

——— (1999): "Language Skills and Earnings Among Legalized Aliens," *Journal of Population Economics*, 12, 63–89.

——— (2007a): "Earnings and Occupational Attainment: Immigrants and the Native Born," Discussion Paper 2676, IZA.

———— (2007b): "Modeling Immigrants' Language Skills," Discussion Paper 2974, IZA.

DUSTMANN, C. (1994): "Speaking Fluency, Writing Fluency and Earnings of Migrants," *Journal of Population Economics*, 7, 133–156.

———— (1999): "Temporary Migration, Human Capital, and Language Fluency of Migrants," *Scandinavian Journal of Economics*, 101(2), 297–314.

DUSTMANN, C., AND F. FABBRI (2003): "Language Proficiency and Labour Market Performance of Immigrants in the UK," *The Economic Journal*, 113, 695–717.

DUSTMANN, C., AND A. VAN SOEST (2001): "Language Fluency and Earnings: Estimation with Misclassified Language Indicators," *Review of Economics and Statistics*, 83(4), 663–674.

———— (2002): "Language and the Earnings of Immigrants," *Industrial and Labor Relations Review*, 55(3), 473–492.

ESPENSHADE, T. J., AND H. FU (1997): "An Analysis of English-Language Proficiency among U.S. Immigrants," *American Sociological Review*, 62(2), 288–305.

FISHE, R., R. TROST, AND P. LURIE (1981): "Labor Force Earnings and College Choice of Young Women: An Examination of Selectivity Bias and Comparative Advantage," *Economics of Education Review*, 1(2), 169–191.

GRENIER, G. (1984): "The Effect of Language Characteristics on the Wages of Hispanic-American Males," *Journal of Human Resources*, 19, 25–52.

HECKMAN, J. (1976): "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement*, 5, 475–492.

———— (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1), 153–161.

KOSSOUDJI, S. (1988): "English Language Ability and the Labor Market Opportunities of Hispanic and East Asian Immigrant Men," *Journal of Labor Economics*, 6(2), 202–228.

LEE, L. F. (1976): "Estimation of Limited Dependent Variables by Two-Stage Methods," Unpublished Ph.D. Dissertation, University of Rochester.

MADDALA, G. (1983): *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press.

MANNING, A. (2003): *Monopsony in Motion: Imperfect Competition in Labor Markets*. Princeton University Press.

MCMANUS, W., W. GOULD, AND F. WELCH (1983): "Earnings of Hispanic Men: The Role of English Language Proficiency," *Journal of Labor Economics*, 1(2), 101–130.

MOHANTY, M. (2001): "Determination of Participation Decision, Hiring Decision, and Wages in a Double Selection Framework: Male-Female Wage Differentials in the U.S. Labor Market Revisited," *Contemporary Economic Policy*, 19(2), 197–212.

MORTENSEN, D. T. (2003): *Wage Dispersion: Why Are Similar Workers Paid Differently?* Cambridge, Massachusetts.

REIMERS, C. (1983): "Labor Market Discrimination against Hispanic and Black Men," *Review of Economics and Statistics*, 65(4), 570–579.

——— (1985): "A Comparative Analysis of the Wages of Hispanics, Blacks and Non-Hispanic Whites," in *Hispanics in the U.S. Labor Market*, ed. by G. Borjas, and M. Tienda, pp. 27–76. Academic Press, Orlando.

RIVERA-BATIZ, F. (1990): "English Language Proficiency and the Economic Progress of Immigrants," *Economics Letters*, 34(3), 295–300.

SHIELDS, M. A., AND S. W. PRICE (2002): "The English Language Fluency and Occupational Success of Ethnic Minority Immigrant Men Living in English Metropolitan Areas," *Journal of Population Economics*, 15, 137–160.

TAINER, E. (1988): "English Language Proficiency and the Determination of Earnings Among Foreign-Born Men," *Journal of Human Resources*, 23, 108–122.

