POLITECNICO DI TORINO
Repository ISTITUZIONALE

L'Union Fait la Force? Evidence for Wage Discrimination in Firms with High Diversity

Original
L'Union Fait la Force? Evidence for Wage Discrimination in Firms with High Diversity / Grinza, Elena; Kampelmann, Stephan; Rycx, François. - In: THE JOURNAL OF ECONOMIC INEQUALITY. - ISSN 1569-1721. - 18:2(2020), pp. 181-211. [10.1007/s10888-019-09433-7]

Availability:
This version is available at: 11583/2790076 since: 2020-06-09T19:26:08Z

Publisher:
Springer Nature

Published
DOI:10.1007/s10888-019-09433-7

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

12 October 2024

L'union fait la force?
Evidence for wage discrimination in firms with high diversity

Elena Grinza[a], Stephan Kampelmann[b], and François Rycx[b,c,d,e,f]

[a] University of Milan [b] Université Libre de Bruxelles (CEB and DULBEA)

[c] IZA [d] IRES [e] GLO [f] humanOrg

Abstract

Measuring the economic impact of coworkers from different countries of origin sparked intense scrutiny in labor economics, albeit with an uncomfortable methodological limitation. Most attempts have involved metrics that eliminate most of the socially and economically relevant heterogeneity among different countries of origin, salient dimension of diversity and critical determinant of labor market outcomes of migrants. The typical examples of such metrics are diversity indicators that divide the firm's workforce into binary categories such as blacks and whites, foreigners and natives, and non-Europeans and Europeans. We propose an entirely novel approach that constructs a firm-level aggregate measure of diversity that explicitly takes into account differences in socio-economic conditions of migrants' countries of origin. To do so, we use the United Nations Development Programme's Human Development Index (HDI), a standard harmonized measure of cross-country variations in levels of socio-economic development that is available for virtually all the countries in the world. By resorting to rich matched employer-employee panel data for Belgium, we use this new aggregate measure of firm-level diversity to estimate state-of-the-art firm-level wage equations, which control for a wide range of observable and time-invariant unobservable factors, including variations in labor productivity between firms and within firms over time. The results suggest that the majority of firms do not discriminate against foreigners. However, we find that firms with high diversity broadly discriminate against them. The wage discrimination in high-diversity firms could be alleviated through a stronger presence of collective bargaining and efforts to de-cluster foreigners from low-HDI countries in these firms.

Keywords: discrimination, diversity, matched employer-employee panel data, wages, workers' origin.

JEL classifications: J15, J16, J24, J31, J7.

Corresponding author:
Elena Grinza
University of Milan
Via Conservatorio 7
20122 Milan, Italy
E-mail: elena.grinza@unimi.it

We are most grateful to Statistics Belgium for giving us access to the data. We thank Maarten Goos, Sem Vandekerckhove, Guy Van Gyes, and the members of the scientific committee of the IPSWICH project for constructive comments on earlier versions of this paper. We also thank the Editor and two anonymous reviewers for helpful comments and suggestions. Funding for this research was provided by the Belgian Science Policy Office (BELSPO). The usual disclaimer applies.

1. Introduction

The recent surge in immigration from Africa and the Middle East into the European Union, especially Germany and Sweden, has renewed controversies about the impact of a more diverse workforce on labor market outcomes, such as working conditions and wages. In Germany, a country that experienced a net increase of more than 1.1 million foreigners in 2015 (Bundesamt für Migration und Flüchtlinge, 2016), this controversy led to renewed debates on the recently introduced federal minimum wage. To improve the employability of newly arrived workers with lower skills, the federal government proposed in early 2017 to exempt asylum seekers from minimum wage rules. This was heavily criticized by trade union representatives, which argued that asylum seekers could be exploited as "cheap labor" by German employers since they are generally less informed about worker rights and minimum standards.[1]

Whereas public debates on labor market impacts of a more diverse workforce tend to focus on the societal or macro level, a series of econometric studies highlight the importance of examining the wage effects of diversity at the firm or micro level. For the USA, for instance, Carrington and Troske (1998) show that ethnic diversity within manufacturing firms has a significant effect on wages, as the remuneration of black workers decreases and that of white workers increases if they have more black coworkers. For Swedish firms, Åslund and Skans (2010) show that ethnic groups with more foreign coworkers earn less than those with more native colleagues. Their results also suggest that the more an individual is exposed to foreigners within his or her workplace, the lower will be the wage. A similar picture emerges from data on German firms in four metropolitan areas analyzed by Dustmann et al. (2015), which find that foreign workers with a higher share of foreign coworkers suffer from a wage penalty.

A drawback of the empirical literature on firm-level diversity is that diversity is generally apprehended in terms of binary categories, such as blacks *versus* whites, foreigners *versus* natives, and non-Europeans *versus* Europeans (all the empirical studies mentioned above use binary categories in their regressions). While this is partly due to methodological constraints (most estimators require a limited number of explanatory variables to allow for statistical identification), such binary categories eliminate most of the socially and economically relevant differences among migrants' countries of origin. This is problematic because differences in levels of socio-economic development in migrants' home countries are salient dimensions of diversity and are well-known crucial determinants of migrants' labor market outcomes, including the degree of bargaining power with employers. Therefore, in reality, diversity is a phenomenon with more facets than dichotomic categories can embed. Thus, rather than considering foreigners as a homogeneous group and opposing them to natives, one would like to obtain a more nuanced picture that distinguishes between, say, foreigners from neighboring countries with relatively similar levels of schooling and economic development, from those that come from countries that radically differ from the host country. In the context of Belgium, for example, the impact of diversity likely differs

[1] Süddeutsche Zeitung, January 2, 2017, Bundesregierung plant Ausnahmen für Flüchtlinge beim Mindestlohn. Retrieved on December 14, 2017, from http://www.sueddeutsche.de/wirtschaft/migranten-bundesregierung-plant-ausnahmen-fuer-fluechtlinge-beim-mindestlohn-1.3317242.

between a firm that employs workers from France or the Netherlands compared to a firm with workers from the Democratic Republic of the Congo or Syria.

In this paper, we develop and apply to rich matched employer-employee panel data for Belgium an innovative approach to measure firm-level diversity, which goes beyond binary classifications and explicitly takes into account heterogeneity in levels of socio-economic development of migrants' countries of origin. Our approach consists of four steps. First, we use the United Nations Development Programme's Human Development Index (HDI) of employees' countries of origin to convert qualitative variables related to diversity (i.e., nationality at birth) into a normalized quantitative variable embedding the level of socio-economic development of the worker's country of origin. For example, a worker of Belgian origin is assigned the value 0.88 (the HDI of Belgium in 2010), a worker of Syrian origin the value 0.64 (the HDI of Syria in 2010), and so forth. In our procedure, we use the HDI to measure socio-economic conditions of migrants' countries of origin because of its long-standing reputation as a standard measure of cross-country differences in social and economic development. Also, the HDI has the advantages of being based on a harmonized and widely accepted methodology and producing a fine-grained picture of socio-economic differences between virtually all the countries in the world. In the second step, we aggregate the individual-level HDIs of all the workers within the same firm into a single measure of firm-level diversity. Our aggregate measure is based on average Euclidean distances that we calculate for all the pair-wise combinations within the workforce of the firm. For instance, we compute the absolute difference between the Belgian and Syrian workers ($\sqrt{(0.88-0.64)^2}=0.24$) and also the differences between all the other possible couples within the firm. Then, we take the average of all these pair-wise differences to obtain a firm-level indicator of diversity. In the third step, we use this aggregate measure of firm-level diversity to explain variations in firm-level wages. Our econometric approach is based on the empirical framework proposed by Bartolucci (2014), which, in turn, builds on Hellerstein et al. (1999). This method has the advantage of addressing a variety of factors that potentially influence the link between diversity and wages, such as productivity, sorting/segregation, and discrimination. In the final step, we examine the moderating role of firm-level collective bargaining and different levels of diversity in the firm.

