IZA/CEPR 11TH EUROPEAN SUMMER SYMPOSIUM IN LABOUR ECONOMICS

**Supported and Hosted by the Institute for the Study of Labor (IZA)**

Buch, Ammersee

17-19 September 2009

# Understanding the Native-Immigrant Wage Gap Using Matched Employer-Employee Data. Evidence from Germany

Cristian Bartolucci

# Understanding the Native-Immigrant Wage Gap using Matched Employer-Employee data. Evidence from Germany*

Cristian Bartolucci

Collegio Carlo Alberto†

September 2009

**Abstract**

Hellerstein and Neumark (1999) developed a very straightforward method to detect wage discrimination using matched employer-employee data. However it may produce biased estimates whenever there is not perfect competition in the labor market or when the discriminated group is segregated into good or bad firms. The purpose of this paper is to develop a test for wage discrimination that completes the Hellerstein and Neumark (1999) approach. To do this I propose to estimate a wage setting equation at the firm level that exploits changes in the native-immigrant composition within firms across time in order to have identification of different wage policies toward those groups. Using matched employer-employee data from Germany. I show that both bias-sources are empirically significant when analysing discrimination against immigrants. I find that immigrants are being discriminated. They receive wages which are 16 percent lower than native workers in the same firm. I also find that immigrants are positively segregated into good firms. I do not find significant evidence in favour of a taste-based discrimination model but I do find evidence against a statistical discrimination model.

JEL Code: J71, J64

KEYWORDS: Labor market discrimination, immigration, matched employer-employee data.

---

*This article is a revised version of a chapter from my PhD dissertation at CEMFI. I would like to thank Manuel Arellano, Stéphane Bonhomme, Raquel Carrasco, Jose María Labeaga, Pedro Mira, Enrique Moral-Benito, Diego Puga, Jean-Marc Robin, and members of the audiences in the EEA-Mettings (2009), Workshop on Wages and Firms: New Research Using Linked Employer-Employee Datasets (2009) and CEMFI for very helpfull comments and suggestions. I am also indebted with Dana Muller and Daniela Hochfellner from the Institute for Employment Research for invaluable support with the data.

†Via Real Collegio, 30 - 10024 - Moncalieri (Torino), Italy. Email: cristian.bartolucci@carloalberto.org.

# 1 Introduction

The Hellerstein and Neumark strategy has been found to be a very direct and popular method to detect wage discrimination using matched employer-employee data. However it may produce biased estimates whenever there is not perfect competition in the labor market or when the discriminated group is segregated into good or bad firms. The purpose of this paper is to develop a test for wage discrimination that completes the Hellerstein and Neumark (1999) approach. To do this I propose to estimate a wage setting equation at the firm level that exploits changes in productivity and changes in the native-immigrant composition within firms across time to have identification of different wage policies toward those groups. Using Matching employer-employee data form Germany I show that this bias is empirically significant when analyzing discrimination against immigrants.

In the Altonji and Blank's handbook chapter (1999), labor market discrimination is defined as a situation in which persons who provide labor market services and who are equally productive in a physical or material sense are treated unequally in a way that is related to an observable characteristic such as race, ethnicity or gender.

The most widely-used approach to test for labor market discrimination takes the unexplained gap in wage regressions as evidence of discrimination. This method, also known as the residual method, estimates Mincer-equations for both groups and then it decomposes the difference of mean wages into "explained" and "unexplained" components. The fraction of the gap that cannot be explained by differences in observable characteristics is considered as discrimination. In the Altonji and Blank's definition spirit, the residual approach may be understood as a comparison of wages and productivity where the last one is approximated by a function of observable characteristics. However, if there are unobservable characteristics that correlate with migration status and that are also correlated with productivity[1], the discrimination measure may be biased.

[1]We typically think on environmental variables, tastes, education quality and language

The availability of matched employer-employee data allowed a response to this potential weakness of the residual approach. Hellerstein and Neumark (1999), and Hellerstein, Neumark and Troske (1999), proposed a method that uses firm-level data to estimate relative marginal products of various workers types to be then compared with their relative wages. Productivity of each worker type is estimated in terms of the proportion of workers of each type in the firm. Whenever perfect competition holds in the labor market any difference in wages that is not driven by a difference in productivity may be considered as discrimination. There have been a number of papers applying this approach for different countries as the already mentioned Hellerstein and Neumark, (1999) with Israeli data, and Hellerstein, Neumark and Troske, (1999) with U.S. data but also Verner (1999) with data from Ghana, Crepon, Deniau and Pérez-Duarte (2003) with French data, Zhang & Dong (2006) with Chinese data, Kawaguchi (2007) with Japanese data, Van Biesebroeck, (2007) with data from three Sub-saharian countries and Campos-Vazquez (2008) with German data.

Although this approach has been found to be a more direct way to test for discrimination it may be criticized in two dimensions: The first one addresses that variation in worker composition is likely to be correlated with heterogeneity in the firm's technology and may be, therefore, endogenous to the model. The second criticism deals with their assumption of perfect competition. If this condition does not hold it is not clear how meaningful is the estimated difference between relative-wages and relative-productivity.

In this paper I propose a method to test for labor market discrimination that completes the Hellerstein and Neumark (1999) approach. I take advantage of a matched employer-employee panel data set to estimate a reduced form wage setting equation at the firm level and to test, controlling for productivity and firm's fixed characteristics, whether the proportion of immigrants is significant. This approach exploits changes in the native-immigrant composition within the firm across time to have identification of different wage policies toward those groups. The longitudinal dimension of the data allows me to have estimates

skills.

that are robust to any correlation of the worker composition with the firm fixed effect. This reduced form permits me to remain agnostic about the true labor market model. It could fit the data in a perfect competition scenario, but also in a labor market with frictions scenario.

The panel data also allows me to estimate firm specific discrimination parameters following the strategy presented in Arellano and Bonhomme (2008). Although these estimates are very noisy - I have a small-T panel -, I can estimate the unbiased correlation with some other firm variables, such as profit or tenure of immigrants, in order to have indirect evidence of different discrimination theories, testing some of their implications.

I use a 1996-2005 panel of matched employer-employee data provided by the German Labor Agency, called LIAB.[2] This dataset is especially useful for this study for two reasons. Firstly, it contains essential data about workers' nationality. Secondly, it is a panel that tracks firms as opposed to individuals, which is necessary to have estimates in the wage setting equation that are robust to a correlated fixed effect.

I find that both sources of bias are important when analyzing discrimination against immigrants in Germany. Immigrants are suffering wage discrimination, and depending on the measure of productivity used, discrimination ranges between 12.8 percent and 16.8 percent. This finding is surprising if we take into account that both the traditional approach and the Hellerstein and Neumark (1999) approach conclude that immigrants are not receiving significantly lower wages. The elasticity of wages to productivity is significantly different from one and hence assuming wages equal to productivity may be dangerous. When estimating by OLS, discrimination was found to be significantly lower which gives evidence of positive segregation of immigrants into good firms. Positive segregation would imply an underestimation of discrimination. Although the reduced-form wage setting equation is very simple, it has an acceptable fit of the wage-bill data. I do not find neither significant evidence of immigrants

[2] This dataset is subject to strict confidentiality restrictions. It is not directly available but only after the IAB has approved the research project, The Research Data Center (FDZ) provides on site use or remote access to external researchers.

moving to less discriminatory firms nor significant evidence in favor of a taste-based discrimination model nevertheless I do find evidence against a statistical discrimination model.

A great part of the empirical literature on discrimination has been focused on gender and racial discrimination. Wage differentials between natives and immigrants have generally been understood as an assimilation process that involves differences in productivity, such as language skills (e.g. Borjas 1994, Chiswick and Miller 1995; Carnevale et al. 2001; Dustmann and van Soest 2002); differences in education quality (Sweetman 2003) or differential returns to foreign schooling and labor market experience (e.g. Friedberg 2000 and Bratsberg and Ragan 2002). As discrimination has normally been detected through the unexplained gap in wage equations and this approach is not able to disentangle differences in productivity and discrimination, there are few papers studying labor market discrimination against immigrants. Some exceptions, also with matched employer-employee data, are Aydemir and Skuterud (2008) that studies the relative importance and sources of immigrants wage differentials within and across establishments in Canada, and Aeberhardt and Pouget (2008). For Germany there is a new working paper by Campos-Vazquez (2008) that uses the same LIAB data than this paper and replicates the Hellerstein and Neumark (1999) analysis to test for discrimination against German immigrants.