A Modeling selection into economic sector and occupation

Theory predicts that workers with higher productivity enter higher paid jobs. Therefore, even having controlled for self-selection into participation and employment, earnings may be affected by worker's occupation and economic sector choice. Therefore, we extend our model to explicitly control for selection through these channels. In the sub-sample of employed persons, we take account of self-selection into the economic sector modeled as the probability of working in basic or high tech industry and of self-selection into occupation modeled as the probability of being a white-collar worker. Analogously to our empirical model discussed in section 2.2, we consider both choices as joint decisions in a first step. Assuming joint normality of the errors, estimation is carried out using a bivariate probit model.

However, one needs to keep in mind that participation-employment decision is a double-hurdle, i.e. employment decision is observed only for participating individuals. In case of occupation-industry decision all combinations are observed: 1. white-collar in basic and high tech industry, 2. blue-collar in basic and high tech industry, 3. white-collar in other industry, and 4. blue-collar in other industry. Let the two choices be modeled as:

$$I_3 = Z_3\gamma_3 + \epsilon_3$$
$$I_4 = Z_4\gamma_4 + \epsilon_4,$$

with $\text{cov}(\epsilon_3, \epsilon_4) = \rho_2$ and I_3 being an indicator variable taking the value one if a person is a white-collar worker and I_4 being an indicator variable taking the value one if a person is employed in basic or high-tech industry. There is an earnings equation defined for each group (see Fishe, Trost, and Lurie (1981)):

$$
\begin{aligned}
w_1 &= X_1\beta_1 + u_1 && \text{if} && I_3 = 1, I_4 = 1 \\
w_2 &= X_2\beta_2 + u_2 && \text{if} && I_3 = 0, I_4 = 1 \\
w_3 &= X_3\beta_3 + u_3 && \text{if} && I_3 = 1, I_4 = 0 \\
w_4 &= X_4\beta_4 + u_4 && \text{if} && I_3 = 0, I_4 = 0.
\end{aligned}
\tag{A.1}
$$

It is assumed that $u_{k \in K}$ with $K = \{1, 2...4\}$ is jointly distributed with ϵ_3 and ϵ_4 with $E(u_{k \in K}\epsilon_{\ell \in L}) = \sigma_{k\ell}$ where $L = \{1, 2\}$. Since w_1 is observed only if $I_3 = 1, I_4 = 1$ then $E(w_1|I_3 = 1, I_4 = 1) = X_1\hat{\beta}_1 + E(u_1|I_3 = 1, I_4 = 1)$. It is shown in Fishe, Trost, and Lurie (1981):

$$
\begin{aligned}
E(u_1|I_3 = 1, I_4 = 1) &= \sigma_{13}M_{12} + \sigma_{14}M_{21} \\
E(u_2|I_3 = 0, I_4 = 1) &= \sigma_{23}M_{34} + \sigma_{24}M_{43} \\
E(u_3|I_3 = 1, I_4 = 0) &= \sigma_{33}M_{56} + \sigma_{34}M_{65} \\
E(u_4|I_3 = 0, I_4 = 0) &= \sigma_{43}M_{78} + \sigma_{44}M_{87}.
\end{aligned}
\tag{A.2}
$$

The Ms are Mills ratios adjusted for bivariate case. The general formula for calculating Ms is

$M_{ij} = (1 - \rho_2^2)^{-1}(P_i - \rho_2 P_j)$ with Ps defined by the following equation:

$$P_1 = \frac{\int_{-\infty}^{Z_4\gamma_4} \int_{-\infty}^{Z_3\gamma_3} \epsilon_3 f(\epsilon_3\epsilon_4)d\epsilon_3 d\epsilon_4}{F(Z_3\gamma_3, Z_4\gamma_4)}, \qquad P_2 = \frac{\int_{-\infty}^{Z_3\gamma_3} \int_{-\infty}^{Z_4\gamma_4} \epsilon_4 f(\epsilon_3\epsilon_4)d\epsilon_4 d\epsilon_3}{F(Z_3\gamma_3, Z_4\gamma_4)},$$

$$P_3 = \frac{\int_{-\infty}^{Z_4\gamma_4} \int_{Z_3\gamma_3}^{\infty} \epsilon_3 f(\epsilon_3\epsilon_4)d\epsilon_3 d\epsilon_4}{F(-Z_3\gamma_3, Z_4\gamma_4)}, \qquad P_4 = \frac{\int_{Z_3\gamma_3}^{\infty} \int_{-\infty}^{Z_3\gamma_4} \epsilon_4 f(\epsilon_3\epsilon_4)d\epsilon_4 d\epsilon_3}{F(-Z_3\gamma_3, Z_4\gamma_4)}, \qquad \text{(A.3)}$$