The rest of the paper is structured as follows. The next section reviews theoretical arguments for the impact of firm-level diversity on wages. Section 3 presents the data set and descriptive statistics. Section 4 briefly summarizes the empirical framework and the estimators that we use to identify the impact of diversity on wages. Section 5 presents and discusses the results and robustness tests. The final section concludes.

2. Diversity and wages

There is a vast literature on the potential causal links between diversity in terms of countries of origin and wages. In this section, we summarize the theoretical arguments for such links by grouping them into three types of effects. We then discuss two moderating factors: firm-level collective bargaining and different levels of diversity in the firm.

2.1. Arguments for a causal link between diversity and wages

Most of the theoretical explanations concerning the impact of diversity on wages can be grouped into three types of effects: productivity, sorting/segregation, and bargaining effects.

In the context of studies on diversity, productivity effects reflect the human-capital perspective on wage setting, according to which differences in wages among different workers are related to differences in productivity (see Garnero et al., 2014a for an overview of these arguments). Such effects are hypothesized and documented both at the individual and collective level. Individual-level productivity differences among individuals from different countries of origin are related to language abilities (Dustmann and Van Soest, 2002; Hellerstein and Neumark, 2003), literacy skills (Ferrer et al., 2006), quality and transferability of foreign education and training (Aeberhardt and Pouget, 2010; Gratsberg and Ragan, 2002), and less valuable labor market experience obtained abroad (Friedberg, 2000). Collective-level productivity effects are associated with positive or negative spillovers that arise when workers with different languages or customs work together (Böheim et al., 2012; Ottaviano and Peri, 2012).

The second set of explanations for wage differences among workers of different origin relates to non-random sorting/segregation effects. Many empirical studies find evidence of non-random sorting/segregation related to ethnicity (Aeberhardt and Pouget, 2010; Åslund and Skans, 2010; Aydemir and Skuterud, 2008; Bayard et al., 1999; Carrington and Troske, 1998; Elliott and Lindley, 2008; Hellerstein and Neumark, 2008; Peri and Sparber, 2009). The most common categories in which non-random sorting/segregation occurs include job types, tasks, occupational nomenclatures, firms with different technologies or capital endowments, and sectors of activity. As summarized by Kampelmann and Rycx (2016), whereas productivity-related effects refer to differences between natives and foreigners within the same category (e.g., unequal productivity within the same occupation), sorting/segregation effects point to differences in the distribution of natives and foreigners across different categories that each capture a certain level of productivity (e.g., overrepresentation of foreigners in occupations characterized by lower productivity).

While much of the literature on diversity focuses on productivity and sorting/segregation effects, in this paper, we focus on a third category that received less attention but can be much relevant, namely, bargaining effects. We argue that bargaining effects can be interpreted as the correlation between diversity and the firm's average wage level that remains unexplained after controlling for productivity and sorting/segregation effects. Once the other effects are accounted for, additional variations in the *ceteris paribus* wage level arguably reflect the result of wage negotiation (bargaining) within the firm.

Some arguments on bargaining effects suggest that diversity could lead to increases above the *ceteris paribus* wage, whereas other mechanisms predict wage depression (i.e., discrimination).

The former set of arguments essentially indicate that a more diverse workforce is in a better position to bargain a larger share of rents that are generated by the firm. Such rent-extraction could be due to a stronger bargaining position of foreigners, particularly high-skilled professionals, which generally have more access to outside options in different countries (e.g., because they know foreign languages better or have a more international

experience). Employers could be forced to share rents with such workers in order to retain them (or attract them in their firms). Moreover, foreigners likely are less anchored in a specific cultural context and could, therefore, more credibly threaten the employer to leave the company in search of better work offers in other countries. Another bargaining-related argument for a positive relationship between diversity and wages concerns the collective element of the wage-setting process. A stronger presence of foreigners within a firm – which often goes hand in hand with higher diversity – could modify the bargaining power to the disadvantage of employers, for example, making it more difficult for them to pay discriminatory wages. Over certain degrees of firm-level diversity, belonging to a minority could actually mean belonging to a majority in the firm, thereby potentially reversing the conventionally assumed higher bargaining power of the "traditional" majority of native workers.

However, there are also strong theoretical arguments in favor of a negative relationship between diversity and the firm's *ceteris paribus* wage level. This means an underpayment of a more diverse workforce, all the other things (i.e., productivity and sorting/segregation) being equal, which is the central element of the definition of discrimination by Heckman (1998). Classical types of discrimination are proposed by Phelps (1972) and Arrow (1973): "statistical discrimination" and "preference-based discrimination". The first refers to the effect of negative stereotypes or a general lack of information by employers on the productivity of foreigners. The second refers instead to a situation in which the tastes of employers (or native employees or customers) translate into a lower demand and (consequently) lower wages for foreign workers. Higher levels of diversity also mean higher numbers of ethnicities in the firm. A less cohesive workforce in terms of language and cultural background could weaken its collective bargaining position *vis-à-vis* employers, for instance, when such diversity makes it more problematic to establish trust, cohesion, and common bargaining positions within the workforce. If, as proclaimed by the Belgian motto, *l'union fait la force* (unity makes strength), then, diversity could weaken the bargaining power of workers and their protection against wage discrimination. Another channel through which diversity could negatively affect wages is represented by social networks, which, as shown by Seidel et al. (2000), tend to work against racial minorities in salary negotiations. In practice, this is because minorities have fewer social ties to the organizations with which they negotiate wages.

Our empirical approach, based on Bartolucci's (2014) empirical framework and matched employer-employee panel data, allows us to account for both productivity and sorting/segregation effects. We will then use the resulting *ceteris paribus* wage to provide evidence on the theoretical arguments above, that is, whether the effect of diversity on wages due to bargaining effects (i.e., discrimination) is positive or negative.

2.2. The moderating role of collective bargaining institutions

Certain institutional factors likely influence the relationship between diversity and the *ceteris paribus* wage. In particular, formal collective bargaining institutions could diminish wage discrimination against minority groups (Freeman, 1980; Plasman et al., 2007). In many countries, including Belgium, trade unions present themselves as advocates of fair pay

for vulnerable groups (Dell'Aringa and Lucifora, 1994; Pillinger, 2014). However, the opposite view also emerges in the literature. Trade unions could prioritize the exclusive interests of native workers, especially in situations and periods in which foreigners are not affiliated to trade union organizations, as it was in the 1960s and 1970s in West Germany, where trade unions openly defended the interests of German workers against wage demands of the foreign *Gastarbeiters* (Kampelmann, 2011). Research on more recent years shows that the question of whether unions are inclusive or exclusive towards workers with a migration background has not been settled yet. For the case of New Zealand, for instance, Harcourt et al. (2008) argue that, contrary to their rhetoric, trade unions are relatively unsuccessful in combating discrimination against foreigners and ethnic minorities.

At the micro level, examining how the impact of diversity on wages varies depending on the degree of collective bargaining specific to the firm can shed more light on the role of trade unions. Our data allow us to investigate this. Thanks to detailed information on the presence of firm-level collective bargaining, we can split the sample between i) firms that are covered only by national- and sectoral-level bargaining and ii) firms that have an additional round of bargaining at the firm level. If collective bargaining prioritizes inclusiveness over the exclusive interests of native workers, we expect that wage discrimination is attenuated in firms with firm-level bargaining (Dell'Aringa and Lucifora, 1994; Plasman et al., 2007).