The rest of this paper is organized as follows. In the next section, I briefly describe the immigration phenomena in Germany. In section 3, I present the model and I formally compare it with the Hellerstein and Neumark (1999) approach. In the fourth section, I present the data-set. Section five presents results and robustness check. In section six, I show how this method can be used to distinguish between different discrimination theories and in the last section, I conclude.

## 2 Background

Germany is a very interesting country to study migration mainly because immigrants represent an important and stable fraction of the population. The proportion of immigrants has had very small changes in the last 15 years ranging between 8.2 percent and 8.9 percent, see Figure (1).

The first Immigration wave that immediately followed the end of the Second World War started when several millions of refugees from the former East Germany and from East European regions resettled in the Federal Republic of Germany and a later one, until the construction of the Berlin Wall, when Germans from the German Democratic Republic were able to enter to West Germany.

The Second immigration wave started in 1955, when Italy and Germany signed a treaty, which allowed organized recruitment of Italian workers to meet the needs of the growing German economy. The recruitment of the foreign labor force intensified dramatically reacting to the sharp increase in the demand for additional labor force. This policy was expanded to the following countries: Spain and Greece (1960), Turkey (1961), Morocco (1963), Portugal (1964), Tunisia (1965) and Yugoslavia (1968).[3] These agreements were intended to meet the needs of the German economy by reducing the movement costs of unskilled workers. Foreign workers were recruited to Germany on a temporary basis.

The practice of the foreign labor recruitment stopped in 1973, following the oil crisis and a sharp decrease in the labor demand. The end of the labor recruitment and new barriers for foreign workers to settling in Germany minimized the short-term immigration and started a new tendency toward permanent settlement among those who entered Germany as temporary workers. This was the start of the next period of immigration to Germany, the one based on family reunifications of guest workers who arrived earlier. The Turkish population was the main one in taking advantage of this possibility and in spite of the halt

[3] See Rudolph (1994) for a good description of this phenomena.

placed on recruitment in 1973, it continued to rise and it now forms the largest foreign minority in Germany.

Since the late 1980s, the inflow of refugees and asylum seekers has increased and marked another phase in the post-war immigration. The number of asylum applicants rose significantly in the second half of the 1980s and peaked at 440,000 in 1992, partly as a result of the war in the former Yugoslavia. Between 1988 and 1992, 1.1 million asylum-seekers filed applications. As a reaction to this, the German Parliament agreed to the "asylum compromise" in 1993, which made applying for political asylum in Germany considerably more difficult. Hence, the number of applications for asylum has declined steadily and the proportion of immigrants has stabilized, see figure (1).

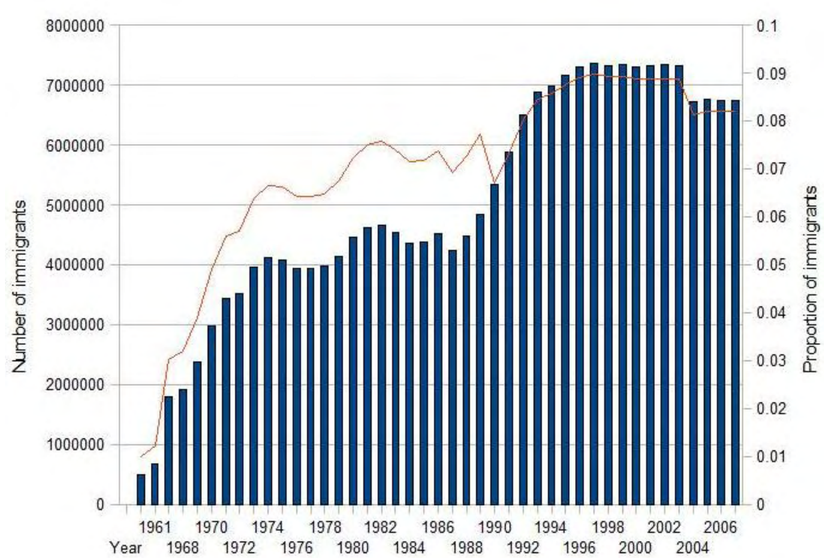


Figure 1: Number and Proportion od Immigrants

The percentage of immigrants in Germany increased from less than 1 percent, 506,000 foreigners in 1955 to 8.2 percent in 2007[4], 6,744,879 registered immigrants[5], see figure (1). In terms of workers, in my sample the fraction of

[4]Data from the Federal Office for Migration and Refugees.

[5]There are no statistics concerning irregular immigration or immigrants staying in Ger-

immigrants is slightly higher and ranges from 9.4 to 10.9 percent between 1996 and 2004. See Section 4 for more details.

# 3 The Model

The Hellerstein and Neumark Approach has been found to be a very direct, and very popular, method to detect wage discrimination. As it was stated in the introduction, this approach may produce biased estimates whenever there is not perfect competition in the labor market or when immigrants are not randomly distributed between firms. In order to formalize these criticisms, let me assume that the wage setting equation takes the following form:

$$w_{ijt} = \alpha_j + \beta p_{ijt} + \gamma I_i + \epsilon_{ijt}, \tag{1}$$

where $w_{ijt}$ is the log-wage of individual $i$, on firm $j$, at time $t$, $p_{ijt}$ is the individual log-productivity and $I_i$ is an immigrant indicator. In this context I interpret $\epsilon_{ijt}$ as an econometric mean-zero residual term due to the imposed linearity in the wage setting equation or to measurement errors.

In $\alpha_j$, the firm fixed effect, there may be observed firm fixed characteristics like region, sector or unionization of the workforce and there could also be unobserved ones as wages policies, risk aversion, technology or managerial quality.

This wage setting reduced form is very flexible, it allows me to be agnostic about the labor market model that is behind the data. It may be compatible with a perfect competition model where the elasticity of wages to productivity is one ($\beta = 1$), but it could also be valid to fit a wage setting equation in a labor market with frictions model, where only a fraction of the marginal product is paid to the worker and hence $\beta < 1$.

As discrimination was defined as a situation in which workers who provide labor market services and who are equally productive in a physical or material

many without a permit. Unofficial estimates, which refer to between 500,000 and one million irregular immigrants residing in Germany, are not based on scientific assessment. As data used in this paper comes from social security records I consider only registered immigrants.

sense are treated unequally in a way that is related with their migration status, a direct test for discrimination would be to test $\gamma \neq 0$.

To directly estimate (1) is not feasible because, generally, individual productivity is unobserved. There are some cases where individual productivity is more easily measured as in academic positions, see Ferber and Green (1982) or in jobs with under-piece contracts, see Milgrom, Petersen and Snartland (2006). Although this kind of studies may have measures of individual productivity, they are likely to be weak in terms of external validity.

This test may be connected to what Hellerstein and Neumark (1999) and Hellerstein, Neumark and Troske (1999) proposed. They used matched employer-employee data from Israel and US to estimate relative marginal products of various workers types. Then they compare productivity differentials ($\rho = \frac{E(P_{ijt}|I_i=1)}{E(P_{ijt}|I_i=0)}$) with wages differentials ($\lambda = \frac{E(W_{ijt}|I_i=1)}{E(W_{ijt}|I_i=0)}$).

In Hellerstein and Neumark (1999), $\lambda$ is estimated using data on total firm's wage bills and proportion of women. They estimate by nonlinear least squares the following equation:

$$\ln(\bar{w}_{jt}) = cons + \ln(1 + (\lambda - 1)\frac{L_W}{L}),$$

where $\bar{w}_{jt}$ is the mean wage paid by firm $j$, $L$ is the total number of workers in the plant $j$, and $L_w$ is the proportion of women.