$$P_5 = \frac{\int_{Z_4\gamma_4}^{\infty} \int_{-\infty}^{Z_3\gamma_3} \epsilon_3 f(\epsilon_3\epsilon_4)d\epsilon_3 d\epsilon_4}{F(Z_3\gamma_3, -Z_4\gamma_4)}, \qquad P_6 = \frac{\int_{-\infty}^{Z_3\gamma_3} \int_{Z_4\gamma_4}^{\infty} \epsilon_4 f(\epsilon_3\epsilon_4)d\epsilon_4 d\epsilon_3}{F(Z_3\gamma_3, -Z_4\gamma_4)},$$

$$P_7 = \frac{\int_{Z_4\gamma_4}^{\infty} \int_{Z_3\gamma_3}^{\infty} \epsilon_3 f(\epsilon_3\epsilon_4)d\epsilon_3 d\epsilon_4}{F(-Z_3\gamma_3, -Z_4\gamma_4)}, \qquad P_8 = \frac{\int_{Z_3\gamma_3}^{\infty} \int_{Z_4\gamma_4}^{\infty} \epsilon_4 f(\epsilon_3\epsilon_4)d\epsilon_4 d\epsilon_3}{F(-Z_3\gamma_3, -Z_4\gamma_4)}.$$

The Ps can be interpreted as conditional expectations of errors in equation (A.1) as $E(\epsilon_3|I_3 = 1, I_4 = 1) = \frac{\int_{-\infty}^{Z_4\gamma_4} \int_{-\infty}^{Z_3\gamma_3} \epsilon_3 f(\epsilon_3\epsilon_4)d\epsilon_3 d\epsilon_4}{F(Z_3\gamma_3, Z_4\gamma_4)} = P_1$ and so on. Equation A.3 can still be evaluated numerically. In fact, the term M_{12} and M_{21} are calculated by the same token as λ_1 and λ_2 for which the numerical evaluation is straightforward. Other terms are calculated in the same fashion.

B   Tables

Tab. B.1: Selection model: results for separate estimation of participation and employment decision

	Language usage		Language proficiency	
	Part.	Emp.	Part.	Emp.
Constant	0.5768	-0.0377	0.5887	-0.0094
Age	0.0055	0.0401	0.0150	-0.0495
Age(squared)	0.0001	-0.0006	0.0001	-0.0005
Time of residence	0.0068	0.0275**	-0.0251	0.0000
Time of residence(squared)	-0.0002	-0.0004*	0.0001	-0.0003
Level of Education (Ref. low skilled)[a]				
Medium skilled	-3.7063*	-3.8598	-2.7681*	-3.9856*
High skilled	-0.3958	3.8949	0.5533	-1.3900
Language Usage (Ref. mainly mother tongue)[b]				
Mainly German	0.3511*	0.5653#	—	—
Partly German	0.3248#	0.3193**	—	—
Language Proficiency	—	—	0.2124**	0.3193#
Exclusion Restrictions (Participation)				
Married	0.4741**	—	0.4835***	—
Children	-0.2159	—	-0.2172	—
Interactions				
Medium skilled*age	0.1702*	0.2222*	0.1653*	0.2160*
Medium skilled*age (squared)	-0.0024*	-0.0029*	-0.0024*	-0.0028*
High skilled*age	-0.0001	0.1138	-0.0704	0.1097
High skilled*age(squared)	0.0003	-0.0016	0.0012	-0.0016
Mainly German*Medium skilled	0.2500	0.3752*	—	—
Mainly German*High skilled	0.4260	-0.4192	—	—
Partly German*Medium skilled	0.0029	0.1518	—	—
Partly German*High skilled	0.2521	-0.3770	—	—
Language Proficiency*Medium skilled	—	—	0.0952	0.1969
Language Proficiency*High skilled	—	—	0.3106*	-0.1225
Linear combinations Medium-skilled				
Mainly German	0.6010***	0.9404#	—	—
Partly German	0.3277**	0.4711#	—	—
Language Proficiency	—	—	0.3076***	0.5162#
Linear combinations High-skilled				
Mainly German	0.7771*	0.1461	—	—
Partly German	0.5769	-0.0576	—	—
Language Proficiency	—	—	0.5230***	0.1968
Year dummies	YES	YES	YES	YES
Observations	2,830	2,575	2,830	2,575
pseudo R^2	0.0534	0.0770	0.0533	0.0784