2.3. No group dynamics without a group?

It is crucial to underline that the bargaining effects reviewed above only operate through mechanisms at the level of *groups*. This implies that it is unlikely that any tangible consequences on wage-setting outcomes manifest themselves if the presence of foreign workers in the firm is restricted to a few individuals. Collective wage bargaining mechanisms can only occur if foreigners are a recognizable group in the firm, with specific characteristics that set them collectively apart from native employees. In other words, a single foreigner in a large company with an otherwise entirely native workforce is unlikely to affect how salaries are negotiated within the company. Instead, if foreigners form a recognizable group within the company, the bargaining process might change. For instance, their presence could reflect the employer's "preference" to pay discriminatory wages to foreigners with weaker bargaining power (as hypothesized by Bloomekatz, 2007). In some cases, the discrimination against foreigners can also be incorporated into collectively negotiated pay scales, for instance, when job categories in which foreigners are overrepresented are specifically designed to allow for wage discrimination (Kampelmann, 2011).

To investigate this, we test whether the effect of diversity on *ceteris paribus* wages differs in firms with low, high, and very high levels of diversity. Given that the mechanisms related to bargaining effects rely on recognizable groups, we expect that the wage effects due to bargaining are stronger in firms with higher levels of diversity and small or inexistent in those with relatively fewer foreigners.

3. Data and descriptive statistics

3.1. Data set

The empirical analysis is based on two large data sets spanning the period 1999-2010. The first is the Structure of Earnings Survey (SES), which covers firms operating in Belgium that employ at least 10 workers and operate in sectors within sections C to K of the NACE nomenclature (Rev. 1). The SES data set contains rich information relating to firm characteristics (e.g., sector of activity, number of workers, level of collective wage bargaining) and to characteristics of the workers employed in the firm (e.g., age, education, gross earnings, paid hours, gender, occupation).[2] These features make SES a matched employer-employee panel data set. The SES data set does not provide financial information and, therefore, it is merged with another firm-level survey, the Structure of Business Survey (SBS). The SBS data set provides information on a variety of financial variables, such as firm-level value added and gross operating surplus per hour. We call the resulting matched employer-employee panel data set "SES-SBS". All the variables in the SES-SBS data set are provided by the firm's management and, therefore, more reliable compared to self-reported information typical of individual- or household-level surveys. Also, the SES does not provide information on workers' origin. This information (i.e., nationality at birth, country of birth, and current nationality) is taken from the National Register (NR) and merged with the SES-SBS data.

The SES, SBS, and NR data sets are all collected and managed by Statistics Belgium, the national institute of statistics in Belgium. The relevant merges are performed by Statistics Belgium, too. In particular, the match between SES and SBS is carried out using the firms' social security numbers.

The earnings measure in SES corresponds to total gross wages, including premia for overtime, weekend or night work, performance bonuses, commissions, and other premia. Work hours represent total actual remunerated hours in the reference period (including paid overtime hours). The firm's value added per hour in SBS is measured at factor costs and based on the total number of hours effectively worked by the firm's employees.

The coverage of SBS differs from that of SES in that it does not include the whole financial sector (NACE J) but only two of its sub-sectors, namely, "Other Financial Intermediation" (NACE 652) and "Activities Auxiliary to Financial Intermediation" (NACE 67). The capital stock of each firm is calculated with the permanent inventory method using annual firm-level information on gross fixed capital formation.

[2] The SES data set is the result of sophisticated stratified sampling design. The stratification criteria refer to the region (NUTS groups), sector of economic activity (NACE groups), and firm size. The sample size in each stratum depends on firm size. Sampling percentages of firms are equal to 10, 50, and 100% when the number of workers is, respectively, lower than 50, between 50 and 99, and above 100. Within the firm, sampling percentages of employees also depend on firm size. Sampling percentages of employees reach 100, 50, 25, 14.3, and 10% when the number of workers is, respectively, lower than 20, between 20 and 50, between 50 and 99, between 100 and 199, and between 200 and 299. Firms that employ 300 workers or more have to report information for a specific number of employees. This number ranges between 30 (for firms with between 300 and 349 workers) and 200 (for firms with 12,000 workers or more). To ensure that firms report information on a representative sample of their workers, they are asked to follow a specific procedure. For a more detailed discussion, see Demunter (2000).

Three filters are applied to the original data set. First, we delete from the sample firms that are publicly controlled and/or operate in predominantly public sectors. The rationale of this filter derives from the standard productivity theory and its requirement that prices have to be economically meaningful. All the regressions are thus applied to privately controlled firms only.[3] Second, in order to ensure that firm-level averages are based on a sufficient number of observations, we filter out firms that provide information on less than 10 employees.[4] Finally, to pursue panel estimation, we restrict the sample to firms observed in at least two consecutive years. This restriction leads to the overrepresentation of medium- and large-sized firms. This is because sampling percentages of firms – and, therefore, the probability of being observed over consecutive years – in the SES data set increase with firm size (see Footnote 2).

The final sample consists of an unbalanced panel of 9,430 firms and 555,963 individuals, yielding 30,355 firm-year observations over 12 years (1999-2010). It is representative of medium-sized and large firms employing at least 10 employees within sections C to K of the NACE (Rev. 1) nomenclature, except for large parts of the financial sector (NACE J) and almost the entire electricity, gas, and water supply industry (NACE E).[5]

3.2. Diversity in terms of human development

As mentioned earlier, we develop an innovative approach to measure workforce diversity in terms of country of origin. Our approach departs from basic binary categories, common in the literature, and develops a firm-level measure of diversity that explicitly takes into account the heterogeneity in levels of socio-economic development of migrants' home countries. This is important because differences in socio-economic (and cultural) backgrounds are crucial dimensions of labor diversity, which significantly impact on migrants' labor market outcomes (e.g., see Adsera and Chiswick, 2007). We use the United Nations Development Programme's Human Development Index (HDI) to convert the qualitative data on workers' origin (i.e., nationality at birth) into a normalized quantitative variable embedding the level of socio-economic development of workers' countries of origin. We use the HDI because of its long-standing reputation as a standard measure of cross-country differences in levels of social and economic conditions. Also, the HDI has the advantage of using a harmonized and widely accepted methodology, which produces a fine-grained picture of socio-economic differences between virtually all the countries in the world. As stated by the United Nations' Human Development Report 2016, the HDI is the geometric mean of three normalized indices, which cover "key dimensions of human development". These dimensions are i) "a long and healthy life", measured through life

[3] More precisely, we eliminate firms for which public financial control exceeds 50%. This exclusion reduces the sample size by less than 2%.
[4] This selection is unlikely to affect the results as it leads to a minimal drop in the sample size.
[5] As highlighted in Table A.1 (see Appendix A), selection criteria reduce the size of the original SES-SBS-NR sample (which is representative of about two-thirds of the total private-sector employment in Belgium) from 49,944 to 30,355 firm-year observations. They also lead to the over- (under-) representation of sectors in which average firm size is bigger (smaller), for instance, the manufacturing industry (hotels and restaurants or wholesale and retail trade).

expectancy at birth; ii) "being knowledgeable", measured through expected years of schooling and average years of schooling; and iii) "a decent standard of living", measured through gross national income per capita in purchasing power parity (UNDP, 2017).[6]

We use the 2010 HDIs for the different national origins in our data set, and associate each worker with the level of human development that corresponds to his or her specific origin, which we operationalize as the individual's nationality at birth. For example, a worker of Belgian origin is assigned the value 0.88 (the HDI of Belgium in 2010), a worker of Syrian origin is assigned the value 0.64 (the HDI of Syria in 2010), and so on.[7]

The average HDI associated with workers in the sample is 0.87, with a standard deviation of 0.03. A closer examination of the different categories of workers reveals that the level of human development associated with their origin is not evenly distributed. For instance, on average, women are associated with a slightly higher HDI than men (0.87 compared to 0.86), which reflects the relatively more substantial proportion of male foreigners from low-HDI countries. Occupations are also differentiated with respect to the HDI of the individuals in them: on average, managers are associated with the highest HDI (0.880), followed by professionals (0.876), technicians (0.875), clerical occupations (0.874), machine operators (0.867), crafts (0.865), service occupations (0.863), and elementary occupations (0.844). We observe less differentiation among sectors of activity, except a markedly low average HDI associated with workers in accommodation and food service activities (0.83). Differences between the three Belgian regions reflect the relatively more significant presence of foreigners from low-HDI countries in the Brussels-Capital Region, in which the average HDI is 0.85 compared to 0.87 in Flanders and Wallonia.