$\rho$ is estimated with production functions, assuming a Cobb-Douglas or translogarithmic functional forms with quality adjusted labor input. In their simpler case, they estimate marginal products of women and men by NLLS in the following equation:

$$Ln(Y_{jt}) = Ln(A_j) + \alpha Ln(K_{jt}) + bLn(M_{jt}) + \gamma Ln(L^Q_{jt}) + g(K_{jt}, M_{jt}, L^Q_{jt}),$$

where $K_{jt}$ is capital, $M_{jt}$ is material, $g(K_{jt}, M_{jt}, L^Q_{jt})$ is the second order term in the production function and $L^Q_{jt}$ is quality of labor aggregate that is defined as:

$$L^Q = L\left\{1 + (\varphi - 1)(L^W/L)\right\}$$

where $L$ is the total number of workers in the plant and $L^W$ is the number of women in the plant.[6]

We may easily interpret their strategy within the framework presented above. Noting that taking averages of equation (1) across groups (immigrants and natives), we have:

$$\frac{\Sigma(w_{ijt}|I_i=0)}{\Sigma 1(I_i=0)} = \frac{\beta\Sigma(p_{ijt}|I_i=0)+\Sigma(\alpha_j|I_i=0)+\gamma I_i+\Sigma(\epsilon_{ijt}|I_i=0)}{\Sigma 1(I_i=0)}, \quad (2)$$

and:

$$\frac{\Sigma(w_{ijt}|I_i=1)}{\Sigma 1(I_i=1)} = \frac{\beta\Sigma(p_{ijt}|I_i=1)+\Sigma(\alpha_j|I_i=1)+\gamma I_i+\Sigma(\epsilon_{ijt}|I_i=1)}{\Sigma 1(I_i=1)}. \quad (3)$$

Subtracting (3) from (2) and noting that $\Sigma(\epsilon_{ijt}|I_i=1)=0$ and $\Sigma(\epsilon_{ijt}|I_i=0)=0$ [7], we have:

$$\begin{aligned} & \lambda-\rho \qquad (4) \\ \simeq \quad & \gamma + \underbrace{\underbrace{(1-\beta)(1-\rho)}_{\text{Perfect Comp. Bias}} + \underbrace{\left[\frac{1}{\Sigma 1(I_i=1)}\Sigma(\alpha_j|I_i=1)-\frac{1}{\Sigma 1(I_i=0)}\Sigma(\alpha_j|I_i=0)\right]}_{\text{Segregation Bias}}}_{\text{Potential Bias}} \end{aligned}$$

where $\lambda$ represents the mean of immigrants' wages relative to the natives' mean-wages and $\rho$ represents the mean of immigrants' productivity relative to the

[6] The Hellerstein and Neumark (1999) model is more complicated because they allow for several population groups. See Hellerstein and Neumark (1999) for details.

[7] This is easily proved noting that:

$$\begin{aligned} & \overbrace{\left[\frac{1}{\Sigma 1(I_i=1)}\Sigma(w_{ijt}|I_i=1)-\frac{1}{\Sigma 1(I_i=0)}\Sigma(w_{ijt}|I_i=0)\right]}^{\to E(w_{ijt}|I_i=1)-E(w_{ijt}|I_i=0)\simeq\lambda-1} - \\ & \beta\overbrace{\left[\frac{1}{\Sigma 1(I_i=1)}\Sigma(p_{ijt}|I_i=1)-\frac{1}{\Sigma 1(I_i=0)}\Sigma(p_{ijt}|I_i=0)\right]}^{\to E(p_{ijt}|I_i=1)-E(p_{ijt}|I_i=0)\simeq\rho-1} \\ = \quad & \gamma+\left[\frac{1}{\Sigma 1(I_i=1)}\Sigma(\alpha_j|I_i=1)-\frac{1}{\Sigma 1(I_i=0)}\Sigma(\alpha_j|I_i=0)\right], \quad \blacksquare \end{aligned}$$

natives' one. $\lambda - \rho$ is a measure of wage discrimination if $\beta = 1$ and $cov(\alpha_j, I_i) = 0$. Hellerstein and Neumark (1999), make inference about discrimination (i.e.: $\gamma$) simply comparing $\lambda$ and $\rho$ estimated at the firm level. Although they are very cautious in their interpretation of this difference arguing that $\hat{\lambda} - \hat{\rho}$ gives evidence in favor of discrimination if $\beta \neq 1$ or $cov(\alpha_j, I_j) \neq 0$ there is not an *a priori* direction of the bias and hence it is not clear how informative are their findings.

In order to be clearer in this explanation let me decompose the bias in two components:

*Perfect Competition Bias:* The first part of the bias is addressing that whenever a change in productivity is not fully transferred to the wage, two groups with different productivity may have larger or smaller differences in wages that are not implying discrimination. As it can be seen in Section 5, $\beta$ is found to be significantly different from one. Depending on the specification and the measure of productivity used, it ranges between 0.25 and 0.45. To show numerically how important may be this bias, let me consider a very simple example where there are two groups $A$ and $B$, where $A$ is 20 percent more productive than $B$. If there is not discrimination against any group and assuming that $\beta = 0.4$ workers of the $A$ group are supposed to have wages only 5 percent higher than workers of the $B$ group. The Hellerstein and Neumark (1999) approach would wrongly imply that workers of the $A$ group are being discriminated.

*Segregation Bias:* The second part of the bias gives us information about segregation. This bias is important if high wages firms hire more or less immigrants than low wages firms. the firm fixed effect there may be observed and unobserved fixed characteristics. Hellerstein and Neumark (1999) try to deal with some of the observed ones clustering the analysis at different levels.

One of the Altonji and Blank (1999) criticisms to the Hellerstein and Neumark (1999) approach is related to this segregation bias, the last term in (4). They argue that the variation in worker composition is likely to be correlated with heterogeneity in the production technology and may be endogenous to the model. In this context the firm's technology may have an effect over the firm's fixed effect, $\alpha_j$. But this criticism go further because in Hellerstein and Neumark (1999) papers, $\rho$ is estimated with cross sectional data and therefore their estimates are not robust to a potential correlation between workers composition and the firm fixed effect in the production function.

### 3.1 Detecting discrimination at the firm-level

Without measures of individual productivity, and having shown that to aggregate at the group level might be dangerous in some cases, a second best would be to aggregate equation (1) within the firm:

$$\bar{w}_{jt} = \alpha_j + \beta \bar{p}_{jt} + \gamma \bar{I}_{jt} + \bar{\epsilon}_{jt}. \tag{5}$$

Therefore $\bar{w}_{jt}$ is the mean of log-wages in firm $j$, $\bar{p}_{jt}$ is the mean of individual log-productivity of firm $j$, and $\bar{I}_{jt}$ is the proportion of immigrants in firm $j$ at time $t$.[8] I estimate equation (5) replacing $\bar{w}_{jt}$ by the log of mean wages of firm $j$ and $\bar{p}_{jt}$ by the log of the output per-worker.[9]

The conceptually relevant measure of productivity should be the marginal productivity of workers in firm $j$. Assuming a Cobb-Douglas production function:

$$Y_{jt} = A_j K_{jt}^{\phi_k} L_{jt}^{\phi_l},$$

where $A_j$ is the firm fixed effect, $Y_{jt}$ is the output of firm $j$ at time $t$, $K_{jt}$ is its capital and $L_{jt}$ is its labor input. The marginal productivity of labor is given by $\phi_l A_j K_{jt}^{\phi_k} l_{jt}^{(\phi_l - 1)}$, that equals $\phi_l Y_{jt}/L_{jt}$. Therefore using the log of mean output or using the log of marginal productivity would only modify the constant term $\alpha_j$ in (5) adding $\log(\phi_l)$.

This specification has the advantage that it does not necessarily involve estimating relative productivity as in Hellerstein and Neumark (1999). That es-

[8] The migration status indicator, $I_i$, is worker specific, but the proportion of immigrant workers is firm and time specific

[9] As in LIAB there are data on total output and total wage bill paid by each firm, to estimate equation (5), I replace the mean of the log by the log of the mean in wages and productivity. The error of approximation goes to the residual and is not supposed to be correlated with regressors. This is deeply discussed in the appendix.

timation usually involves the estimation of a cobb-douglas production function with quality adjusted labor input. In order to have estimates robust to any correlation of inputs, including labor input composition, with the firm fixed effect, the production function should be estimated by differenced-GMM as in Arellano and Bond (1989) or by SYSTEM-GMM using the set of instruments proposed in Arellano and Bover (1995). When estimating this kind of production function by GMM, I significantly loose precision and the quality-adjustment parameters are almost non-informative. This problem is usual in this production function specifications. In the appendix I present the Non Linear Least Squares estimates of the production function in levels to compare my results with those obtained with the Hellerstein and Neumark (1999) approach.

A final issue would be to allow for heterogeneity in $\gamma$s. This would imply a difference if $\bar{I}_{jt}$ is correlated with $\gamma_j$ and then $E(\gamma_j) \neq \gamma$. If immigrants are self selected into less discriminatory firms this correlation would be positive and then, $E(\gamma_j) \neq \gamma$. This possibility will be discussed in the last section of this paper.