All estimates are displayed in terms of coefficients. Significance is indicated as follows: # denoting the 0.1%, *** the 1%, ** the 5% and * the 10% level.

[a] Medium-skilled are people with professional training, high-skilled are people with advanced technical college or university degree.

[b] Language proficiency is estimated by the ordered probit regression. See section 4.4 for details.

Tab. B.2: Earnings equation and language usage in the household[a]

	Model 1	Model 2	Model 3	Model 4	Model 5
Constant	1.6260#	1.5633#	1.4986#	1.5469#	1.5264#
Age	0.0280***	0.0306**	0.0310**	0.0307**	0.0301**
Age(squared)	-0.0003**	-0.0004**	-0.0004	-0.0004**	-0.0004**
Time of residence	0.0135	0.0150#	0.0188#	0.0156#	0.0182#
Time of residence(squared)	-0.0002**	-0.0002**	-0.0003#	-0.0003#	-0.0003#
Level of Education[b](Ref. low skilled)					
Medium skilled	-0.0085	0.0003	0.0003	0.0070	0.0018
High skilled	-0.0920	-0.0609	-2.0326	-0.0716	-0.0609
Language usage (Ref. mainly mother tongue)					
Mainly German	0.0531*	0.0750	0.0985*	0.0921*	0.0855
Partly German	-0.0226	-0.0171	-0.0111	-0.0084	-0.0116
Sector (Ref. industry & manufacturing)					
Transportation	-0.1264#	-0.1254#	-0.0596	-0.1423#	-0.0587
Construction	-0.0092	-0.0082**	0.0664**	-0.0092	0.0725**
Trading services	-0.1218#	-0.1206#	-0.0763	-0.1536#	-0.0724*
Social services and health	-0.0874**	-0.0899**	-0.0517	-0.1172**	-0.0343*
Firm size (Ref. < 20 employees)					
20 - 199 employees	0.0996#	0.0994#	0.0983	0.1019#	0.0991#
200 - 1999 employees	0.1917#	0.1923#	0.1822#	0.1817#	0.1730#
2000 and more employees	0.2466#	0.2454#	0.2311#	0.2441#	0.2311#
Location[c](Ref. south)					
North	-0.0102	-0.0116	-0.0185	-0.0110	-0.0186
Center	-0.0335*	0.2868*	-0.0315	-0.0356*	-0.0306
Interactions					
Mainly German*Medium skilled	0.0177	0.0106	0.0131	0.0016	0.0050
Mainly German*High skilled	0.3326#	0.2868**	0.1706	0.0331	0.0106
Partly German*Medium skilled	0.0750**	0.0789**	0.0965**	0.0737**	0.0832**
Partly German*High skilled	0.1931**	0.1603**	0.1524*	0.1216	0.1047
Linear combinations Medium-skilled					
Mainly German	0.0709**	0.0855	0.1116*	0.0937	0.0905
Partly German	0.0525*	0.0617*	0.0854*	0.0653	0.0716
Linear combinations High-skilled					
Mainly German	0.3857#	0.3618#	0.2690**	0.1253	0.0961
Partly German	0.1706**	0.1431*	0.1412*	0.1131	0.0932
Selection terms[d]					
λ_1 (Participation)	—	0.2105	0.2762	0.2823	0.3023
λ_2 (Employment)	—	-0.2263	-0.2704*	-0.2498*	-0.2822*
M1	—	—	0.0374*	—	—
M2	—	—	-0.0678	—	—
M3	—	—	-0.2227	—	-0.0162
M4	—	—	0.0763	—	0.0497
M5	—	—	-0.0845	—	—
M6	—	—	0.3124	—	—
M7	—	—	-0.1343*	—	0.0130
M8	—	—	-0.0509	—	-0.0742*
Year dummies	YES	YES	YES	YES	YES
Observations	2,021	2,021	2,021	1,865	1,865
R^2	.2894	.2915	.3264	.2789	.2908

Significance is indicated as follows:# denoting the 0,1%, *** denoting the 1.0%, ** the 5% and * the 10% level.

