Institut für Volkswirtschaftslehre
Universität Augsburg

Volkswirtschaftliche Diskussionsreihe

How Different are Wages from Wage Potentials?

Analyzing the earnings disadvantage of immigrants in Germany

Günter Lang

Beitrag Nr. 256, Januar 2004

How Different are Wages from Wage Potentials?

Analyzing the earnings disadvantage of immigrants in Germany

Günter Lang
Department of Economics
University of Augsburg
D-86135 Augsburg
GERMANY

guenter.lang@wiwi.uni-augsburg.de

Abstract

Immigrants in Germany display a poor earnings performance relative to natives. Arguing that human capital endowments identify earnings potentials rather than actual earnings, this paper estimates a stochastic earnings frontier and searches for systematic differences between natives and migrants in terms of distance to the frontier. GSOEP-Data of the year 2000 is used for estimation. The empirical results clearly support the frontier assumption, but – surprisingly – find natives and immigrants at about the same distance to the frontier. Assuming a half-normal distribution of the wage-inefficiency term, both groups transformed on average a modest 81% share of their potential income into market earnings. Due to the similar positions of natives and immigrants relative to the frontier, the wage discrimination hypothesis is rejected. Actually, human capital differentials are clearly the most important source for wage inequality. The earnings frontiers of immigrants from Eastern Europe as well as from Turkey are steeper than the respective frontier of natives, which supports the assimilation hypothesis. No assimilation is found for migrants from the European Union and from the former Yugoslavia.

JEL classification: J31, J61, J71

Keywords: Stochastic Wage Frontier, Inefficiency, Immigration, Assimilation, Discrimination, Human Capital Approach

1 Introduction

Worldwide immigrant flows have dramatically increased with the beginning of the nineteen-eighties. In Germany, for example, more than 3 million persons from Eastern Europe have arrived since 1985. Although this influx of ethnic German immigrants has recently reduced from its peak at the end of the eighties, still more than 100,000 persons per year are currently arriving. Being the descendants of former emigrants, the majority has legitimate claim for German citizenship and will settle permanently. Looking back into the past, a strong migration flow could also be observed between 1960 and 1975, when more than 3 million foreign guest workers from the southern regions of the EU (Spain, Italy, Greece), the former Yugoslavia and Turkey moved to West Germany. Asylum seekers are not considered in these figures. For details see *Loeffelholz and Köpp (1998).*

The focus of this paper is on the substantial wage inequality between the different ethnic groups of immigrants and the native population. In 2000, for example, monthly gross earnings of male immigrants from Turkey, the former Yugoslavia and Eastern Europe are about 20% below the level of natives. On an hourly basis, the differential narrows somewhat to 15% for Turkish migrants and to 17% for migrants from Yugoslavia and Eastern Europe. In contrast, immigrants from EU countries turned out to be a relative successful group on the German labor market: Their wage disadvantage is 10% on a monthly and 7% on an hourly basis. Details are presented in *Table A-1* in the appendix.

Within the framework of the human capital approach, the observed wage inequality is interpreted as inequality of productivity relevant characteristics, i.e. different stocks of human capital. However, incomplete information about prospective wages (*Polachek and Robst, 1998*), discrimination (*Oaxaca, 1973*) or other market imperfections like limited regional mobility of the workers may also play an important role for the determination of individual earnings. As a result, some of the employees earn less than their human capital stock allows for.

This paper is following a small but growing literature using frontier approaches to determine the earnings function. Important contributions are from *Daneshvary et al. (1992), Herzog et al (1985), Hofler and Polachek (1985), Hunt-McCool and Warren (1993), Polachek and Yoon (1996)* and *Polachek and Robst (1998)*. The idea of estimating a frontier relationship was originally introduced for the analysis of firms, where the quality of the management is probably not homogenous but differs among firms (*Farrell, 1957*). As a result of this heterogeneity, not all firms are able to extract the maximum possible output from a given bundle of inputs. The output loss from not being on the production frontier is called inefficiency.

Transferred to the labor market, the earnings frontier describes the highest potential income associated with a given stock of human capital. All individuals are located either on or below this frontier. Workers who translate their potential wages one-by-one into market wages are enjoying a fully efficient position. In contrast, individuals who earn less than their potential wage are suffering from some kind of "wage inefficiency". Wage inequality is therefore not restricted to be a result of differentials in productivity relevant characteristics, but may also be a result of different positions relative to the frontier. A worker with a low wage potential may easily overtake a "high potential" due to a better transformation into market earnings. Other than the *Oaxaca (1973)*-decomposition with separate estimations for groups, all observations are jointly used to define the frontier. Of course, the distance to the frontier is then individually different. Actually, frontier approaches allow for a ranking of individuals by their relative distance to the frontier.

A two-step procedure is used in this paper. In a first step, the wage function is estimated on the basis of the stochastic frontier approach (see *Fried et al., 1993*, for an overview). From the wage frontier maximum earnings of all individuals and therefore the heterogeneity of the human capital endowment can be determined. Possible assimilation processes (*Chiswick, 1978*) are considered for in the wage function. The results allow for a comparison of the wage potentials of natives relative to the wage potentials of migrants. In a second step, the degree of wage inefficiency is calculated for each individual. A higher degree of wage inefficiency is expected for migrants, which could also explain at least a part of the wage gap. Discrimination, for example, would drive an additional wedge between potential income and observed income for a disadvantaged ethnic group (*Robinson and Wunnava, 1989*). This hypothesis is tested for by a second-stage regression, where wage efficiency levels are explained by a set of variables including ethnicity dummies catching up possible discrimination.

The structure of the paper is as follows: Chapter 2 describes the specification of the model and the estimation method. In chapter 3 the dataset and the estimation results on the earnings frontier are presented. Chapter 3 is rounded up by drawing conclusions about the explanatory power of the human capital approach for the native-immigrant wage gap. Chapter 4 estimates the individual deviations from the frontier, and a regression tries to explain the different wage-inefficiency levels. Finally, chapter 5 sums up.

2 Specification

Following the prevailing literature in labor economics, a semilogarithmic human capital production function of the extended *Mincer (1974)* type

$$\ln E_i = \alpha + \beta' X_i + \varepsilon_i \ , \qquad (1)$$

is assumed. E_i denotes hourly earnings of individual i, X_i is a vector of classical ('Mincerian') human capital variables, α as well as the β-vector represent unknown parameters, and ε is the error term. Equation (1) states that wages are determined by the endowment with human capital *X*. In the following empirical application a bundle of proxy variables like years of schooling are used to measure *X*.