# 4 Data

The data I use for the present study refer to West-German workers[10] contained in the linked employer-employee dataset of the IAB (LIAB) covering the period 1996-2005. LIAB is created by matching the data of the IAB establishment panel and the process-produced data of the Federal Employment Services (Social security records).

The IAB Establishment Panel is an annual survey of German establishments, which started in western Germany in 1993 and was extended to eastern Germany in 1996. The sample of selected establishments is random and stratified by industries, establishment size and regions. The sample unit is the establishment as the local business unit. The establishments asked in the survey are selected

[10]All employees and trainees subject to social security are included, while the selfemployed, family workers, a subgroup of civil servants ("Beamte"), students enrolled in higher education and those in marginal employment are excluded

from the parent sample of all German establishments that employ at least one employee covered by social security. Participation of establishments is voluntary, but the response rates are high, they exceed 70 percent.[11] The firm's data gives details on total sales, value added, investment, depreciation, number of workers and sector. I only consider firms with strictly positive output. To ensure a consistent comparison of results across specifications, the data used for each specification exclude observations with missing values for any of the independent variables used in the regressions. Firms in the financial and public sectors are excluded from my subsample, see Table 1 for some descriptive statistics.

Table 1: Firms

| Firm's Descriptive Statistics | |
|---|---|
| Workers | 211.86 |
| output* | 54.3 |
| depreciated Capital* | 1,38 |
| value Added* | 23.0 |
| Total Wage Bill* | 8.24 |
| Unionized | 63.4% |
| Single Establishment | 70.2% |
| Immigrant's Proportion | 7.7% |
| Imm. Prop within-firm Std. Dev. | 2.6% |
| Observations | 20.886 |

Note: * per annum in millions of euros. Descriptive statistics obtained from the panel of firms.

The distinctive feature of this data is the combination of information about individuals and details concerning the firms in which these people work. The workers source contains valuable data on age, sex, nationality, daily wage (censored at the upper earnings limit for social security contributions), schooling/training, occupation based on a 3-digit code and the establishment number.

In Table 2, I present some descriptive statistic of both immigrants and natives, estimated from my sample. The proportion of women is significantly higher in the native population. Immigrants are younger and they have less tenure and experience. There are important differences in term of occupations

[11] For a more precise description of this dataset, see Alda et al (2005)

Table 2: Demographic Differences

| | Immigrants | Natives |
|---|---|---|
| Sex (%) | 25.6 | 31,2 |
| Age (years) | 39.6 | 40.4 |
| Tenure (years) | 10.5 | 11.1 |
| Experience (years) | 15.1 | 16.7 |
| Unskilled (%) | 80.9 | 52.4 |
| Part-time jobs (%) | 9.2 | 12.8 |
| Agriculture (%) | 2.5 | 3.9 |
| Manufacturing (%) | 70.3 | 59.1 |
| Construction (%) | 3.0 | 3.3 |
| Trade (%) | 3.5 | 6.9 |
| Services (%) | 20.6 | 26.7 |
| Daily Wages (€) | 109.0 | 94.7 |
| Observations | 1.185.362 | 11.832.370 |

Note: Descriptive statistics estimated from the panel of workers. As wages are censored at the upper earnings limit for social security contributions, mean-wages are obtained by Maximum-Likelihood assuming log-normality.

and sectors. Immigrants are more concentrated in the manufacturing sector and low-skill occupations[12] than natives.

# 5 Results

Although the conceptually proper measure of productivity is value added, in this data set this measure may have some reliability problems.[13] Here I opted to report results with both measures.

In Table 3, I present the results. In Columns (1) and (2), I report the estimates without including any measure of productivity. The OLS estimate of $\gamma$ in column (1) is understood as the unconditional wage gap, only controlling for time effects. This wage gap is obtained from firm-level data and it is not statistically different from the unconditional wage gap obtained from worker-

[12]Following the FDZ's criteria, I have considered as unskilled jobs the following groups: Agrarian occupations, manual occupations, services and simple commercial or administrative occupations, While I have considered as skilled jobs Engineers, professional or semi-professional occupations, qualified commercial or administrative occupations, and managerial occupations.

[13]See Adisson, Schank, Schnabel and Wagner (2003) for a good discussion on this issue.

Table 3: Wage Setting Equation

| $\overline{w}_{jt}$ | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| $\beta_{\overline{output}}$ | - | - | 0.426 | 0.486 | - | - |
| | - | - | (0.013) | (0.006) | - | - |
| $\beta_{\overline{Value-Added}}$ | - | - | - | - | 0,392 | 0.233 |
| | - | - | - | - | (0.013) | (0.005) |
| $\gamma$ | **-0.181** | **-0.258** | **0.070** | **-0.126** | **0.032** | **-0.168** |
| | (0.044) | (0.043) | (0.040) | (0.041) | (0.039) | (0.049) |
| Fixed Effects | no | yes | no | yes | no | yes |
| obs. | 29,943 | 29,943 | 29,943 | 29,943 | 29,943 | 29,943 |
| $R^2$ | 1.2% | - | 37% | - | 33% | - |
| $R^2$-With. | - | 1.4% | - | 28% | - | 13% |
| $R^2$-Betw. | - | 0.2% | - | 38% | - | 35% |

Note: Each column represents a single linear regression using the panel of firms. Time dummies are included in every specification. Standard errors in parentheses. In OLS regressions (column 1, 3 and 5) standard errors are calculated clustering by firm.

level data.[14] On the other hand, $\gamma$ estimated by within groups, without further controls, refers to the average unconditional wage gap within firm. The difference between the overall wage differential and the within firm wage gap is informative about sorting of immigrants into firms, this issue will be discussed deeply in the next subsection.

In Columns (2) and (3), I report estimates using total output as a measure of productivity. In this specification the estimated premium for being an immigrant is 7 percent, marginally significant (p-value = 0.082). But estimating the same specification by within groups it is noteworthy that the discrimination parameter is significantly lower -12.6 percent, that would imply that immigrants are being discriminated. This finding is surprising if we take into account that, using the same data, both the traditional approach and the Hellerstein and Neumark (1999) approach conclude that immigrants are not receiving significantly lower wages than natives.[15]

Estimating equation (5) but including value added as a measure of produc-

[14] The unconditional wage gap obtained from worker-level data is -13.1 percent, see Table 8
[15] See sections B and C in the appendix

tivity, $\hat{\beta}$ is lower in both estimations, OLS and WG. I find the same pattern in terms of $\gamma's$ than in columns (3) and (4). The lower punctual estimate of $\beta_{Value-Added}$, and the lower $R^2$ may be understood as evidence of measurement error in value added as it was pointed out by Addison et al (2003).

It is important to note that $\beta$, the elasticity of wages to productivity, is found to be significantly different from one in every specification and hence to assume wages equal to productivity, as in Hellerstein and Neumark (1999) may be critical.

Note that without including any measure of productivity there is obviously a very poor fit of the wage data. But including productivity in these regressions the $R^2$ become acceptable and similar to the $R^2$ obtained in individual level wage regressions.

## 5.1 Segregation

The positive difference $\hat{\gamma}_{OLS} - \hat{\gamma}_{WG}$ may be understood as evidence in favor of positive segregation of immigrants into firms with higher fixed effect. This positive segregation implies an underestimation of discrimination when the within-firm variation is not isolated. In the appendix B I present discrimination measures estimated with Mincer-Equations and a Oaxaca-Blinder decomposition. I find that the unexplained wage gap is 2 percent that would mean that immigrants have a positive premium, this is also found in this analysis with OLS. My hypothesis is that this positive premium is mainly due to positive segregation.

In general, the concept of segregation aims to capture systematic sorting by workers belonging to different groups. Segregation becomes interesting when this sorting is associated with job characteristics that finally affect wages. Different job characteristics or segregation dimensions have different interpretations. Whenever the concentration of workers is higher in some regions or in some sectors, it may be revealing self selection of immigrants. On the other hand if immigrants are systematically sorted into the worst payer firms, within a region, sector and firm's size cell, it may be giving us evidence in favor of structural differences between both groups.

The role of various dimensions of immigrant segregation in generating differences in wages may be empirically investigated. I estimate equation ( 5) including different sets of controls for firm characteristics. Results are reported in Table 4. As these characteristics are fixed or they have very little variation across time, the specification when all these controls, and other non-observed firm fixed characteristics, are included is equivalent to the within group regression reported in column (5).