[a] Model 1 - no selection control; Model 2 - controlling for selection into participation and employment; Model 3 - additional controls for selection into economic sector and occupation; Model 4 - excluding white-collar workers, controlling for selection into participation and employment; Model 5 - excluding white-collar workers, additional controls for selection into economic sector and occupation.

[b] Medium skilled are people with professional training, high skilled are people with advanced technical college or university degree.

[c] *North* contains the Federal States of Schleswig-Holstein, Hamburg, Lower-Saxony, Bremen, and Berlin. *Center* are the Federal States North Rhine-Westphalia, Rhineland-Palatinate, and Saarland. *South* comprises Hessen, Bavaria, and Baden-Wuerttemberg.

[d] λ_1 and λ_2 are calculated using estimates for the participation and employment given in Table B.1.

Tab. B.3: Earnings equation and self-assessed language proficiency[a]

	Model 1	Model 2	Model 3	Model 4	Model 5
Constant	1.6306$^{\#}$	1.6594$^{\#}$	1.5863$^{\#}$	1.6172$^{\#}$	1.6215$^{\#}$
Age	0.0283***	0.0274**	0.0288**	0.0286**	0.0276**
Age(squared)	-0.0003**	-0.0003**	-0.0003**	-0.0003**	-0.0003**
Time of residence	0.0073	0.0081	0.0109	0.0091**	0.0124***
Time of residence(squared)	-0.0001	-0.0001	-0.0002**	-0.0002**	-0.0002***
Level of Education^{b}(Ref. low skilled)					
Medium skilled	0.0057	0.0137	0.0110	0.0134	0.0126
High skilled	-0.1412**	-0.1255	-0.0846	-0.0473	-0.0413
Language usage					
Language Proficiency	0.0297**	0.0315	0.0441*	0.0346*	0.0313
Sector (Ref. industry & manufacturing)					
Transportation	-0.1286$^{\#}$	-0.1278$^{\#}$	-0.0601*	-0.1415$^{\#}$	-0.0567*
Construction	-0.0110	-0.0097	0.0683**	-0.0092	0.0743**
Trading services	-0.1230$^{\#}$	-0.1219$^{\#}$	-0.0744*	-0.1511$^{\#}$	-0.0691*
Social services and health	-0.0746*	-0.0756*	-0.0411	-0.1514	-0.0299
Firm size (Ref. < 20 employees)					
20 - 199 employees	0.1038$^{\#}$	0.1041$^{\#}$	0.1024$^{\#}$	0.1057$^{\#}$	0.1028$^{\#}$
200 - 1999 employees	0.1939$^{\#}$	0.1946$^{\#}$	0.1847$^{\#}$	0.1862$^{\#}$	0.1767$^{\#}$
2000 and more employees	0.2460$^{\#}$	0.2453$^{\#}$	0.2330$^{\#}$	0.2473$^{\#}$	0.2346$^{\#}$
Location^{c}(Ref. south)					
North	-0.0081	-0.0099	-0.0168	-0.0103	-0.0177
Center	-0.0348*	-0.0367*	-0.0323*	-0.0362*	-0.0325*
Interactions					
Language Proficiency*Medium skilled	0.0131	0.0088	0.0166	0.0088	0.0093
Language Proficiency*High skilled	0.1401$^{\#}$	0.1276$^{\#}$	0.1167*	0.0205	0.0150
Linear combinations Medium-skilled					
Language Proficiency	0.0428**	0.0403	0.0978***	0.0455	0.0406
Linear combinations High-skilled					
Language Proficiency	0.1699$^{\#}$	0.1591$^{\#}$	0.0608*	0.0572	0.0463
Selection terms^{d}					
λ_1 (Participation)	—	0.0959	0.1569	0.1578	0.1644
λ_2 (Employment)	—	-0.1896	-0.2141	-0.2005	-0.2304
M1	—	—	-0.1066	—	—
M2	—	—	-0.0649	—	—
M3	—	—	-0.2861	—	0.0041
M4	—	—	0.0602	—	0.0022
M5	—	—	0.1429	—	—
M6	—	—	-0.0992	—	—
M7	—	—	-0.1931*	—	-0.0010
M8	—	—	-0.0619	—	-0.0899*
Year dummies	YES	YES	YES	YES	YES
Observations	2,021	2,021	2,021	1,865	1,865
R^2	.2906	.2921	.3231	.2728	.2855