The focus of this paper is on the earnings differentials between natives and immigrants. As for immigrants, it is important to consider assimilation processes, which may also play an important role for wage levels. The assimilation hypothesis assumes a depreciation of the human capital stock at the time of entry, which can be compensated by adjusting to the new working environment during the next periods (*Chiswick, 1978).* Similar to *Chiswick*, assimilation is introduced by expanding (1) into

$$\ln E_i = \alpha + \beta' X_i + \sum_{g=1}^{4} \gamma_g \left(YSM_i \cdot D_i^g \right) + \sum_{g=1}^{4} \eta_g D_i^g + \varepsilon_i \quad . \qquad (2)$$

In (2), the dummy variable D_i^g equals one if individual i belongs to immigrant group g, zero otherwise. Actually, four categories of immigrants are differentiated in this paper: Immigrants from Eastern Europe, the European Union, Turkey, and the former Yugoslavia. For all immigrants, the years since migration to Germany are measured by the variable YSM_i. In equation (2), the parameters η measures the percentage loss of human capital at the time of entry to Germany, differentiated by the source countries of immigrants. This reflects the lack of Germany-specific skills. The parameter γ_g measures the rate at which immigrants from source country g do indeed catch up to the native counterparts. The assimilation hypothesis is supported for group g if γ_g is significantly positive and γ_g is significantly negative.

In equation (2) earnings are related to human capital plus a stochastic error term. However, as described in the introduction, individuals may not always be able to transform their human capital endowment into earned income. The theoretical earnings function is an ideal - the maximum of earnings one can realize. From an econometric point of view, the residual variable ε_i has a negative mean value. *Aigner, Lovell and Schmidt (1977)* proposed a formulation within which the error term ε_i is split up into tow parts: A non-negative "inefficiency" term u_i generated by a stochastic process, and a white noise variable $\upsilon_i \sim N(0, \sigma_\upsilon^2)$. As a result, the earnings function (2) can be written as

$$\ln E_i = \underbrace{\underbrace{\underbrace{\alpha + \beta' X_i + \sum_{g=1}^{4} \gamma_g \left(YSM_i \cdot D_i^g \right) + \sum_{g=1}^{4} \eta_g D_i^g}_{\text{wage potential (= wage frontier)}} + \upsilon_i}_{\text{stochastic wage frontier}} - u_i}_{\text{observed wage}} \quad . \qquad (3)$$

Figure 1:
Earning Pattern of Natives and Immigrants with Assimilation

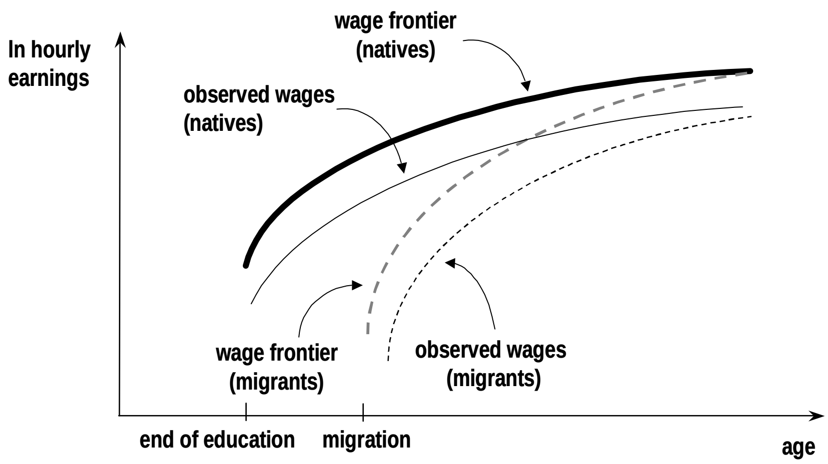


Figure 1 is graphically illustrating the frontier earnings function and observed earnings for a representative native and for a representative migrant. The deterministic part of the earnings function $\alpha + \beta' X + \sum \gamma (YSM \cdot D) + \sum \eta D$ defines an envelope associated with specific amounts of human capital. Both earnings frontiers are concave with respect to the age[1] variable. As in contrast to *Robinson and Wunnava (1989),* wage inefficiency is allowed to occur for natives as well as for migrants. This increases the plausibility of the approach, since some reasons for a shortfall from the potential earnings curve may also be relevant for natives (e.g. regional or occupational immobility, incomplete information). However, the distance to the frontier may be wider for an immigrant than for a native. Finally, in *Figure 1* an assimilation process of immigrants is assumed. The result is an earnings frontier, which is steeper than the respective frontier of natives.

To estimate the parameters of the underlying function, the stochastic distribution of the inefficiency term u_i has to be specified. The most popular assumption is a half normal distribution for u_i, introduced by *Aigner et al. (1977).* The log-likelihood function for the half-normally distribution is

1 Estimations in the empirical part of the paper are based upon actual labor market experience.

$$\ln L(\alpha, \beta, \gamma, \eta, \sigma^2, \lambda) = \sum_{i=1}^{N} \left[-\ln \sigma \quad + \frac{1}{2} \ln \frac{2}{\pi} - \frac{1}{2} \frac{\varepsilon_i^2}{\sigma^2} + \ln \left(1 - \Phi \left(\frac{\varepsilon_i \lambda}{\sigma} \right) \right) \right], \tag{4}$$

where $\Phi(\cdot)$ is the standard normal distribution function, $\sigma^2 = \sigma_\upsilon^2 + \sigma_u^2$, $\lambda = \sigma_u / \sigma_\upsilon$, and $\varepsilon_i = -u_i + \upsilon_i$.

To control for the robustness of the results, estimations are also run for an exponentially distributed inefficiency term (see *Meeusen and van den Broeck, 1977).* The corresponding log-likelihood function is defined as

$$\ln L(\alpha, \beta, \gamma, \eta, \sigma_\upsilon^2, \theta) = \sum_{i=1}^{N} \left[\ln \theta + \frac{1}{2} \theta^2 \sigma_\upsilon^2 + \theta \varepsilon_i + \ln \Phi \left(\frac{-\varepsilon_i}{\sigma_\upsilon} - \theta \sigma_\upsilon \right) \right]. \tag{5}$$

ε_i is defined as before. Numerical maximization of (4) and (5) produces the compounded residual ε_i, which is the sum over the random component υ_i and the negative of the inefficiency term u_i. Following *Jondrow et al. (1982),* the conditional expectation value of u_i given ε_i can be calculated as

$$E(u_i | \varepsilon_i) = \left(\frac{\phi[\varepsilon_i \lambda / \sigma]}{1 - \Phi[\varepsilon_i \lambda / \sigma]} - \frac{\varepsilon_i \lambda}{\sigma} \right) \frac{\sigma \lambda}{1 + \lambda^2} \tag{6}$$

for the half-normal assumption. In equation (6), $\phi(\cdot)$ and $\Phi(\cdot)$ represent the density and the distribution of the standard normal distribution. All information necessary for the computation of E(u_i) is available from the parameters of the maximized likelihood-function.