Table 4: Segregation

| $\overline{w}_{jt}$ | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| $\beta_{output}$ | 0.426 | 0.419 | 0.446 | 0.427 | 0.486 |
| | (0.014) | (0.014) | (0.015) | (0.016) | (0.006) |
| $\gamma$ | **0.070** | **0.026** | **0.039** | **-0.066** | **-0.126** |
| | (0.040) | (0.042) | (0.038) | (0.040) | (0.042) |
| Region | no | yes | no | yes | - |
| Sector | no | no | yes | yes | - |
| Firm Characteristics | no | no | no | yes | - |
| Firm Fixed Effects | no | no | no | no | yes |
| *obs* | 24.943 | 24,943 | 20,886 | 23,720 | 19,663 |
| $R^2$ | 0.372 | 0.372 | 0.433 | 0.445 | 0.369 |
| | (1) | (2) | (3) | (4) | (5) |
| $\beta_{Value-Added}$ | 0.392 | 0.385 | 0.372 | 0.355 | 0.233 |
| | (0.014) | (0.014) | (0.014) | (0.014) | (0.005) |
| $\gamma$ | **0.033** | **0.014** | **0.010** | **-0.081** | **-0.168** |
| | (0.040) | (0.041) | (0.038) | (0.040) | (0.046) |
| Sector | no | yes | no | yes | - |
| Region | no | no | yes | yes | - |
| Firm Characteristics | no | no | no | yes | - |
| Firm Fixed Effects | no | no | no | no | yes |
| *obs* | 24,943 | 24,943 | 20,886 | 24,943 | 19,663 |
| $R^2$ | 0.335 | 0.334 | 0.369 | 0.387 | 0.347 |

Note: Each column represents a single within-group linear regression using the panel of firms. Time dummies are included in every specification. Standard errors in parentheses. In OLS regressions (column 1,2,3 and 4) standard errors are calculated clustering by firm. $R^2$ do not take into account the variation in firm fixed effects.

The first and last columns replicate the results reported in Table 3. In Column (2), I report results when only controls for region are included. When I control for region, I observe that $\gamma$ is smaller but not significantly. This finding connects with Borjas (1999) which argues that immigrants are not randomly assigned to regions, presumably they choose areas which provide them the best opportunities. To find a lower $\gamma$ when controlling by region is consistent with this assessment, because is telling us that part of this positive premium that immigrants are obtaining is due to their choice of region. Positive segregation in term of region has also been found in Canada by Aydemir and Skuterud (2008).

Column (3) reports results when I only control for Industry. This control may be important if we take into account that the sectoral composition is significantly different across migration status, see Table 2. A comparison of $\gamma$ from columns (1) and (3) is informative about the effect of the industrial segregation over wages. The difference in $\gamma$ is -3.1 percent when using output and -2.3 percent when using value-added.[16] The negative difference in $\gamma$ is providing evidence in favor of positive segregation of immigrants into better industries. There are several studies trying to find what proportion of the gender and racial wage gap is due to interindustry differences in worker composition. For immigrants this literature is smaller, a good example is again Aydemir and Skuterud (2008) with Canadian matched employer-employee data. They surprisingly found that immigrants are employed in industries with slightly lower wage effects.

In Column (4), I present results when region-effects, sector-effects and other firm characteristics are included. I have considered firm size, an indicator of unionization[17] and an indicator that takes the value one if the firm is a single-establishment. We observe that $\gamma$ is significantly lower than the one reported in column (1). This finding suggest that the positive wage premium that immigrants are receiving in column (1) is consequence of their choices of sector, regions and firm's type, once we control for them, immigrants receive wages between 7 percent and 8 percent lower than natives.

It is surprising that there is a great part of the wage differential that is not accounted by this "observable" segregation. Comparing $\gamma$ from columns (4) and

[16] These differences are not significantly different from zero.

[17] In the IAB Establishment survey there is an explicit question that ask if the establishment is bound by industry-wide wage agreements, a company agreement concluded by the establishment and trade unions or not bound by collective agreements.

(5), we observe that once I control for observable and unobservable firm fixed characteristics the wage premium for immigrants is lower. This difference is significant when using value-added and almost significant when using output. This finding would imply that immigrants are hired in better firms than natives also within each region, sector and firm's characteristics cell. But within the firm, they are receiving wages between 13 percent and 17 percent lower than natives.

A last issue related with segregation is the difference between segregation among establishment and segregation within establishments. I have referred to segregation meaning segregation among establishments. $\gamma$, the measure of discrimination considered in this paper, capture both direct wage discrimination and segregation within establishments. In general firms cannot have explicit differences in wage policies towards different groups but they are allowed to have as many occupations, and wage categories, as they need and hence to concentrate some group to specific wage categories that is conceptually equivalent to set discriminatory wages, but harder to be proved. In this paper I skip this debate, and both sources of within-firm wage-differences are considered discrimination.

## 5.2 Worker's types

These results have been obtained using firm level data, hence I cannot include a broad set of variables to characterize workers. This may involve a problem meanwhile this method may be capturing different wage policies towards other groups that correlate with the migration status. To illustrate this point, let me assume that immigrants are not discriminated but women are, as the gender composition is significantly different between natives and immigrants[18], I would find that immigrants are receiving higher salaries than natives.

In Table 2, I show that the migration status is highly correlated with gender and job-qualification. To asses the importance of this issue I estimate the model controlling for gender and job qualification.

*Gender Composition:* To analyze if previous results are driven by gender

[18]See Table 2

composition differences, I estimate equation (5) but decomposing the workforce into four groups in terms of migration status and gender:

$$\bar{w}_{jt} = \alpha_j + \beta \bar{p}_{jt} + \gamma_{IM} P_{jt}^{IM} + \gamma_{NW} P_{jt}^{NW} + \gamma_{IW} P_{jt}^{IW} + \bar{\epsilon}_{jt}, \tag{6}$$

where $P_{jt}^{IM}$ is the proportion of immigrant-males in firm $j$ at time $t$, $P_{jt}^{NW}$ is the proportion of native-females and $P_{jt}^{IW}$ is the proportion of immigrant-females.

Table 5: Migration and Gender

| $\overline{w}_{jt}$ | Output | Value-Added |
|---|---|---|
| $\beta$ | 0.486 | 0,233 |
| | (0.006) | (0.005) |
| $\gamma_{IM}$ | -0.087 | -0,133 |
| | (0.053) | (0.058) |
| $\gamma_{NW}$ | -0.078 | -0,039 |
| | (0.032) | (0,035) |
| $\gamma_{IW}$ | -0.283 | -0,271 |
| | (0.071) | (0.077) |
| Fixed Effects | yes | yes |
| obs, | 24,943 | 24,943 |
| $R^2$ | 0.3829 | 0.3452 |

Note: Each column represents a single within-group linear regression using the panel of firms. Male-Natives are the reference group. Time Dummies are included in every specification. Standard errors in Parentheses. $R^2$ do not take into account the variation in firm fixed effects.

Results are presented in Table 5. We observe that male immigrants receive wages between 9 percent and 13 percent lower than male natives and female immigrants receive wages between 20 percent and 23 percent lower than female natives. Women are receiving lower wages that men. This difference ranges between 4 percent and 8 percent for natives but it is not always significant. This findings are consistent with the results presented in Bartolucci (2009), where estimating and a structural model to study gender wage gaps with the same data-set, women are not found to have significantly lower bargaining power in all the sectors.

*Skilled-Unskilled Composition*To understand if previous results are driven by job-qualification composition, I estimate equation (5) also decomposing the workforce into four groups in terms of migration status and job-qualification:

$$\bar{w}_{jt} = \alpha_j + \beta \bar{p}_{jt} + \gamma_{IS} P_{jt}^{IS} + \gamma_{NU} P_{jt}^{NU} + \gamma_{IU} P_{jt}^{IU} + \bar{\epsilon}_{jt}, \tag{7}$$

where $P_{jt}^{IS}$ is the proportion of skilled-immigrants in firm $j$ at time $t$, $P_{jt}^{NU}$ is the proportion of unskilled-natives and $P_{jt}^{IU}$ is the proportion of unskilled-immigrants. The reference group are the natives in skilled occupations. Results are presented in Table 6.

Table 6: Migration and Job-Qualification

| $\overline{w}_{jt}$ | Output | Value-Added |
|---|---|---|
| $\beta$ | 0.478 | 0,228 |
| | (0.006) | (0.005) |
| $\gamma_{IS}$ | 0.266 | 0,355 |
| | (0.106) | (0.115) |
| $\gamma_{NU}$ | 0,272 | 0,453 |
| | (0.032) | (0.035) |
| $\gamma_{IU}$ | -0,105 | -0,109 |
| | (0.046) | (0.050) |
| Fixed Effect | yes | yes |
| obs, | 24,943 | 24,943 |
| $R^2$ | 0.3829 | 0.316 |

Note: Each column represents a single within-group linear regression using the panel of firms. Skilled-Natives are the reference group. Time dummies are included in every specification. Standard errors in Parentheses. $R^2$ do not take into account the variation in firm fixed effects..