3.3. Firm-level diversity and descriptive statistics

To move from individual-level observations to a firm-level indicator of diversity, HDIs within each firm's workforce need to be aggregated into a single measure of firm-level diversity. The aggregate measure of diversity used in this paper is based on the Average Euclidean Distance (AED), which we calculate for all the pair-wise combinations within the firm's workforce. To continue with the example above, we compute the absolute difference between the Belgian and Syrian workers ($\sqrt{(0.88-0.64)^2}=0.24$) and also the differences between all the other possible couples of workers within the firm; we then take the average of all these pair-wise differences.

Descriptive statistics for the 30,355 firm-year observations in our sample are shown in Table 1. The average AED is 0.03, with a standard deviation of 0.04. The AED is much

[6] For a detailed description of the HDI's underlying calculations, see the technical notes included in the United Nations' Human Development Report 2016: http://hdr.undp.org/sites/default/files/hdr2016_technical_notes_0.pdf (accessed on January 8, 2018).

[7] Given that the HDI is updated on an annual basis since 1990, one could alternatively associate the value of the HDI for a given country in a given year to each individual. For instance, an employee of Belgian origin could be assigned the value 0.87 (the HDI of Belgium in 2000) if he or she is observed in 2000, and the value 0.88 (the HDI of Belgium in 2010) if he or she is observed in 2010. However, this has the inconvenience of generating missing observations since the HDI covers virtually all the countries of the world only since 2010. Moreover, the focus of this paper lies on cross-country variations in human development rather than within-country variations over time, which is why we prefer to use the values of the HDI at a fixed point in time.

differentiated across regions, with firms based in Brussels being considerably more diverse than firms in Flanders and Wallonia (AEDs of 0.05 in Brussels and 0.02 in Flanders and Wallonia).[8]

[Insert Table 1 about here]

An effective way to explore the firm-level distribution of AEDs is by looking at the relationship between the AED and the proportion of native Belgians in each firm (Figure 1). The graph includes a histogram of the share of native Belgians (i.e., those born in Belgium and with Belgian nationality), whose relative frequencies are shown on the left-hand axis. As one can see, only a few firms have less than 80% of native Belgians.

[Insert Figure 1 about here]

The second information included in the figure is the relationship between the share of native Belgians, on the one hand, and the firm's AED, on the other hand. This relationship is captured by a fractional polynomial whose confidence interval is also depicted in the graph (the scale is given on the right-hand axis). The curve shows that firms with around 50% of foreigners are the most diverse. When the proportion of Belgian workers is above 50%, the relationship between the AED and the share of Belgian workers is negative: as the proportion of foreigners approaches 0, the diversity also goes to 0. Conversely, when the proportion of Belgian workers approaches 0, the diversity is high, though a bit lower than at the maximum, presumably because these firms employ at least some foreigners from the same origin (and therefore the same HDI). The left-hand tail of the distribution is, however, very rare in our sample. Thus, for the vast majority of observations, we observe a negative relationship between the AED and the proportion of Belgian workers in the firm.[9]

4. Estimation strategy

4.1. Wage-setting equations at the firm level

For several decades, the contributions by Oaxaca (1973) and Blinder (1973) provided the most commonly used tools for studying wage discrimination against foreigners. However, in recent years, the standard version of the Oaxaca-Blinder decomposition has attracted increasingly sharp criticism. First, by definition, the resulting residual gap conveys any unobserved intrinsic productivity differences or sorting into discrimination. Second, this method only controls for differences in occupational or sectoral composition between natives and foreigners, rather than also explaining the process of sorting into groups with

[8] Descriptive statistics, reported in Table A.2 (see Appendix A), show that the filters that have been applied to the original SES-SBS-NR data have a minimal impact on the firm-level mean, standard deviation, average Euclidean distance, and coefficient of variation of workers' HDIs.

[9] Additional descriptive statistics on the composition of the workforce by origin within firms (and relative HDIs) are reported in Appendix B.

different productivity. It is, therefore, subject to a "potential selectivity bias" (Aeberhardt and Pouget, 2010). Third, the individual-level equations of the Oaxaca-Blinder framework ignore productivity spillovers that occur at the firm level. The conclusion that Bartolucci (2014) draws from this is harsh: "As discrimination has normally been detected through the unexplained gap in wage equations and this approach is not the best option for disentangling differences in productivity and discrimination, there are few papers that address labor market discrimination against immigrants" (p. 3).

The increasing availability of firm-level matched employer-employee data facilitated the emergence of an alternative approach to measure discrimination. This new method was developed by Hellerstein et al. (1999) and refined by Hellerstein and Neumark (2006), Vandenberghe (2011), and Van Ours and Stoeldraijer (2011), among others. It is now standard in the literature on productivity and wage effects of labor heterogeneity (Cardoso et al., 2011, Devicienti et al., 2018; Garnero et al., 2014a,b; Giuliano et al., 2017; Göbel and Zwick, 2012; Vandenberghe, 2013). It is based on the separate estimation of a production function and a wage equation at the firm level. The production function yields estimates for the average marginal product of each category of workers (natives, foreigners, and so forth), while the wage equation estimates the impact of each group on the average wage paid by the firm. Estimating both equations with the same set of explanatory variables allows comparing the parameters associated with contributions to the (average) marginal product and (average) wage of each category of workers.

Such method captures compositional and sorting effects that are ignored in the Oaxaca-Blinder framework; crucially, the productivity differences associated with observable characteristics are directly measured instead of being assumed. However, these advantages often deliver potential rather than actual mileage: while the firm-level wage equations in the Hellerstein et al.'s (1999) framework are generally robust to different specifications and provide precise estimates, the identification of production functions is often far more problematic due to high standard errors and noise in the productivity measures (Göbel and Zwick, 2012; Vandenberghe, 2013). Bartolucci (2014) argues that it is difficult to obtain precise estimates of the relative productivity parameter. Indeed, the search for the appropriate form of the production function is a long-standing quest in the microeconometric literature (Ackerberg et al., 2015). While empirical studies focusing only on firm-level production functions are more flexible in the choice of both the functional form and statistical estimator, the Hellerstein et al.'s (1999) method imposes symmetry between the wage-setting and production function equations. This explains most studies' use of simple constant elasticity of substitution or Cobb-Douglas functional forms and Fixed Effects (FE) or Generalized Methods of Moments-Instrumental Variable (GMM-IV) estimators for both equations. The problem is that the compelling theoretical reasons to use Olley-Pakes (Olley and Pakes, 1996) or Levinsohn-Petrin (Levinsohn and Petrin, 2003) estimators for production functions often lack a theoretical rationale in the case of wage equations. The fact that some firm-level studies on immigration estimate only production functions (Nicodemo, 2013; Paserman, 2013), and others only wage equations (Böheim et al., 2012) is a way to circumvent this issue but comes at the price of renouncing from measuring wage discrimination.