Similarly, for the exponential case u_i is computed as

$$E(u_i | \varepsilon_i) = (-\varepsilon_i - \sigma_\upsilon^2 \theta) + \left(\frac{\sigma_\upsilon \phi[(\varepsilon_i + \theta \sigma_\upsilon^2) / \sigma_\upsilon]}{\Phi[(-\varepsilon_i - \theta \sigma_\upsilon^2) / \sigma_\upsilon]} \right). \tag{7}$$

The u_i-values are transformed into the efficiency ratio EFF_i, measuring the gap between the wage frontier and the stochastic wage frontier as percentage of the wage frontier actually realized by person *i*. Due to the semilogarithmic form of the earnings equation (1), EFF_i can be calculated as

$$EFF_i = \frac{\exp\left(\alpha + \beta' X_i + \sum_{g=1}^{4} \gamma_g \left(YSM_g \cdot D_i^g\right) + \sum_{g=1}^{4} \eta_g D_i^g - u_i\right)}{\exp\left(\alpha + \beta' X_i + \sum_{g=1}^{4} \gamma_g \left(YSM_g \cdot D_i^g\right) + \sum_{g=1}^{4} \eta_g D_i^g\right)} = \exp(-u_i) \cdot \qquad (8)$$

EFF_i is restricted to the interval]0,1], with the upper boundary representing a worker who transforms his human capital endowment perfectly into market income. A value of less than one - e.g. 0.90 - indicates that the observed employee is realizing 90% of his frontier wage. With other words, this person could increase his earnings by $1/0.90 - 1 = 11.1\%$ without increasing his human capital endowment.

In a second step, the estimated EFF_i–values are explained by some household characteristics of the employee, supplemented by information about the employer and a source-country dummy for immigrants:

$$EFF_i = \begin{array}{lll} \delta_0 & & \langle \text{Interpretation:} \rangle \\ + \delta_1 * MARRIAGE_i & + \delta_2 * CHILD_i & \langle \rightarrow \text{spatial immobility} \rangle \\ + \delta_3 * SMALL_i & + \delta_4 * MEDIUM_i & \langle \rightarrow \text{market power of employer} \rangle \\ \sum_{g=1}^{4} \delta_{4+g} D_i^g & & \langle \rightarrow \text{discriminationan against immigrants} \rangle \\ + \varepsilon_i^1 & & \end{array} \qquad (9)$$

MARRIAGE gives the marital status, *CHILD* the number of children living in the household of individual *i*. To be married and to educate children are important determinants of spatial immobility and should therefore have impacts on the degree of wage efficiency. The sign of these parameters is expected to be negative. The dummy variables *SMALL* and *MEDIUM* take the value one if the corresponding firm employs less than 20 respectively between 20 and 2000 workers. To be employed at a large firm with more than 2000 employees is the reference scenario. When assuming that unions are less powerful in small and medium sized firms, δ_3 and δ_4 should also have a negative sign.

As for migrants, their actual earnings levels may also be influenced by discrimination. Discrimination is captured by a set of dummy variables D_i^g equal to one if individual *i* is a migrant from region *g*. If a specific ethnic minority is systematically suffering from lower EFF_i–

values, then the corresponding D_i^g parameter(s) will take a negative sign. ε_i^1 represents a regular error term.

3 Human Capital Approach: Empirical Results

Empirical basis of this paper is the German Socio-Economic Panel (GSOEP), where a representative number of natives and immigrants are interviewed. Starting in 1984 with about 6000 household interviews, the panel was several times extended to compensate for panel attrition, the German unification, and the resurgence of migration in Germany. In 1994 and 1995 the so-called “immigration waves” were introduced, broadening the sample to Ethnic German migrants from Eastern Europe. This study uses the recent 2000-wave where 6000 additional households were interviewed, nearly doubling the sample size against 1999 to about 13,000 households with 34,000 persons. Many households with immigrants are among this recent extension. However, this is also the reason that the estimation of cohort effects as in *Borjas (1985, 1994)* is problematic. The panel length especially for immigrants from Eastern Europe is too short. Significant panel attrition, which enforces continuous “refreshments” of the interviewed households, is further worsening the problem. Detailed information about the GSOEP and the sample sizes is provided by *Pannenberg (2000).*

For this study, the baseline sample – males between 18 and 64 years of age living or working in West Germany – consists from about 9,000 observations. In a second step, all persons who were unemployed in the year 2000, self-employed or not active on the labor market are excluded (-2,000 observations). More than 1,000 have to be deleted due to missing information about wages, measured as gross earnings per actual labor hour. Finally, to reduce distortions from extreme outliers or data errors, the research population is confined to a sub sample of employees working at least 15 hours per week (-1,300 observations) and hourly wages in the range from 2.5 EUR to 100 EUR per hour (-200 observations). After these corrections, the final sample consists from 4456 employees, of which 3852 are natives.

Their country of birth differentiates natives and immigrants, with immigrants neither being born in West Germany nor in East Germany. Four ethnic groups are defined: Immigrants from Eastern Europe, the European Union (defined as EU-15), from Turkey, and from the former Yugoslavia. Eastern Europe is the main source region of ethnic Germans ("Aussiedler"). Turkey is the largest source country of foreign guest workers, with immigration starting at the beginnings of the sixties. The EU member countries as a whole represent the second largest source region. Most of the EU immigrants are from the southern member countries (Italy, Greece, Spain). Although the influx from Yugoslavia started with some lag (recruitment offices were opened at the end of the sixties), the share of non-German immigrants is third after Turkey and the EU (see *Loeffelholz and Köpp, 1998*, for more details). Migrants from other regions of the world – actually a small and heterogeneous sub sample of the GSOEP – are deleted. From the total of 4456 persons, 604 persons are immigrants and 3852 are natives (see also *Table A-1*).