We observe that skilled-immigrants surprisingly receive salaries higher than skilled natives and that unskilled immigrants receive wages significantly lower than unskilled-natives. It is noteworthy that although unskilled workers receive wages 40 percent lower than skilled ones[19], once I control for productivity, native-unskilled workers have a positive wage differential, receiving wages that are 35 percent higher that native workers with equivalent productivity, in skilled occupations. This finding is also consistent with Bartolucci (2009), where un-

[19]The conditional mean of wages is 40.9 percent higher for skilled workers than for unskilled ones. See table (8) in the Appendix.

skilled workers are found to have higher bargaining power in every sector, for both, women and men.[20]

# 6 Testing implications of Discrimination models

Other interesting feature of this approach is that I can estimate (5) with firm specific $\gamma$. Hence I have a firm specific measure of wage discrimination against immigrants and I can look for evidence of different discrimination models, testing for their implications.

There are two mains branches in the theoretical discrimination literature: Taste Based Discrimination and Statistical Discrimination. These models emphasize two broad types of discrimination. The first is prejudice, which Gary Becker (1971) formalizes as a "taste" by at least some members of the majority group against interacting with members of the minority group. The second is statistical discrimination by employers in the presence of imperfect information about the skills or behavior of members of the minority group. Even though it is difficult to clearly distinguish between different theoretical hypotheses, some lessons can be drawn.

Becker and many others have discussed the fact that his model implies that discriminating employers earn lower profits than non-discriminators, since the non-discriminators will pay less for their labor by hiring discriminated workers. This implication may be directly tested in this framework. If taste-based discrimination were the true model we should observe a positive correlation between the $\gamma' s$ and firm profits.

The first papers discussing statistical discrimination were Phelps (1972) and Arrow (1973). The basic premise of this literature is that firms have limited information about the skills and turnover propensity of applicants, hence they have an incentive to use easily observable characteristics such as race or gender to "statistically discriminate" among workers if these characteristics are

[20] I cannot conclude that skilled workers are being discriminated because I am basically comparing different jobs.

correlated with performance[21]. There are two main branches in the statistical discrimination literature.

The first one investigate whether biased racial and gender stereotypes might be self confirming when the payoff for hard-to-observe worker investments depends on employer beliefs. Therefore an *a priori* unfounded belief about a group performance may be *a posteriori* confirmed. This issue, that was mainly addressed by Arrow (1973) and Coate and Loury (1993), is not analyzed in this paper because it should be captured controlling for productivity.

The second branch concerns the consequences of group differences in the precision of the information that employers have about individual productivity. It was mainly developed by Aigner and Cain (1977) with subsequent papers by Lundberg and Startz (1983) and Lundberg (1991). If this is true, and assuming that the signal that the firm extracts to infer productivity is more precise when the tenure increases, we should observe a positive correlation between tenure and the discrimination parameter $\gamma$.

Having firm specific discrimination parameters also allow us to have a better understanding of immigrant self selection into less discriminatory employers. If there is self-selection of immigrants into these employers, the expected value of $\gamma_j$ obtained here should be different from $\gamma$ obtained in the previous section.

To estimate firm specific discrimination parameters I follow Arellano and Bonhomme (2008). I estimate equation (5) in two simple steps: I firstly obtain the common parameters as follows: I regress the residual of firm specific regressions of the total wage bill and the variables with constants coefficients, on the proportion of immigrants and a constant term:

$$Q_j \bar{w}_{j,t} = (Q_j \overline{Z}_{j,t})' \delta + Q_j \epsilon_{jt},$$

where $Q_j = (I_{Tj} - X_j(X_j{}' X_j)^{-1} X_{ji})$, $X_j$ is a $2 \times T_j$ matrix with a column of ones that identifies the firm fixed effect and a column with the firm proportion of immigrants, $T_j$ the individual length of the panel (As my data-set is an

[21]Although it is illegal to make hiring, pay, or promotion decisions based on predictions about worker performance by gender or migration status, such behavior would be hard to detect in many circumstances.

unbalanced panel $T$ is firm-specific) and $Z_j$ is a matrix that contains those variables with constant coefficients, that are time dummies, output, value added or total number of workers depending on the specification.

Once I have estimated $\delta$, $\gamma_j$ is easily recovered:

$$\left( \begin{array}{c} \alpha_j \\ \hat{\gamma}_j \end{array} \right) = X_j({X_j}'X_j)^{-1}(\bar{w}_{j,t} - Z_j'\delta) = \gamma_j + X_j({X_j}'X_j)^{-1}\epsilon_{jt}.$$

Note that the estimated firm-specific fixed effect as the firm specific discrimination parameter are equal to the true parameters plus a term that is $O(1/T_j^{0,5})$. See Arellano and Bonhomme (2008) for more details.

Table 7: Wage Setting Equation with Random Coeficcients

| $\overline{w}_{jt}$ | (1) | | (2) | |
|---|---|---|---|---|
| $\beta_{\overline{output}}$ | 0,455 | | - | |
| | (0.009) | | - | |
| $\beta_{\overline{Value-Added}}$ | - | | 0.175 | |
| | - | | (0,006) | |
| SECOND STAGE REGRESSIONS | $\alpha_j$ | $\gamma_j$ | $\alpha_j$ | $\gamma_j$ |
| PROFITS | **-2.83** | 2.16 | **-2.88** | 2.15 |
| | **(0.99)** | (13.30) | **(1.02)** | (14.17) |
| TENURE OF IMMIGRANTS | 0.013 | **-0.211** | **0.019** | *-0.203* |
| | (0.008) | **(0.11)** | **(0.008)** | *(-0.117)* |
| PROPORTION OF IMMIGRANTS | 0.013 | 5.233 | -0.154 | *7.88* |
| | (0.290) | (3.90) | (0.299) | *(4.15)* |
| UNIONIZED WORKFORCE | **0.183** | 0.932 | -0.088 | 0.217 |
| | (0.099) | (1.33) | (0.102) | (1.42) |
| SINGLE ESTABLISHMENT | *-0.151* | 1,458 | **-0,324** | 1.71 |
| | *(0.095)* | (1.28) | **(0.098)** | (1.36) |
| CONSTANT | **1.66** | -1.06 | **4.84** | -0.73 |
| | (0.358) | (4.81) | **(0.369)** | (5.13) |
| OBSERVATIONS | 1,964 | | 1,964 | |

Note: Time & sector dummies included. Std. Errors in parentheses

Results are presented in Table 7. The second step is mostly imprecise. I find that the firm fixed effect is negatively correlated with the firm profit. As this fixed effect represent wages, given productivity this finding is not surprising. Although it has a coefficient marginally significant when using output as a measure of productivity, single establishment are in general lower payers. Firm with unionized workforce have higher fixed effect, this is not found when using value added.

I finally regress the firm specific discrimination parameter in the firm's mean profits and the firm's mean tenure of immigrants in order to look for evidence in favor of the most popular discrimination theories: Taste based discrimination and Statistical Discrimination. Profits have positive correlation with the discrimination parameter, that means that firms with higher profit discriminate less, as predicted by the taste-based discrimination literature. I find that the mean-tenure of immigrants in the firm is negatively and significantly associated with the discrimination parameter. I find that these firms with higher immigrant's tenure, where differences in the precision of the productivity signal should be less significant, are discriminating more. This may be understood as evidence against the statistical discrimination literature.

# 7 Conclusion

The Hellerstein and Neumark strategy has been found to be a very direct and popular method to detect wage discrimination using matched employer-employee data but it may produce biased estimates whenever there is not perfect competition in the labor market or when the discriminated group is segregated into good or bad firms. The purpose of this paper is to develop a test for wage discrimination that completes the Hellerstein and Neumark (1999) approach. To do this I propose a wage setting equation at the firm level that exploits changes in productivity and changes in the native-immigrant composition within firm across time to have identification of different wage policies toward those groups. Using Matching Employer-Employee data form Germany I show that this bias is empirically significant when analyzing discrimination against immigrants.