In this paper, we build on the method developed by Bartolucci (2014) that i) avoids the specification of the functional form of the production function equation and directly uses firm-level productivity data to measure discrimination against foreigners; ii) neither assumes perfect competition in the labor market nor a linear relationship between wages and productivity (it allows for non-unitary wage-productivity elasticities); and iii) produces a measure of wage discrimination against foreigners that is robust to labor market segregation.[10]

The wage-setting equation proposed by Bartolucci (2014) is similar to the wage equation in the Hellerstein et al.'s (1999) framework but directly estimates a parameter for the logarithm of average firm-level productivity. The integration of measured productivity yields the following wage equation:

$$\log(\overline{w}_{j,t}) = \alpha_j + \beta \log(\overline{p}_{j,t}) + \gamma \mathrm{AED(HDI)}_{j,t} + \lambda X_{j,t} + \varepsilon_{j,t}, \quad (1)$$

where the dependent variable, $\log(\overline{w}_{j,t})$, is the logarithm of the average hourly wage in firm j in year t; $\log(\overline{p}_{j,t})$ is the logarithm of average hourly productivity (defined as value added per hour worked); $\mathrm{AED(HDI)}_{j,t}$ is the average Euclidean distance based on HDIs (our measure of diversity), and γ is the parameter that captures potential wage discrimination; $X_{j,t}$ is a vector containing a set of observable characteristics of firm j and its workforce in year t, including controls for the occupational, educational, and sectoral composition of the firm (the full set of control variables is shown in Table 1).

4.2. Estimation methods

Equation (1) can be estimated with different methods. Basic pooled Ordinary Least Squares (OLS) estimations of productivity models are criticized for their potential "heterogeneity bias" (Vandenberghe, 2013). This is because firm productivity and mean wages depend in no small extent on firm-specific characteristics that are not measured in micro-level surveys. As a consequence, these estimators likely are biased since unobserved firm characteristics may simultaneously affect the firm's value added (or wages) and the composition of its workforce.

Empirical studies show that firm-level fixed effects are important determinants of wage differentials between male foreigners and natives, so that including firm fixed effects attenuates the problem of unobserved firm characteristics (Aydemir and Skuterud, 2008). However, the FE estimator does not address the potential endogeneity of the explanatory variables. For several reasons, the composition of a firm's workforce is potentially endogenous. First, the average wage offered by the firm might influence its attractiveness for workers, and a relatively higher wage could attract workers with better unobserved skills. Second, shocks in productivity levels or wages might generate correlated changes in the firm's workforce composition. For instance, in periods of economic downturns, firms might lay off more foreigners than natives. In order to tackle both firm fixed unobserved

[10] We refer the reader to Bartolucci (2014) for the proofs of these properties.

heterogeneity and potential endogeneity, we estimate all the equations also using a GMM-IV specification in first differences with instrumental variables (Black and Lynch, 2001; Dearden et al., 2006). We use two types of instruments. Following Van Ours and Stoeldraijer (2011) and Göbel and Zwick (2012), the first instruments first-differenced AEDs with their lagged levels. The implicit assumption is that changes in wages in one period, although possibly correlated with contemporaneous variations in AEDs, are unrelated to lagged levels of the latter. Moreover, changes in AEDs are assumed to be sufficiently correlated with their past levels. The second instrument is the annual average AED in the sector in which firm j operates.[11] The rationale for this instrument is that sectoral AEDs are correlated with the AED of firm j, while being unrelated to the productivity (and wage) of firm j and the error term (Garnero, 2015).

In order to assess the soundness of this approach, we perform an array of statistical tests. The first measures whether the correlation between the instrumental variables and the endogenous variables is sufficiently strong, that is, that the instruments are not weak. For this purpose, we used the Kleibergen-Paap rk Wald F statistic. Under the null hypothesis, the instruments are weak. A standard rule of thumb is to reject the null hypothesis if the F statistic has a value of at least 10 (Van Ours and Stoeldraijer, 2011). The second test is the Kleibergen-Paap rk LM statistic, whose null hypothesis is that the equation is under-identified. The third test concerns the validity of the instruments and uses the Hansen's (1982) test of over-identifying restrictions. Under the null hypothesis, the instruments are valid, that is, uncorrelated with the error term. A fourth indicator tests whether AEDs are indeed endogenous so that an IV approach is warranted. Under the null hypothesis, the explanatory variables can be treated as exogenous.

5. Results

In this section, we confront the different theoretical arguments presented in Section 2 with empirical evidence based on estimations of Equation (1). We present results for three variations: the benchmark results for the entire sample (Section 5.1); separate estimations for firms with and without firm-level collective bargaining (Section 5.2); and separate estimations for firms with low, high, and very high levels of diversity (Section 5.3).

5.1. Benchmark results

Table 2 presents the results for the three different estimators discussed in Section 4.2: OLS, FE, and GMM-IV.

To examine the effect of the control variables in the model, we also show OLS estimates without individual and firm controls (Column (1)), with individual but no firm controls (Column (2)), and including both individual and firm controls (Column (3)). The effect of the control variables on the magnitude of the diversity coefficient is considerable: the size of the coefficient associated with the AED drops from -1.14 (without any controls)

[11] The average is calculated excluding the firm j.

to -0.08 (including all controls). This suggests that higher diversity is strongly correlated with variables that are relevant for firm-level wage setting. The substantial difference between the AED coefficients in the first two columns of Table 2 indicates that individual control variables (gender, age, education, tenure, and contract type) are responsible for the bulk of this effect.

[Insert Table 2 about here]

OLS estimators are biased if unobservable variables play a role in the wage-setting process. The FE estimator (Column (4)) accounts for unobservable firm-level time-invariant variables. The resulting AED coefficient is higher compared to the OLS equations (Columns (2) and (3)). Contrary to the *observable* individual and firm characteristics, the *unobservable* time-invariant characteristics of the firm and its workforce increase the size of the negative effect of diversity on wages. In other words, diversity appears to be stronger in firms with unobservable characteristics that have a positive effect on average wages.

The AED coefficient derived from the GMM-IV estimator equals -0.37 and is, therefore, higher compared to both the OLS and FE coefficients. The post-estimation tests reported at the bottom of the table (Column (5)) suggest that the variables of interest are not endogenous. We have confidence in the validity of these endogeneity tests since the instruments that we use in the GMM-IV are sound: the tests for weak-, over-, and under-identification suggest that the instrumentation is satisfactory. Given that the AED appears not to be endogenous and that the FE estimator is more efficient compared to the GMM-IV one, the FE estimator is our preferred estimator.

How should the FE diversity coefficient of -0.15 be interpreted? Following Bartolucci's (2014) interpretation of Equation (1), the coefficient captures a *ceteris paribus* effect of diversity on the firm's average wage, which controls for the workforce composition, firm characteristics and, crucially, firm-level productivity. The latter is statistically significant in the OLS and FE models (Columns (3) and (4), respectively). This means that the negative coefficients that we observe in Table 2 are evidence for wage discrimination: the negative association between firm-level diversity and wages is not justified by productivity differences or other wage-related factors. Higher diversity is correlated with lower *ceteris paribus* wages and, therefore, higher discrimination.

How sizable is the observed wage discrimination? Should a coefficient of -0.15 be regarded as heavy discrimination against a more diverse workforce? To answer these questions, it is useful to go back to the descriptive statistics presented in Table 1. In the sample, the average HDI associated with workers equals 0.87 (with a standard deviation of 0.03), and the average AED in the firms equals 0.03 (with a standard deviation of 0.04). This means that the average Euclidean distance between two workers is 0.03. This is a relatively low gap in HDIs. Compared to the HDI of a Belgian worker (0.88), only 24 out of 195 countries classified by UNDP (2017) yield a distance in terms of human development equal or below 0.03. Most of these countries are Belgium's Western European neighbors or OECD countries.