In a first step of the empirical analysis, the income frontier (2) specified as

$$\begin{aligned}\ln E_i = \; & \alpha \\ & + \beta_1 \cdot SCHOOL_i \quad + \beta_2 \cdot OCCUPATION_i \quad + \beta_3 \cdot COLLEGE_i \qquad \langle \rightarrow \text{education} \rangle \\ & + \beta_4 \cdot EXPERIENCE_i \quad + \beta_5 \cdot EXPERIENCE_i^2 \quad + \beta_6 \cdot UNEMPLOYED_i \qquad \langle \rightarrow \text{career variables} \rangle \\ & + \beta_7 \cdot STATUS_i \quad + \beta_8 \cdot HOURS_i \qquad \langle \rightarrow \text{job variables} \rangle \\ & + \beta_9 \cdot LANGUAGE_i \quad + \beta_{10} \cdot RETURN_i \qquad \langle \rightarrow \text{other characteristics} \rangle \\ & + \sum_{g=1}^{4} \gamma_g YSM_i \cdot D_i^g \quad + \sum_{g=1}^{4} \eta_g D_i^g \qquad \langle \rightarrow \text{assimilation} \rangle \\ & + \varepsilon_i \end{aligned} \tag{10}$$

is determined. *SCHOOL* measures the years of schooling in Germany or abroad by five categories (seven, nine, ten, twelve or thirteen years). Not the actual years of schooling, but the school degrees decide about the category of each person. Immigrants finishing their education in the home country are assumed to suffer from a discount on their school education. The years of schooling assigned to are seven, nine or – at the maximum - ten years (see the GSOEP handbook on the computation of the $PGEN variables). *OCCUPATION* is a dummy variable identifying occupational training; *COLLEGE* is a dummy variable identifying a successful college degree. *EXPERIENCE* measures the cumulated labor experience either in Germany or abroad, *UNEMPLOYED* the cumulated times of unemployment. The *STATUS* dummy takes one if the person is a blue-collar worker. *HOURS* of working are entering the earnings function due to the possibility of a declining per-hour-productivity. The *LANGUAGE* dummy identifies problems with the German language; the *RETURN* dummy differentiates between permanent migrants and temporary migrants. As usual, *YSM* measures the years since migration to Germany. A description of the variables, their mean values and their standard deviations is provided in *Table A-1* in the appendix.

The functional specification of equation (10) allows for concavity in the experience variable ($\beta_5 < 0$). The parameters β_9, β_{10}, γ and η control for the specific situation of immigrants, with four ethnicity dummies D_i^g capturing possible differences in the assimilation process among ethnic groups.

Estimation results are presented in *Table A-2* in the appendix. The first column listed in this table contains the OLS estimates for the parameters of the wage function, assuming that there is only white noise in the wage function (i.e. $u_i = 0$). From OLS, the overall goodness-of-fit is $\overline{R}^2 = 0.30$, which is quite satisfactory for an earnings function. These OLS estimates are used as starting values in the iterative process to obtain the ML estimates[2]. Both the half-normal as well as the exponential model are estimated in two versions: One based on the native-only data and without the assimilation variables, which is a control estimation, and one based on the full specification and the whole sample, i.e. natives and immigrants. The ML-iteration computations turned out to be stable and converging. Most of the estimated parameters are found to be statistically significant; all parameters with the exception of the insignificant *LANGUAGE* dummy show the expected sign. A Likelihood-ratio test that all slope parameters are zero can clearly be rejected.

---

2 All calculations were run by GAUSS.

The empirical results clearly support the use of a frontier approach for the determination of wage functions. As the *t*-statistics suggests, the λ- as well as the θ-parameter are significantly different from zero, rejecting the hypothesis that only random error occur. The advantage of the stochastic frontier approach can also be calculated from variance decomposition: Following *Greene (1993)*, the contribution of the variance of *u* to the variance of the composite error term ε is given by $[(\pi/2)-1]\sigma_u^2/\{[(\pi/2)-1]\sigma_u^2+\sigma_v^2\}$ for the half-normal model. For the exponential model, the counterpart is $[1/\theta^2]/[1/\theta^2+\sigma_v^2]$. Inserting the estimates from *Table A-2*, approximately 39% respectively 26% of the variance of ε can be assigned to earnings inefficiency. This result is line with the findings for the US, where *Hunt-McCool and Warren (1993)* estimate a 27% share. The advantage of a frontier approach is therefore not just marginally; using a stochastic frontier enhances the explanatory power of the wage function significantly.

Interestingly, the slope parameters from the full models and from the native-only estimations are very similar and therefore confirm the robustness of the result. The most notably exception is a small increase of λ and θ, indicating a somewhat more important role of inefficiency in the full sample. A likelihood ratio test on the joint-insignificance of the migration-specific variables (i.e. $\beta_9=\beta_{10}=0, \gamma=0, \eta=0$) was conducted for the full sample. The test statistics take the values 23.3 and 23.6 for the half-normal and the exponential model, respectively, therefore supporting the full model[3]. Similarly, the differences between the exponential and the half-normal model are also small. Due to this similarity, the following analysis is concentrating on the full model estimated by the half-normal stochastic frontier.

From an economic point of view, the role of education for generating income is clearly underlined. The calculations suggest that every year of schooling is shifting the wage frontier upwards by roughly 1.7%. An academic degree turned out to be a key variable for economic suc-