Significance is indicated as follows: $^{\#}$ denoting the 0,1%, *** denoting the 1.0%, ** the 5% and * the 10% level.

[a] Model 1 - no selection control; Model 2 - controlling for selection into participation and employment; Model 3 - additional controls for selection into economic sector and occupation; Model 4 - excluding white-collar workers, controlling for selection into participation and employment; Model 5 - excluding white-collar workers, additional controls for selection into economic sector and occupation.

[b] Medium skilled are people with professional training, high skilled are people with advanced technical college or university degree.

[c] *North* contains the Federal States of Schleswig-Holstein, Hamburg, Lower-Saxony, Bremen, and Berlin. *Center* are the Federal States North Rhine-Westphalia, Rhineland-Palatinate, and Saarland. *South* comprises Hessen, Bavaria, and Baden-Wuerttemberg.

[d] λ_1 and λ_2 are calculated using estimates of the participation and employment equation given in Table B.1.

Tab. B.4: Selection model: results for joint estimation of occupation type and economic sector[a]

	language usage	language proficiency	language usage	language proficiency
	Type of occupation		Economic Sector	
Constant	-2.8096	-2.2869	-1.6932**	-1.6898**
Age	-0.0045	-0.0378	0.0663	0.0632
Age(squared)	0.0002	0.0004	-0.0008	-0.0008
Time of residence	0.1095**	0.0387	0.0380*	0.0553**
Time of residence(squared)	-0.0012	-0.0003*	-0.0008**	-0.0010**
Level of education (Ref. low skilled)[b]				
Medium skilled	0.1578	0.1192	-0.2533	-0.1560*
High skilled	0.9728	-0.1912**	-0.0035	0.0113*
Language usage (Ref. mainly mother tongue)[c]				
Mainly German	0.7715**	—	-0.2693	—
Partly German	0.2800	—	-0.1739	—
Language proficiency	—	0.3083*	—	-0.1327*
Interactions				
Mainly German*medium skilled	-0.0448	—	0.2150	—
Mainly German*high skilled	1.1354*	—	-0.5889	—
Partly German*medium skilled	0.2610	—	0.3122	—
Partly German*high skilled	0.5314	—	-0.2317	—
Language proficiency*medium skilled	—	0.0400	—	0.0871
Language proficiency*high skilled	—	0.8473***	—	-0.1633
Linear combinations medium-skilled				
Mainly German	0.7266**	—	-0.0543**	—
Partly German	0.5411**	—	0.1383	—
Language proficiency	—	0.3483**	—	-0.0727
Linear combinations high-skilled				
Mainly German	1.9069#	—	-0.8583*	—
Partly German	0.8115	—	-0.4057	—
Language proficiency	—	1.1557#	—	-0.1598*
Exclusion restriction (Ref. north)[c]				
Center	0.5218**	0.5174**		
South	0.2035	0.2008		
ρ	-0.4470#	-0.5521#		
Observations	2,021		2,021	

All estimates are displayed in terms of coefficients. Significance is indicated as follows: # denoting the 0.1%, *** the 1%, ** the 5% and * the 10% level.

[a] medium-skilled are people with professional training, high-skilled are people with advanced technical college or university degree.

[b] Language proficiency is estimated by the ordered probit regression. See section 4.4 for details.

[c] *North* contains the Federal States of Schleswig-Holstein, Hamburg, Lower-Saxony, Bremen, and Berlin. *Center* are the Federal States North Rhine-Westphalia, Rhineland-Palatinate, and Saarland. *South* comprises Hessen, Bavaria, and Baden-Wuerttemberg.