I find that Immigrants are suffering wage discrimination. Depending on which measure of productivity is used, discrimination ranges between 12.8 percent and 16.8 percent. This finding is surprising if we take into account that both the traditional approach and the Hellerstein and Neumark (1999) approach conclude that immigrants are not receiving significantly lower wages in Germany.

The elasticity of wages to productivity is significantly different from one and hence assuming wages equal to productivity may be dangerous. Although the reduced-form wage setting equation is very simple, it has an acceptable fit of the wage data and without controlling for firm fixed characteristics, I obtain similar results to those that would be obtained with employee-level data. When estimating by OLS, discrimination was found to be significantly lower, which gives evidence of positive segregation of immigrants into good firms. To understand the nature of this segregation I included different set of controls and I found that most of the segregation is accounted by differences in region, sector and firm size.

I find that female-immigrants are more discriminated than male ones. They receive wages between 20 and 23 percent lower than female natives while male immigrants receive wages between 9 and 13 percent lower than male natives. Unskilled immigrants receive salaries lower than unskilled natives but skilled immigrants receive higher wages than skilled natives.

I do not find significant evidence of immigrants moving to those less discriminatory firms neither significant evidence in favor of a taste-based discrimination model but I do find evidence against a statistical discrimination model.

# A Mean of log-productivity and log of mean productivity

As It is not possible recover the mean of log-productivity using output data, nor through a production function, I use the log of the mean productivity. In the specification shown in section 3, I need the mean log-productivity but with the production function estimations I can only have the log-mean of productivity.

To work with variables in logs have some desirable features: Firstly Labor economist have generally thought on wages as a log-normal variable. Secondly $\gamma$, the discrimination parameter, has a better interpretation as proportional premium.

Assuming that wages and productivity are log-normally distributed, it is possible to correct for the differences between the mean of log-productivity and the log-mean of productivity including a measure of the within-firm variance of productivity. Omitting this correction, estimates would be still correct if I assume that this within-firm variance remains constant across time and then these differences become part of the firm fixed effects. The main weakness of this approach is that I have to assume that the firm-specific variance of productivity do not change when composition changes.

For a more robust, and complex, solution to this problem, I can rearrange equation (1):

$$W_{ijt} = e^{\alpha_j}(P_{ijt})^{\beta}(\gamma)^{I_i}e^{\varepsilon_{ijt}},$$

and then, solving for $P_{ijt}$.

$$P_{ijt} = e^{(\alpha_j/\beta)}(W_{ijt})^{(1/\beta)}(\gamma/\beta)^{I_i}e^{(\varepsilon_{ijt}/\beta)}.$$

Aggregating within firm:

$$\bar{P}_{ijt} = e^{(\alpha_j/\beta)}\sum_{i\in j}(W_{ijt})^{(1/\beta)}(\gamma/\beta)^{I_i}e^{(\varepsilon_{ijt}/\beta)}.$$

This can not be estimated directly, because $W_{ijt}$ is endogenous to the model and it is then correlated with the error term $\varepsilon_{ijt}$. But the wage setting equation provides me proper instruments as $\bar{P}_{ijt}$ and $I_i$ to estimate it by GMM. This alternative is one of the main point in the research agenda.

# B Detecting Discrimination - Traditional Approach

In order to compare different strategies to detect wage discrimination. I perform the traditional approach using Mincer-type wage equations. As it can be seen in

Table 8, immigrants have positive wage differentials. Controlling for observed characteristics, they receive wages, on average, 7.2 percent higher than natives.

**Oaxaca-Blinder Decomposition**

Using results presented on Table 8, I perform a Oaxaca-Blinder decomposition which is to simply decompose the wage-gap between differences in observable and unobservable characteristics.

Oaxaca-Blinder decomposition's results are presented in Table 9. The counterfactual immigrants mean-wage has to be interpreted as the mean-wage that immigrants would have if they had the native's distribution of observable characteristics. Therefore the difference between the counterfactual immigrants mean-wage and the observed immigrants mean-wage is the portion of the gap that is due to differences in observable characteristics.

The portion of the unconditional wage-gap that is not accounted for observable characteristics has usually been interpreted as wage discrimination. In this case I would have that immigrants are not being discriminated. They are receiving wages 2 percent higher than similar natives.

# C Detecting Discrimination - Hellerstein and Neumark (1999) Approach

In order to compare my results with results found using the Hellerstein and Neumark (1999) approach to detect wage discrimination using my data-set, I estimate the firm production function and the firm wage equation. The production function is given by:

$$Ln(Y_{jt}) = const. + \alpha_k Ln(K_{jt}) + \alpha_l Ln(L^Q_{jt}), \tag{8}$$

using firm level data, where $Y_{jt}$ is the value added by firm $j$ at time $t$, $K_{jt}$ is depreciated capital[22] of firm $j$ at time $t$, and $L^Q_{jt}$ is the quality adjusted labor

[22]The survey gives information about investment made to replace depreciated capital. Assuming that a constant fraction ($d$) of capital depreciates by unit of time: $K^d_{jt} = d \times K_{jt} \Rightarrow \log(K^d_{jt}) = \log(d) + \log(K_{jt})$. Therefore $\alpha_k \log(d)$ goes to the constant term.

Table 8: Mincer Wage Equations - Censored-Normal Regression. Maximum Likelihood Estimates

| | General | Natives | Immigrants |
|---|---|---|---|
| Sex | -0.185 | -0.186 | -0.150 |
| | (0.0004) | (0.0003) | (0.0009) |
| Immigrant | 0.072 | - | - |
| | (0.0004) | - | - |
| Age | 0.061 | 0.065 | 0.035 |
| | (0.0001) | (0.0001) | (0.0003) |
| Primary Education | 0.237 | 0.241 | 0.204 |
| | (0.0008) | (0.0004) | (0.0008) |
| College | -0.246 | -0.246 | -0.202 |
| (incomplete) | (0.0009) | (0.0009) | (0.0025) |
| Technical College | 0.370 | 0.376 | 0.314 |
| (completed) | (0.0007) | (0.0007) | (0.0026) |
| College | 0.583 | 0.588 | 0.516 |
| | (0.0008) | (0.0008) | (0.0033) |
| University Degree | 0.709 | 0.716 | 0.648 |
| | (0.0007) | (0.0007) | (0.0023) |
| Tenure | 0.020 | 0.020 | 0.014 |
| | (0.0001) | (0.0001) | (0.0002) |
| Experience | 0.026 | 0.025 | 0.033 |
| | (0.0001) | (0.0001) | (0.0002) |
| Skilled | 0.407 | 0.411 | 0.357 |
| | (0.0010) | (0.0010) | (0.0055) |
| Part-time jobs | -0.696 | -0.703 | -0.616 |
| | (0.0004) | (0.0004) | (0.0013) |
| Constant | 2.381 | 2.319 | 2.894 |
| | (0.0019) | (0.0020) | (0.0031) |
| Pseudo R2 | 46.5% | 50.8% | 46.3% |
| Observations | 13,017,732 | 11,832,370 | 1,185,362 |

Note: Each column represents a single Maximum-Likelihhod linear regression using the panel of workers. Standard errors are given in parentheses. Native-men with no formal education in low-qualification occupations are the reference group. Time and Sector Dummies included.

Table 9: Oaxaca-Blinder Decomposition

| (a) Observed Natives Mean Daily Wage | (b) Observed Immigrants Mean Daily Wage | (c) Counterfactual Immigrants Mean Daily Wage |
|---|---|---|
| 109.0 € | 94.7 € | 111.2 € |
| Total W-Gap ((b)-(a))/(a) | Explained W-Gap=((b)-(c))/(a) | Unexplained W-Gap=((c)-(a))/(a)) |
| -13.1% | -15.1% | 2.0% |

input.

$$L^Q = \{L_{jt}^{mns} + \gamma_w L_{jt}^{wns} + \gamma_i L_{jt}^{mis} + \gamma_u L_{jt}^{mnu} + \gamma_i \gamma_w L_{jt}^{wis} + \gamma_w \gamma_u L_{jt}^{wnu} + \gamma_i \gamma_u L_{jt}^{miu} + \gamma_w \gamma_i \gamma_u L_{jt}^{wiu}\}$$

where $L_{jt}^{mns}$ is the number of male, native and skilled workers, $L_{jt}^{wns}$ is the number of female, native and skilled workers, $L_{jt}^{mis}$ is the number of male, immigrant and skilled workers, $L_{jt}^{mnu}$ is the number of male native and unskilled workers, $L_{jt}^{wis}$ is the number of female, immigrants and skilled workers, $L_{jt}^{wnu}$ is the number of female, native and unskilled workers, $L_{jt}^{miu}$ is the number of male, immigrant and unskilled workers, and $L_{jt}^{wiu}$ is the number of female, immigrants and unskilled workers in firm $j$ at time $t$.