---

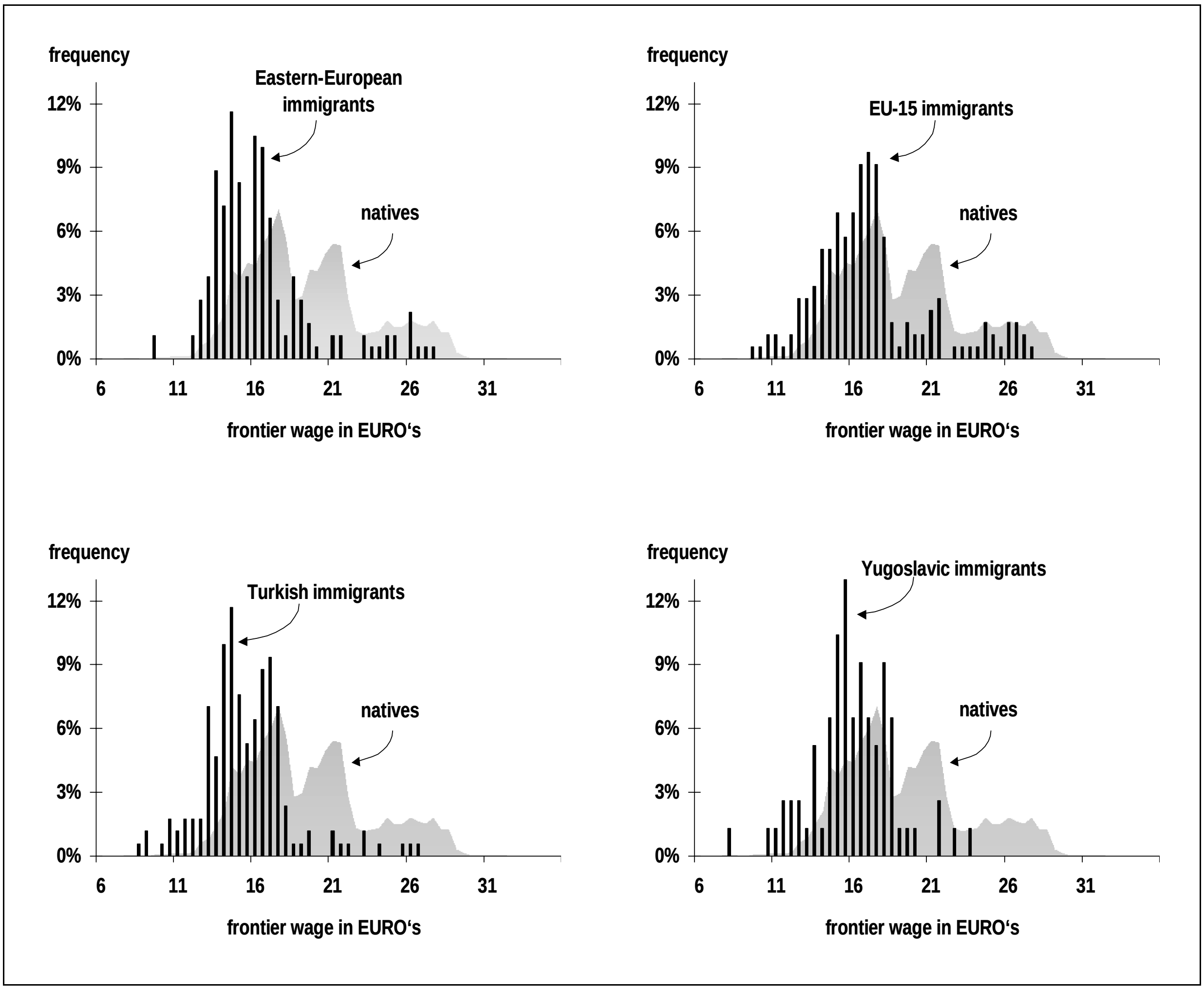
3 The critical value of the Chi-Squared distribution with 10 degrees of freedom is 18.3 for a 95% probability.

cess: The corresponding wage increase is estimated to be 24%, far beyond the role of the insignificant occupation variable. Every year of unemployment is decreasing the earnings potential by 4.0%, being a blue-collar worker by about 19%. Aside from schooling, practical experience turned out to be the most important income determinant. Ceteris paribus, 25 years of labor market experience are driving the earnings potential upwards by somewhat more than 20% relative to a beginner.

The estimates also support the assimilation hypothesis, but this support is limited to immigrants from Eastern Europe and Turkey. The corresponding γ as well as the η parameters are significantly different from zero and show the expected sign. Both groups are estimated to suffer from a negative income shock at the time of entry at about –16%. Disappointingly, the following assimilation process shifts the wage frontier by just 0.8% per year upwards and is therefore very slow. About 19 years are necessary to compensate for the disadvantage from migrating. The assimilation hypothesis is not supported for foreign guest workers from the European Union and from Yugoslavia, however.

As a main result of the frontier earnings function, lower earnings of migrants are in line with the human capital approach. Modest human capital endowment of ethnic Germans and of foreign guest workers reduce their income potential by a value between 9% and 17% relative to natives (see *Table A-3*). Employees from Turkey and from Yugoslavia are at the lower end of this interval, with the frontier wage differential relative to the native frontier turning out to be a nearly perfect forecast for the actual wage differential. As for immigrants from the European Union, the frontier wage differential is estimated somewhat more pessimistic than the actual wage differential. *Table A-3* provides details on the sources of the poor human capital endowment. The status as a blue-collar worker, relatively long times of unemployment and a missing college degree are the main reasons for the shortfall of immigrants relative to native employees.

Figure 2:
The distribution of frontier wages



Calculations for the half-normal model.
The percentage-frequency describes the share of persons in the relevant category; each category width is 0.5€.

Figure 2 is graphically illustrating the results. The percentage frequency of frontier wages over the estimated bandwidth is plotted for immigrants relative to natives. From the distribution of the frequency masses, one can immediately see the earnings disadvantage of immigrants. Especially the right half of the tail, representing high-skilled employees with a college degree and labor market experience, is barely covered by immigrants. What also can be seen is a much more heterogeneous wage distribution among natives than among migrants.

4 Earnings Efficiency

The previous part analyzed the relationship between human capital stocks and earnings ideals. In this chapter, EFF_i -values measuring realized earnings as percentage of potential earnings are calculated and analyzed. *Table 1* shows the main results of the calculations. The figures clearly underline the relevance of the frontier concept: At the average, employees are realizing a modest 80.5% share of their potential earnings as actual earnings (half-normal model). A representative employee could therefore increase the wage level by 24% without any investments into his human capital endowment.

Table 1:
Wage efficiency by group

	Half-normal Distribution		Exponential Distribution	
	mean value of EFF_i	standard deviation	mean value of EFF_i	standard deviation
whole sample	0.805	0.072	0.857	0.070
natives	0.804	0.074	0.856	0.073
immigrants from				
- Eastern Europe	0.810	0.058	0.864	0.052
- European Union	0.811	0.057	0.865	0.053
- Turkey	0.811	0.053	0.866	0.049
- former Yugoslavia	0.815	0.042	0.871	0.033

The difference between the half-normal and the exponential model is mainly a difference in the level of EFF_i. With the exponential model, the gap between actual earnings and potential earnings is smaller (wage discount of 14.3% versus 19.5% with the half-normal model). However, the ranking of the individuals in terms of EFF_i is nearly identical. Actually, the ranking coefficient between the exponential and the half-normal model is 0.99, supporting the robustness of the results.