We can illustrate the magnitude of potential changes in firms' AEDs through hypothetical examples. If a Belgian firm were composed of 1 individual from each country

for which we have HDIs, the diversity would be much higher than the current value of 0.03: in such a hypothetical multicultural firm, the AED would equal 0.21. A more realistic example would be a firm composed of 30 workers of Belgian origin (and hence an AED equal to 0), which recruits 10 individuals with Moroccan background (one of the most common foreign nationality in Belgium: around 20% of foreigners in the sample are born in Morocco). In this case, the AED of the firm would increase to 0.07. The wage change due to discrimination – that is, the decrease *beyond* variables related to productivity, sorting, and segregation – would equal 0.07 x -0.15 = -0.01, or -1 percentage point. These examples suggest that only relatively abrupt changes in the diversity of a firm's workforce lead to sizable wage discrimination effects. This numeric exercise also illustrates that such abrupt changes are not improbable in countries with increasingly diverse workforces. In Germany, for instance, the labor force increased by 1.5% in 2017, mostly due to foreigners. Moreover, we will show below that, in particular firms, the discriminatory effects are more than 3 times higher than the benchmark estimations; this means that relatively small changes in diversity can still lead to sizable changes in firm-level wages.

5.2. The role of collective bargaining

A testable hypothesis that we formulated in Section 2.2 postulates that collective bargaining institutions could attenuate the link between firm-level diversity and wage discrimination. We now present estimation results on the moderating role of a specific type of collective bargaining in the Belgian labor market, namely, the renegotiation of national and sectoral wage agreements through firm-level bargaining. We obtain these estimates by dividing the full sample into firms that renegotiated collective agreements at the firm level (this concerns 20% of the observations) and those that apply only national and sectoral agreements (the remaining 80%). Under the hypothesis that trade unions work towards wage equality and the reduction of any discriminatory pay that was not eliminated in national and sectoral agreements that are binding for all companies, the renegotiation at the local level would reduce the extent of wage discrimination picked up by the diversity coefficient in Equation (1). The results presented in Table 3 provide evidence that this is indeed the case. The negative impact of the AED on the average hourly wage equals -0.13 in firms that renegotiate wages (Column (7)); this effect is significantly lower than that estimated for the full sample (Column (6)) and the sample of firms without firm-level renegotiation (Column (8)).[12]

[Insert Table 3 about here]

[12] *T*-tests for equality of regressions coefficients are equal to 18.8*** and 37.3*** when comparing coefficients associated with the AED(HDI) variable, respectively, in Columns (7) and (6) (i.e., in firms that renegotiate wages *versus* in the full sample) and in Columns (7) and (8) (i.e., in firms that renegotiate wages *versus* in firms without firm-level renegotiation) of Table 3. They thus confirm that differences in regression coefficients are statistically significant.

5.3. The impact of the level of diversity

A second testable hypothesis relates to the importance of group dynamics for wage discrimination to become evident within firms. According to the reasoning presented in Section 2.3, it seems likely that such group dynamics unfold only if a certain critical mass of diversity is reached. In order to test for the presence of such thresholds, we estimate separate wage equations for three sub-samples that differentiate between firms with low, high, and very high diversity.

The definition of the three subsamples is based on the percentile distribution of AEDs in the full sample. We define firms with relatively low diversity when the diversity indicator is equal or below percentile 75 (AED = 0.04). High-diversity firms are defined as those with AEDs between percentiles 75 and 95 (AED = 0.11). Firms with very high diversity are defined as firms with AEDs that lie above percentile 95. The three subsamples, therefore, include, respectively, 75, 20, and 5% of the firm-year observations.

The estimation results of our preferred FE estimator are presented in Table 4. They provide evidence that wage discrimination related to firm-level diversity can only be detected in firms with high or very high diversity. In those firms that lie below percentile 75, the diversity coefficient is also negative but not statistically different from zero. Another remarkable result is that the size of the effect increases with the extent of diversity: compared to the full sample (Column (9)), the effect of the same change in diversity on wages is 4 percentage points greater in firms with high diversity (Column (11)) and 35 percentage points greater in firms with very high diversity (Column (12)).[13] This suggests that most firms are not concerned with a link between diversity and wage discrimination. In the average firm, diversity is arguably not visible or influential enough to affect wage-setting processes beyond those that are associated with other variables included in the model. Instead, in firms with higher levels of diversity, a powerful mechanism towards wage depression seems to be at work. A relatively modest change in the composition of firms with higher diversity leads to sizable changes in average wages. In sum, diversity does not entail wage discrimination in most firms – except for firms with higher levels of diversity, in which it is very problematic indeed.[14,15]

---

[13] *T*-tests for equality of regressions coefficients are equal to 55.1***, 68.1***, and 59.8*** when comparing coefficients associated with the AED(HDI) variable, respectively, in Columns (9) and (11) (i.e., in the full sample *versus* in firms with high diversity), in Columns (9) and (12) (i.e., in the full sample *versus* in firms with very high diversity), and in Columns (11) and (12) (in firms with high diversity *versus* in firms with very high diversity) of Table 4. They thus confirm that differences in regression coefficients are statistically significant.

[14] Descriptive statistics provided in Appendix B enable to interpret these effects in more detail. Most importantly, the strong negative coefficients associated with the AED(HDI) variable in Table 4 for firms with high and very high diversity (which would belong to the fourth and fifth quintiles of Table B.3, see Appendix B) are presumably related to the clustering of groups of non-EU15 immigrants in these firms. In firms with very high diversity – defined in the paper as the five percentiles with the highest AED(HDI) – the proportion of Belgians sinks to 44%, whereas Western Europeans represent 14% of the workforce in the average firm. By contrast, immigrants from North Africa and Sub-Saharan Africa represent 27% and 7%, respectively.

[15] As a robustness test, we re-estimated the preferred FE model by quartiles, that is, fixing cut-off points at percentiles 25, 50, and 75 of the AED(HDI) variable at the firm level. The corresponding estimates (available upon request) corroborate our conclusion that there is no wage penalty in low-diversity firms (i.e., in firms belonging to the three lower quartiles). The significant negative wage effect in firms with high diversity is confirmed in the subsample covering the upper quartile. As expected, the coefficient lies between the estimates

[Insert Table 4 about here]

5.4. Robustness tests

Given the novelty of our approach to convert qualitative data on workers' origin into a quantitative indicator of diversity that takes into account heterogeneity in levels of socio-economic development of migrants' home countries, we carry out several robustness tests to assess the sensitivity of the results to different methodological choices. None of these tests alter the main observations borne out by the discussion of results in the previous three sections.

The first robustness test concerns the choice of the reference year for HDIs. For reasons presented in Section 3.2, the benchmark regressions presented above use the 2010 HDIs (i.e., those referring to our last year of observation). We re-estimate the regressions using the reference years 2000 and 1990, which are the two previous years for which the United Nations' Development Programme published HDIs for a large sample of countries. Probably due to the strong inertia in human development values, the estimation coefficients are not affected by the choice of the reference year (complete regression outputs of these robustness tests are available upon request).

Another robustness test concerns the impact of methodological choices related to the aggregation of individual HDIs into a firm-level indicator of diversity. As summarized in Kampelmann (2016), methodological choices underlying such aggregation can impact the numerical values of diversity indicators. In addition to the use of average Euclidean distances on which we based the regression results above, we re-estimate all the regressions using an alternative indicator of diversity, namely the standard deviation of HDIs within each firm. Results are presented in Appendix C. We find that the signs and significance levels of OLS, FE, and GMM-IV estimates are almost not sensitive to whether we use the standard deviation or the average Euclidean distance to aggregate information on diversity within firms. Moreover, the magnitude of our preferred estimate (i.e., that obtained by FE, as the *p*-value associated with the endogeneity test is equal to 0.16) does not vary substantially: it decreases from -0.15 to -0.11 when using the standard deviation of HDIs instead of the average Euclidean distance at the firm level. As an alternative measure of diversity, we also used the coefficient of variation of HDIs at the firm level. Again, the estimated coefficients (available upon request) are close enough to the benchmark estimations to conclude that the benchmark results are robust to different methodological choices of aggregating HDIs at the firm level.