The wage equation is given by:

$$Ln(W_{jt}) = const. + \kappa Ln(L_{jt}^Q), \tag{9}$$

where $W_{jt}$ is the total wage bill paid by firm $j$ at time $t$. In Table 10, I report the results from the estimations of the production function and wage equations using the total wages and salaries reported in the LIAB as paid by the establishment between 1996 and 2004. In column (1) I present parameter estimated from equation (8) by non-linear least squared, In column (2) I report parameter estimated from equation (9) by non-linear least squared, and in column (3) I

Table 10: Hellerstein et al Approach

| | (1) Output | (2) Wage | (3) p-value (1)-(2) |
|---|---|---|---|
| *Immigrants* | 0.99 | 0.98 | 54.2% |
| | (0.09) | (0.03) | - |
| *Women* | 0.34 | 0.38 | 3.7% |
| | (0.02) | (0.01) | - |
| *Unskilled* | 0.33 | 0.47 | 0.0% |
| | (0.01) | (0.01) | - |
| $\alpha_k$ | 0.16 | - | - |
| | (0.01) | - | - |
| $\alpha_l$ | 0.89 | - | - |
| | (0.01) | - | - |
| $\kappa$ | - | 1.05 | - |
| | - | (0.002) | - |
| *constant* | 9.29 | 7.47 | - |
| | (0.62) | (0.02) | - |
| $R^2$ | 0.82 | 0.92 | - |
| *Observations* | 12,259 | 17,224 | - |

Note: Columns (1) and (2) represent single non-linear regressions using the panel of firms. Male–Skilled-Natives are the reference group. Time Dummies are included in every specification. Standard errors are given in Parentheses.

report p-values from test of equality between parameter reported in column (1) and (2).

Looking first at the production function estimates in column (1), I find that the coefficient for immigrants indicates that foreign workers are somewhat equally productive than natives with an estimate of $\gamma_i$ that is 0.99, not significantly different form one. I also find that productivity of women is surprisingly low and that workers in unskilled occupation produced two thirds less than workers in skilled occupations. Looking at the wage equation I find similar patterns in terms of immigrants and women. Workers in unskilled occupation receive salaries 53 percent lower than workers in skilled ones.

Column (3) of Table 10 reports the p-values of tests of the equality of the coefficients from the production function (column (1)) and the wage equation (column (2)). The results for immigrants are not conclusive, productivity gap and the wage gap are not significantly different from zero. The results for women show that the productivity gap between men and women exceeds the wage gap. The wedge between relative wages and relative productivity is -0.04 (0.34 - 0.038), and the p-value of the test of the equality of relative wages and relative productivity for women is 3.7 percent. This approach would conclude that men are being discriminated.

# References


[1] Abowd, J., and Kramarz F. & Margolis D. "High wages workers and high wages firms" Econometrica, vol. 67, No 2 (march 1999), pp: 251-333

[2] Abowd, J., and Kramarz F. 1999. "The Analysis of Labor Markets Using Matched Employer-Employee Data". Orley C. Ashenfelter and David Card, eds., Handbook of Labor Economics, Vol. 3B (Amsterdam: Elsevier Science Publishers), pp. 2629-710.

[3] Aeberhardt R. & Pouget J. "National Origin Wage Differentials in France" Working paper, March 2006.

[4] Alda H., Bender S. & Gartner H. "The linked employer-employee dataset of the IAB (LIAB)", IAB Discussion Paper No. 6/2005, 2005.

[5] Altonji J. and Blank R. "Race and Gender in the Labor Market" in Orley Ashenfelter and David Card, eds., Handbook of Labor Economics. Vol. 3, Amsterdam: North Holland, 1999.

[6] Arellano, M. and S.R. Bond (1991), "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations", Review of Economic Studies, 58, 277–297.

[7] Arellano, M. and S. Bonhomme (2008) "Identifying Distributional characteristics in Random Coefficient Panel Data Models" Working Paper.

[8] Arellano, M., and O. Bover (1995), “Another Look at the Instrumental-Variable Estimation of Error-Components Models,” Journal of Econometrics, 68, 29–52.

[9] Arrow, K. (1973), ”The theory of discrimination”, in: O.A. Ashenfelter and A. Rees, eds., Discrimination in labor markets (Princeton University Press, Princeton, NJ) pp. 3-33.

[10] Aydemir, A. and M. Skuterud (2008). ”The immigrant wage differential within and across establishments”. Industrial and Labor Relations Review 61(3), 334-352.

[11] Aigner, DJ. and G. Cain (1977), ”Statistical theories of discrimination in labor markets”, Industrial and Labor Relations Review, 30: 175-187.

[12] Bartolucci, C (2009) ”Gender Wage Gaps Reconsidered: A Structural Approach using Matched Employer-Employee Data” Working Paper

[13] Becket, G. (1971) The economics of discrimination, 2nd edition (The University of Chicago Press, Chicago).

[14] Bratsberg, B. and J. Ragan (2002), “The Impact of Host-Country Schooling on Earnings” Journal of Human Resources, 37(1), 63-105.

[15] Borjas, G. (1999) ”The Economic Analysis of Immigration” O. Aschenfelter & D. Card (ed.), ”Handbook of Labor Economics”, Elsevier edition 1, Volumen 3, Number 3.

[16] Carnevale, A., R. Fry and B. Lowell (2001), “Understanding, Speaking, Reading, Writing, and Earnings in the Immigrant Labor Market” American Economic Review, 91(2), 159-64.

[17] Chiswick, B. and W. Miller (1995), “The Endogeneity Between Language and Earnings: International Analyses” Journal of Labor Economics, 13(2), 246-87.

[18] Coate, S. and G. Loury (1993), "Antidiscrimination enlbrcement and the problem of patronization". American Economic Review 83(2): 92 98.

[19] Dustmann, C. and A. van Soest (2002), "Language and the Earnings of Immigrants" Industrial and Labor Relations Review, 55(3): 473-92.

[20] Ferber, M. and c. Green (1982) "Traditional or Reverse sex discrimination? A Case Study of a large public university" Industrial and Labor Relations Review, 35(4), 550-564.

[21] Friedberg, R. (2000), "You Can't Take It With You? Immigrant Assimilation and the Portability of Human Capital," Journal of Labor Economics, 18(2), 221-51.

[22] Heckman J. "Detecting discrimination" Journal of Economic Perspectives, 12(2), 1998, pp 101-116.

[23] Hellerstein J. & Neumark D. "Ethnicity, Language, and Workplace Segregation: Evidence from a New Matched Employer-Employee Data Set", Annales d'Economie et Statistique, 2003, no. 71-72, pp. 19-78.

[24] Hellerstein J., Neumark D. & Troske K."Wages, Productivity, and Worker Characteristics: Evidence from Plant-Level Production Function and Wage Equations" Journal of Labor Economics, 17(3), June 1999, pp. 409-446.

[25] Hellerstein J. & Neumark D. "Sex, Wages, and Productivity: An Empirical Analysis of Israeli Firm-Level Data," International Economic Review, 40(1), February 1999, pp. 95-123.

[26] Hellerstein J. & Neumark D."Using Matched Employer-Employee Data to Study Labor Market Discrimination" IZA Discussion Paper No. 1555, April 2005.

[27] Lundberg, S. (1991), "The enforcement of equal opportunity laws under imperfect information: affirmative action and alternatives", Quarterly Journal of Economics 106 (1): 309-326.

[28] Lundberg, S. J, and R. Startz (1983), ”Private discrimination and social intervention in competive labor markets”, American Economic Review 73: 340-347.

[29] Kawaguchi, D. (2007) ”A Market Test of Sex Discrimination: Evidence from Japanese Firm-Level Panel Data”, forthcoming in International Journal of Industrial Organization .

[30] Milgorm E., T. Pettersen and V. Snartland (2006) ”Are Female Workers Less Productive than Male Workers”, Institute for Research on Labor and Employment Working Paper 139-06.

[31] Rudolph, H. (1994) “Dynamics of Immigration in a Non-immigrant Country: Germany. In Fassman, He. and Münz, R. (eds.) European Migration in the Late Twentieth Century. Laxenburg, Edward Elgar, pp. 113-127.