Surprisingly, the absolute differences in EFF_i between natives and migrants are in the range of 1% and therefore insignificant. Due to the assumption of more serious information problems of migrants, eventually coupled with discrimination effects, the expectation was to find a lower, more heterogeneous EFF_i -value. What was found instead is a less heterogeneous efficiency variable for immigrants and a somewhat better transformation of the wage potential into realized wages. The significantly lower standard deviations for immigrants confirm the results from the previous chapter, where natives were found to be much more heterogeneous than migrants.

Table 2:
Tobit estimation of wage efficiency equation

	EFF_i variable from			
	half-normal model	exponential model	half-normal model	exponential model
σ^2	0.0047*** (0.001)	0.0046*** (0.001)	0.0046*** (0.001)	0.0046*** (0.001)
CONST	0.8044*** (0.0025)	0.8558*** (0.0025)	0.8043*** (0.0025)	0.8556*** (0.0025)
MARRIED	0.0197*** (0.0024)	0.0173*** (0.0023)	0.0196*** (0.0024)	0.0169*** (0.0023)
CHILD	0.0051*** (0.0011)	0.0057*** (0.0011)	0.0051*** (0.0011)	0.0056*** (0.0011)
SMALL	-0.0440*** (0.0032)	-0.0395*** (0.0039)	-0.0443*** (0.0032)	-0.0399*** (0.0031)
MEDIUM	-0.0169*** (0.0024)	-0.0143*** (0.0024)	-0.0172*** (0.0024)	-0.0147*** (0.0024)
$D_{EASTERN\ EUROPE}$			0.0030 (0.0053)	0.0059 (0.0052)
D_{EU}			0.0040 (0.0053)	0.0067 (0.0052)
D_{TURKEY}			-0.0028 (0.0055)	0.0007 (0.0054)
$D_{YUGOSLAVIA}$			0.0124 (0.0079)	0.0167 (0.0078)
Log likelihood	5602.7	5679.6	5604.5	5683.1
LR-statistics: all slopes are 0	333.8***	292.0***	337.4***	299.0***
Observations	4456	4456	4456	4456

Explained variable is EFF_i, which is bounded by the value 1.
*, ** and *** represent a significance level of 90%, 95% and 99%, respectively.

Following equation (9), the relationship between earnings efficiencies EFF_i and some possible sources is estimated by Tobit.[4] This technique is more appropriate than OLS due to the

4 As an alternative to this two-step procedure, a single-stage procedure implementing the sources of wage inefficiency into the earnings function can be used. For panel data, *Battese and Coelli (1995)* have proposed an integrated model. See also *Wang and Schmidt (2002)* for a discussion of the single-stage procedure versus the two-step procedure.

bounded EFF_i-value. As the estimation results, presented in Table *2*, indicate, the size of the firm is the most important variable for EFF_i. Especially workers at small firms with less than 20 employees are obviously accepting wage offers significantly below their potential wages. This result is in line with other studies and may be due to a smaller influence of unions, for example. Interestingly, married men with children in their households are not found to be less, but more wage efficient. A low spatial mobility of these persons is obviously overcompensated by – e.g. – a higher intrinsic motivation.

No evidence was found on the important question of ethnic discrimination. As Table *2* shows, all ethnicity dummies are insignificant. Furthermore, the likelihood ratio test statistics that all ethnicity dummies are jointly equal to zero cannot be rejected. The test statistics are 3.6 for the half-normal and 7.0 for the exponential model, respectively, which are both below the critical value of 9.5 (95% probability). However, this result is not very surprising given the insignificant differences of the mean values of EFF_i in *Table 1*.

Finally, it should be noted that equation (9) can explain only a very small part of the observed variance in EFF_i. The residual variance of the computed models is about 0.0047, which is just slightly below the variance of the observed EFF_i-values (0.0051). Less than 10% of the variation in EFF is accounted for by variation in the regressors. That unsatisfactory result leaves a lot of room for speculation about the sources, which prevent a full conversion of the human capital stock into market earnings.

5 Conclusion

This paper is following recent research in labor market economics estimating the wage function on the basis of a stochastic frontier approach. Maximum potential earnings were calculated for natives as well as for migrants living in West Germany. In a second step, wage inefficiencies were estimated and put into relationship to a vector of explaining variables, including proxy variables for discrimination. As for the data set, the paper relies on the year-2000 wave of the GSOEP, which has much more observations than the previous waves.

The first main result is that lower earnings of immigrants are mainly a consequence of relatively low human capital endowments. Earnings differentials between the native population and immigrants are therefore sustainable. For example, frontier wages of Yugoslav immigrants are predicted to be 17% below the frontier wage of native Germans – an exact match to what we observe on the labor market. Somewhat lower is the explanatory power for ethnic Germans from Eastern Europe, where the earnings frontier predicts a wage gap of 11%, which is substantially better than the observed 17%-differential. Assimilation was found to play a significant role for immigrants for Eastern Europe, confirming the results of *Bauer and Zimmermann (1997)*. However, the catch-up process is very slow. Somewhat surprising is the support of the assimilation hypothesis for immigrants from Turkey, which is not confirmed by the existing literature (see e.g. *Licht and Steiner, 1994, Pischke, 1993, Schmidt, 1997*). As for immigrants from the former Yugoslavia and from EU member countries, assimilation processes could not be detected.

As the second main result, statistical tests clearly support the use of frontier techniques. Actually, up to 39% of the OLS residual variance is assigned to earnings inefficiency. At the average, employees are able to transform 81% (half-normal model) respectively 85% (exponential model) of their frontier earnings into market earnings. Most surprising, the differences between natives and migrants are insignificant. As a consequence, the empirical test on discrimination is clearly rejected for all ethnic groups of immigrants. Although the Tobit estimations show a significant role of firm size and the family status of the individuals, only a small part of the variation in the efficiency variable can be explained.

Which are the implications of these results? First, as a advise for policy making, incentives to increase the human capital stock of migrants are promising higher yields than the introduction of anti-discriminatory measures. Discrimination was not detected for immigrants in Germany. Second, the lack of assimilation capabilities of many guest workers is disappointing. An appropriate selection of future migrants by their willingness to assimilate seems therefore to be very urgent. And, third, this paper provides evidence in favor of the stochastic wage frontier as a useful instrument to analyze wage inequality. A full exploitation of the productivity potential as measured by the current human capital endowment would increase hourly earnings by 18% (exponential model) and 23% (half-normal model), respectively. Even a partial realization of the wasted earnings potentials would drive up earnings much faster than the macroeconomic trend growth of technical progress.