To enable comparability with earlier results in the literature and, in particular, with those of Bartolucci (2014) for Germany, we further re-estimate the model using the share of immigrants as the main explanatory variable instead of the AED(HDI). In the case of Belgium, evidence (Corluy et al., 2015; FPS Employment and Unia, 2017; Higher Council for Employment, 2018; Institute for Equality of Women and Men, 2010; Martens et al., 2005)

---

for firms with high diversity (i.e., with AEDs between percentiles 75 and 95) and very high diversity (i.e., with AEDs above percentile 95). It is equal to -0.22 (p-value < 0.01).

suggests that we can address the problem of heterogeneity among immigrants by distinguishing between individuals from the EU15 and those from outside the EU15. Accordingly, we first compare the wage outcomes of workers born in the EU15 (natives) with those born outside the EU15 (immigrants). However, in order to enable stricter comparability with the study of Bartolucci (2014), we further define natives as people born with the Belgian nationality and compare their earnings with those of immigrants (i.e., workers born with a non-Belgian nationality). Results are reported in Appendix D. When using the share of non-EU15 workers as the main explanatory variable, we find that the regression coefficient associated with this variable is systematically negative and significant. It stands at -0.04 when using the full specification with firm fixed effects. This is our preferred estimate as the *p*-value associated with the endogeneity test indicates that the null hypothesis of no endogeneity should not be rejected. The second robustness test uses the share of Belgian-born workers as the reference category (and the share of non-Belgian-born workers as the main explanatory variable). In this set-up, our preferred estimate is the one obtained by GMM-IV given that the endogeneity test is found to be significant and that the over-identification test suggests that the instruments are valid. The regression coefficient associated with the share of non-Belgian workers is equal to -0.11, which is very close to the outcome reported by Bartolucci (2014). Indeed, Bartolucci (2014) obtained a coefficient of -0.12 for the share of foreign workers in Germany.[16]

6. Conclusions

In most OECD countries, it is increasingly common that firms employ a diverse workforce, with employees from different countries of origin. The impact of this diversity was the object of intense study in labor economics, albeit with the uncomfortable methodological limitation that diversity was measured with metrics that eliminate most of the socially and economically relevant differences among workers' countries of origin, which are crucial dimensions of labor diversity and well-known key determinants of migrants' labor market outcomes.

The approach used in this paper is entirely novel and, based on the conversion of the qualitative information on individuals' countries of origin into an aggregate firm-level diversity indicator built from HDIs, it explicitly takes into account heterogeneity in levels of socio-economic development of migrants' home countries. We used this new aggregate measure of firm-level diversity in state-of-the-art wage equations that control for a range of observable and time-invariant unobservable factors, including differences in labor productivity between firms and within firms over time.

The estimated parameters for the Belgian labor market can be interpreted as evidence for a significant impact of firm-level diversity on average wages. Since these effects occur beyond other factors related to productivity as well as individual and firm characteristics, they can be interpreted as wage discrimination. Compared to the levels of aggregate

[16] We also checked the sensitivity of the findings using an index of polarization instead of the diversity indicator AED(HDI). Detailed results are presented and discussed in Appendix E.

diversity at the firm-level that are observed in Belgium, the absolute magnitude of this wage discrimination is relatively modest: an abrupt shift in the firm's level of diversity is required to yield economically meaningful downward changes in the wage level. The extent of wage discrimination lowers further if we account for the moderating role that collective bargaining could play. For the case of Belgium, where national and sectoral collective agreements apply to virtually all firms but in which local collective renegotiation happens only in some firms, we are able to provide evidence that collective bargaining at the firm-level attenuates the link between diversity and wage discrimination.

Should one conclude from these results that diversity-related wage discrimination is a minor nuisance, for which, even if it occurs, an effective remedy exists in the form of institutionalized collective bargaining? Additional evidence based on diversity thresholds suggests that diversity is indeed relatively unproblematic for wages in most Belgian firms. However, in firms with higher levels of diversity, diversity-related wage discrimination is intense. We interpret this finding that it is not diversity *per se* that sparks a mechanism of discriminatory wage setting, but *levels* of diversity that lie far above the composition of the average firm. This can be accounted for by the hypothesis that group dynamics linked to diversity appear only in high-diversity firms that are characterized by a sizable share of individuals with sub-standard HDIs.

In terms of policy implications, our results suggest to worry less about diversity in general and instead focus on specific spots in the economy that concentrate particularly high levels of diversity. The discrimination in these locations could be alleviated through a stronger presence of collective bargaining or interventions aimed at avoiding or reducing the clustering of foreigners from low-HDI countries in such settings.

References

Ackerberg, D., Caves, K., & Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, 83(6), 2411-2451.

Adsera, A., & Chiswick, B. R. (2007). Are there gender and country of origin differences in immigrant labor market outcomes across European destinations? *Journal of Population Economics*, 20(3), 495-526.

Aeberhardt, R., & Pouget, J. (2010). National origin differences in wages and hierarchical positions. *Annals of Economics and Statistics/Annales d'Économie et de Statistique*, 99-100, 117-139.

Alesina, A., & La Ferrara, E. (2005). Ethnic diversity and economic performance. *Journal of Economic Literature*, 43(3), 762-800.

Arrow, K. (1973). The theory of discrimination. In Ashenfelter, O., & Rees, A. (Eds.), *Discrimination in labor markets* (pp. 3-33). Princeton University Press, Princeton, NJ, USA.

Åslund, O., & Skans, O. N. (2010). Will I see you at work? Ethnic workplace segregation in Sweden, 1985-2002. *Industrial and Labor Relations Review*, 63(3), 471-493.

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

Bartolucci, C. (2014). Understanding the native-immigrant wage gap using matched employer-employee data: Evidence from Germany. *Industrial and Labor Relations Review,* 67(4), 1166-1202.

Bayard, K., Ilellerstein, J., Neumark, D., & Troske, K. (1999). Why are racial and ethnic wage gaps larger for men than for women? Exploring the role of segregation using the new worker-establishment characteristics database. In Haltiwanger, J. (Ed.), *The creation and analysis of employer-employee matched data* (pp. 175-203). Emerald Group Publishing Limited, Amsterdam, NL.

Black, S. E., & Lynch, L. M. (2001). How to compete: The impact of workplace practices and information technology on productivity. *The Review of Economics and Statistics*, 83(3), 434-445.

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources*, 8(4), 436-455.

Bloomekatz, R. (2007). Rethinking immigration status discrimination and exploitation in the low-wage workplace. *UCLA Law Review*, 54(6), 1963-2010.

Böheim, R., Horvath, T., & Mayr, K. (2012). Birthplace diversity of the workforce and productivity spill-overs in firms. WIFO Working Paper No. 438.

Bundesamt für Migration und Flüchtlinge. (2016). Migrationsbericht 2015 des Bundesamtes für Migration und Flüchtlinge im Auftrag der Bundesregierung. Bundesministerium des Innern, Berlin, DE.

Cardoso, A. R., Guimaraes, P., & Varejao, J. (2011). Are older workers worthy of their pay? An empirical investigation of age-productivity and age-wage nexuses. *De Economist*, 159(2), 95-111.

Carrington, W. J., & Troske, K. R. (1998). Interfirm segregation and the black/white wage gap. *Journal of Labor Economics*, 16(2), 231-260.

Corluy, V., Haemels, J., Marx, I., & Verbist, G. (2015). The labour market position of second-generation immigrants in Belgium. NBB Working Paper No. 285.

Dearden, L., Reed, H., & Van Reenen, J. (2006). The impact of training on productivity and wages: Evidence from British panel data. *Oxford Bulletin of Economics and Statistics,* 68(4), 397-421.

Dell'Aringa, C., & Lucifora, C. (1994). Wage dispersion and unionism: Do unions protect low pay? *International Journal of Manpower*, 15(2/3), 150-169.