References

Aigner, D., Lovell, C., Schmidt, P. (1977), Formulation and Estimation of Stochastic Frontier Production Models, Journal of Econometrics 6, 21-37.

Bauer, T., Zimmermann, K.F. (1997), Unemployment and Wages of Ethnic Germans, Quarterly Review of Economics and Finance 37, 361-377.

Battese, G.E., Coelli, T.J. (1995), A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data, Empirical Economics 20, 325-332.

Blinder, A.S. (1974), Toward an Economic Theory of Income Distribution, Cambridge: MIT Press.

Borjas, G.J. (1985), Assimilation, Changes in Cohort Quality, and the Earnings of Immigrants, Journal of Labor Economics 4, 463-489.

Borjas, G.J. (1994), The Economics of Immigration, Journal of Economic Literature 32, 1667-1717.

Chiswick, B.R. (1978), Americanization and the Earnings of foreign-born Men, Journal of Political Economy 86, 897-921.

Daneshvary, N., Herzog, H.W. jr., Hofler, R.A., Schlottmann, A.M. (1992), Job Search and Immigrant Assimilation: An Earnings Frontier Approach, Review of Economics and Statistics 74, 482-492.

Farrell, M.J. (1957), The Measurement of Productive Efficiency, Journal of the Royal Statistical Society, A120, 253-281.

Fried, H.O., Lovell, C.A.K., Schmidt, S.S. (1993), The Measurement of Productive Efficiency: Techniques and Applications, Oxford: Oxford University Press.

Greene, W.H. (1993), The Econometric Approach to Efficiency Analysis, in*: Fried, H.O. et al. (eds.),* The Measurement of Productive Efficiency: Techniques and Applications, 68-119, Oxford: Oxford University Press.

Herzog, H.W. jr., Hofler, R.A., Schlottmann, A.M. (1985), Life on the Frontier: Migrant Information, Earnings and Past Mobility, Review of Economics and Statistics 67, 373-382.

Hofler, R.A., Polachek, S. (1985), A New Approach for Measuring Wage Ignorance in the Labor Market, Journal of Economics and Business 37, 267-276.

Hunt-McCool, J.C., Warren, R.S. jr. (1993), Earnings Frontiers and Labor Market Efficiency, in: *Fried, H.O. (ed.),* The Measurement of Productive Efficiency: Techniques and Applications, Oxford: Oxford University Press, 197-209.

Jondrow, J., Lovell, C.A.K., Materov, I.S., Schmidt, P. (1982), On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model, Journal of Econometrics 19, 233-238.

Licht, G., Steiner, V. (1994), Assimilation, Labour Market Experience, and Earnings Profiles of Temporary and Permanent Immigrant Workers in Germany, International Review of Applied Economics 8, 130-56.

Loeffelholz, H.D.v., Köpp, G. (1998), Ökonomische Auswirkungen der Zuwanderungen nach Deutschland, Duncker & Humblodt: Berlin.

Meeusen, W., van den Broeck, J. (1997), Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error Term, International Economic Review 8, 435-444.

Mincer, J. (1974), Schooling, Experience and Earnings, New York: Columbia University Press.

Oaxaca, R. (1973), Male-Female Wage Differentials in Urban Labor Markets, International Economic Review 14, pp. 693-709.

Pannenberg, M. (2000), Documentation of Panel Sizes and Panel Attrition in the German Socio Economic Panel (GSOEP). Deutsches Institut für Wirtschaftsforschung, discussion paper 196.

Pischke, J.-S. (1993), Assimilation and the Earnings of Guestworkers in Germany, mimeo Massachusetts Institute of Technology.

Polachek, S.W., Robst, J. (1998), Employee Labor Market Information: Comparing direct World of Work Measures of Workers' Knowledge to stochastic Frontier Estimates, Labour Economics 5, 231-242.

Polachek, S.W., Yoon, B.J. (1996), Panel Estimates of a two-tiered Earnings Frontier, Journal of Applied Econometrics 11, 169-178.

Robinson, M.D., Wunnava, P.V. (1989), Measuring Direct Discrimination in Labor Markets Using a Frontier Approach: Evidence from CPS Female Earnings Data, Southern Economic Journal 56, 212-218.

Schmidt, C.M. (1997), Immigrant Performance in Germany: Labor Earnings of Ethnic German Migrants and Foreign Guest Workers, The Quarterly Review of Economics and Finance 37, 379-397.

Wang, H.-J., Schmidt, P. (2002), One-step and Two-step Estimation of the Effects of Exogenous Variables on Technical Efficiency Levels, Journal of Productivity Analysis 18, 129-144.

Appendix

Table A-1
Description of the Primary Data (Mean Values)

		Natives		Immigrants							
				Eastern Europe		EU members		Turks		Former Yugoslavia	
Variable	**Description**	Mean	Std.dev.	Mean	Std.dev.	Mean	Std.dev.	Mean	Std.dev.	Mean	Std.dev.
EARNINGS	Gross EUR-income from wages/salaries per hour	16.16	6.620	13.38	4.160	15.06	4.427	13.67	3.761	13.41	2.880
SCHOOL	Years of schooling by five categories (7. 9. 10. 12 or 13 years). The school degree decides about the classification.	10.38	1.586	9.57	1.005	9.09	1.379	9.17	1.193	8.99	1.275
OCCUPATION	Equals 1 if employee received occupational training; 0 otherwise	0.75	0.432	0.81	0.392	0.49	0.501	0.58	0.495	0.69	0.466
COLLEGE	Equals 1 if employee has university degree. 0 otherwise	0.21	0.405	0.14	0.352	0.11	0.319	0.06	0.246	0.04	0.195
EXPERIENCE	Cumulated years of work experience in the past (without time of apprenticeships)	17.91	11.670	18.24	10.339	24.11	10.666	17.64	9.889	24.05	13.068
UNEMPLOYED	Cumulated years of unemployment in the past	0.64	1.611	1.22	1.873	1.03	1.885	1.56	2.639	1.61	2.591
STATUS	Occupational status: equals 1 if blue-collar worker; 0 otherwise	0.41	0.493	0.82	0.383	0.76	0.428	0.91	0.284	0.90	0.307
HOURS	Number of actual working hours per week	43.06	6.397	41.90	5.486	41.25	6.168	40.92	5.428	41.62	5.654
MARRIED	Equals 1 if married; 0 otherwise	0.66	0.474	0.85	0.357	0.79	0.405	0.90	0.308	0.73	0.448
CHILD	Number of dependent children living in household	0.63	0.929	1.083	1.227	0.89	1.014	1.49	1.267	0.623	0.987
SMALL	Equals 1 if firm has less than 20 employees. 0 otherwise	0.17	0.377	0.20	0.404	0.19	0.397	0.10	0.300	0.18	0.388
MEDIUM	Equals 1 if firm has more than 20 and less than 2000 employees. 0 otherwise	0.54	0.498	0.64	0.483	0.537	0.500	0.63	0.484	0.68	0.471
YSM	Years since immigration to West-Germany	-	-	12.59	6.049	26.74	9.666	22.58	7.339	25.36	8.545
LANGUAGE	Equals 1 if individual has modest or no knowledge of oral German. 0 otherwise	0.001	0.028	0.20	0.400	0.29	0.453	0.374	0.485	0.25	0.434
RETURN	Equals 1 if individual intends to re-migrate. 0 otherwise	0.01	0.077	0.03	0.180	0.41	0.494	0.29	0.453	0.46	0.501
	Number of observations	3852		181		175		171		77	

Male employees only; all data for 2000. Source: German Socio-Economic Panel (GSOEP); own calculations.

Table A-2

OLS and Maximum Likelihood Estimates of Earnings Equation

	OLS	**frontier earnings function**			
	(full sample)	**(natives only)**		**(full sample)**	
		half-normal	exponential	Half-normal	exponential
CONST	2.7988***	3.0321***	2.9651***	2.9979***	2.9347***
	(0.0606)	(0.0679)	(0.0671)	(0.0614)	(0.0605)
SCHOOL	0.0151***	0.0135***	0.0146***	0.0170***	0.0183***
	(0.0045)	(0.0050)	(0.0050)	(0.0045)	(0.0045)
OCCUPATION	0.0088	0.0101	0.0098	0.0087	0.0079
	(0.0125)	(0.0146)	(0.0145)	(0.0124)	(0.0123)
COLLEGE	0.2399***	0.2519***	0.2552***	0.2399***	0.2423***
	(0.0178)	(0.0199)	(0.0200)	(0.0178)	(0.0179)
EXPERIENCE	0.0147***	0.0139***	0.0136***	0.0138***	0.0135***
	(0.0014)	(0.0015)	(0.0015)	(0.0013)	(0.0013)
EXPERIENCE^2 /100	-0.0269***	-0.0246***	-0.0237***	-0.0245***	-0.0236***
	(0.0033)	(0.0036)	(0.0036)	(0.0033)	(0.0032)
UNEMPLOYED	-0.0414***	-0.0460***	-0.0458***	-0.0403***	-0.0400***
	(0.0028)	(0.0032)	(0.0032)	(0.0027)	(0.0027)
STATUS	-0.1883***	-0.2016***	-0.2014***	-0.1937***	-0.1939***
	(0.0117)	(0.0127)	(0.0125)	(0.0118)	(0.0116)
HOURS	-0.0079***	-0.0075***	-0.0077***	-0.0076***	-0.0079***
	(0.0008)	(0.0008)	(0.0008)	(0.0008)	(0.0008)
LANGUAGE	0.0170			0.0173	0.0146
	(0.0294)			(0.0291)	(0.0282)
RETURN	-0.0322			-0.0351	-0.0356
	(0.0285)			(0.0283)	(0.0277)
$YSM_{EAST\ EUROPE} \cdot D_{EAST\ EUROPE}$	0.0088**			0.0084**	0.0080**
	(0.0039)			(0.0038)	(0.0038)
$YSM_{EU} \cdot D_{EU}$	0.0036			0.0033	0.0031
	(0.0025)			(0.0024)	(0.0024)
$YSM_{TURKEY} \cdot D_{TURKEY}$	0.0084**			0.0085***	0.0085***
	(0.0033)			(0.0033)	(0.0032)
$YSM_{YUGO} \cdot D_{YUGO}$	0.0019			0.0017	0.0017
	(0.0042)			(0.0042)	(0.0042)
$D_{EAST\ EUROPE}$	-0.1608***			-0.1610***	-0.1574***
	(0.0557)			(0.0551)	(0.0539)
D_{EU}	-0.0410			-0.0366	-0.0338
	(0.0723)			(0.0715)	(0.0778)
D_{TURKEY}	-0.1565*			-0.1651**	-0.1662**
	(0.0809)			(0.0800)	(0.0778)
D_{YUGO}	-0.0045			-0.0110	-0.0133
	(0.1129)			(0.1126)	(0.1115)
σ^2		0.1540***		0.1462***	
		(0.0088)		(0.0076)	
σ_v^2	0.0910	0.0740	0.0776***	0.0685	0.0714***
			(0.0033)		(0.0028)
λ		1.0392***		1.0648***	
		(0.1095)		(0.0994)	
θ			6.2705***		6.3120***
			(0.4221)		(0.3720)
$\overline{R}^2$ / Log Likelihood	$\overline{R}^2$=0.304	-1083.6	-1072.0	-1109.8	-1093.5
LR-test that all slopes are 0		1422.1***	1437.7***	1683.0***	1704.3***
Observations	4456	3852	3852	4456	4456

Dependent variable is ln hourly (gross) earnings.

*, ** and *** represent a significance level of 90%, 95% and 99%, respectively.

Table A-3

Differences in Human Capital Endowment

Δ human capital endowment	**immigrants from …** Eastern Europe	EU	Turkey	Yugoslavia
SCHOOL	-1.1%	-1.7%	-1.6%	-1.9%
COLLEGE	-1.6%	-2.3%	-3.6%	-4.3%
EXPERIENCE	0.9%	2.8%	0.8%	1.3%
UNEMPLOYED	-2.7%	-1.8%	-4.2%	-4.4%
STATUS	-8.2%	-6.9%	-9.9%	-9.6%
HOURS	0.9%	1.3%	1.6%	1.1%
else	0.4%	-0.4%	0.6%	-0.9%
total human capital differentials*	-11.4%	-9.0%	-16.3%	-16.9%
For comparison: observed wage differentials	-17.2%	-6.8%	-15.4%	-17.0%

Impact of human capital variables on wage frontier, estimated from half-normal model.
All estimates from mean values of each group relative to a representative native employee.