Demunter, C. (2000). Structure and distribution of earnings survey: Analysis 1995. Statistics Belgium Working Paper.

Dustmann, C., Glitz, A., Schönberg, U., & Brücker, H. (2015). Referral-based job search networks. *The Review of Economic Studies*, 83(2), 514-546.

Dustmann, C., & Van Soest, A. (2002). Language and the earnings of immigrants. *Industrial and Labor Relations Review,* 55(3), 473-492.

Devicienti, F., Grinza, E., & Vannoni, D. (2018). The impact of part-time work on firm productivity: Evidence from Italy. *Industrial and Corporate Change*, 27(2), 321-347.

Elliott, R. J., & Lindley, J. K. (2008). Immigrant wage differentials, ethnicity and occupational segregation. *Journal of the Royal Statistical Society: Series A (Statistics in Society),* 171(3), 645-671.

Esteban, J., & Ray, D. (1994). On the measurement of polarization. *Econometrica*, 62(4), 819-582.

Ferrer, A., Green, D. A., & Riddell, W. C. (2006). The effect of literacy on immigrant earnings. *Journal of Human Resources*, 41(2), 380-410.

FPS Employment, & Unia (2017). Monitoring socio-économique 2017: Marché du travail et origine. Report 2017, Brussels, BE.

Freeman, R. B. (1980). Unionism and the dispersion of wages. *Industrial and Labor Relations Review*, 34(1), 3-23.

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

Garnero, A. (2015). *Institutions and heterogeneity in the labour market.* Ph.D. Thesis, École Normale Supérieure and Université Libre de Bruxelles, Paris, FR and Brussels, BE.

Garnero, A., Kampelmann, S., & Rycx, F. (2014a). The heterogeneous effects of workforce diversity on productivity, wages, and profits. *Industrial Relations: A Journal of Economy and Society*, 53(3), 430-477.

Garnero, A., Kampelmann, S., & Rycx, F. (2014b). Part-time work, wages, and productivity: Evidence from Belgian matched panel data. *Industrial and Labor Relations Review,* 67(3), 926-954.

Giuliano, R., Mahy, B., Kampelmann, S., & Rycx F. (2017). Short notice, big difference? The effect of temporary employment on firm competitiveness across sectors. *British Journal of Industrial Relations*, 55(2), 421-449.

Göbel, C., & Zwick, T. (2012). Age and productivity: Sector differences. *De Economist*, 160(1), 35-57.

Gratsberg, B., & Ragan, J. F. (2002). The impact of host-country schooling on earnings. *Journal of Human Resources*, 37(1), 63-105.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4), 1029-1054.

Harcourt, M., Lam, H., Harcourt, S., & Flynn, M. (2008). Discrimination in hiring against immigrants and ethnic minorities: The effect of unionization. *International Journal of Human Resource Management*, 19(1), 98-115.

Heckman, J. J. (1998). Detecting discrimination. *Journal of Economic Perspectives*, 12(2), 101-116.

Hellerstein, J. K., & Neumark, D. (2003). Ethnicity, language, and workplace segregation: Evidence from a new matched employer-employee data set. *Annals of Economics and Statistics/Annales d'Économie et de Statistique*. 71-72, 19-78.

Hellerstein, J. K., & Neumark, D. (2006). Using matched employer-employee data to study labor market discrimination. In Rodgers, W. M. (Ed.), *Handbook on the economics of discrimination*, (Vol. 3, pp. 29-60). Edward Elgar Publishing, Cheltenham, UK.

Hellerstein, J. K., & Neumark, D. (2008). Workplace segregation in the United States: Race, ethnicity, and skill. *The Review of Economics and Statistics*, 90(3), 459-477.

Hellerstein, J. K., Neumark, D., & Troske, K. R. (1999). Wages, productivity, and worker characteristics: Evidence from plant-level production functions and wage equations. *Journal of Labor Economics*, 17(3), 409-446.

Higher Council for Employment (2018). Immigrants born outside the European Union on the Belgian labour market. Report 2018, Brussels, BE.

Institute for the Equality of Women and Men (2010). How to close the gender pay gap? Strategies and good practices of states and social partners in Europe. Report 2010, Brussels, BE.

Kampelmann, S. (2011). *The socio-economics of pay rules.* Ph.D. Thesis, Université de Lille 1, Lille, FR.

Kampelmann, S., & Rycx, F. (2016). Wage discrimination against immigrants: Measurement with firm-level productivity data. *IZA Journal of Migration*, 5(15), 1-24.

Levinsohn, J., & Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies* 70(2), 317-341.

Martens, A., Ouali, N., Van de Maele, M, Vertommen, S., Dryon, P., & Verhoeven, H. (2005). Discrimination des étrangers et des personnes d'origine étrangère sur le marché du travail de la Région de Bruxelles-Capitale. Rapport de synthèse pour l'ORBEM, Brussels, BE.

Montalvo, J., & Reynal-Querol, M. (2005). Ethnic polarization, potential conflict, and civil wars. *American Economic Review*, 95(3), 796-816.

Nicodemo, C. (2013). Immigration and labor productivity: New empirical evidence for Spain. IZA Discussion Paper No. 7279.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, 14(3), 693-709.

Olley, S., & Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6), 1263-1297.

Ottaviano, G. I., & Peri, G. (2012). Rethinking the effect of immigration on wages. *Journal of the European Economic Association*, 10(1), 152-197.

Paserman, M. D. (2013). Do high-skill immigrants raise productivity? Evidence from Israeli manufacturing firms, 1990-1999. *IZA Journal of Migration*, 2(6), 1-31.

Peri, G., & Sparber, C. (2009). Task specialization, immigration, and wages. *American Economic Journal: Applied Economics*, 1(3), 135-169.

Phelps, E. S. (1972). The statistical theory of discrimination. *American Economic Review,* 62(4), 659-661.

Pillinger, J. (2014). Bargaining for equality: How collective bargaining contributes to eliminating pay discrimination between women and men performing the same job or job of equal value. European Trade Union Confederation, Brussels, BE.

Plasman, R., Rusinek, M., & Rycx, F. (2007). Wages and the bargaining regime under multi-level bargaining: Belgium, Denmark and Spain. *European Journal of Industrial Relations*, 13(2), 161-180.

Seidel, M.-D. L., Polzer, J. T., & Stewart, K. J. (2000). Friends in high places: The effects of social networks on discrimination in salary negotiations. *Administrative Science Quarterly*, 45(1), 1-24.

UNDP (2017). Human development report 2016. Human development for everyone. United Nations Development Programme, New York, NY, USA.

Van Ours, J. C., & Stoeldraijer, L. (2011). Age, wage and productivity in Dutch manufacturing. *De Economist*, 159(2), 113-137.

Vandenberghe, V. (2011). Firm-level evidence on gender wage discrimination in the Belgian private economy. *Labour*, *25*(3), 330-349.

Vandenberghe, V. (2013). Are firms willing to employ a greying and feminizing workforce? *Labour Economics*, 22, 30-46.

Table 1: Firm-level descriptive statistics, 1999-2010

	Full sample		Firms with low diversity [e]		Firms with high diversity [f]		Firms with very high diversity [g]	
Variable	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Human Development Index (HDI)	0.87	0.03	0.88	0.01	0.83	0.04	0.78	0.04
Diversity indicators								
AED(HDI) [a]	0.03	0.04	0.01	0.01	0.08	0.03	0.13	0.02
SD(HDI) [b]	0.03	0.04	0.02	0.02	0.09	0.03	0.13	0.02
CV(HDI) [c]	0.04	0.05	0.02	0.02	0.11	0.04	0.17	0.03
Worker characteristics								
Women	0.27	0.24	0.27	0.24	0.27	0.24	0.27	0.24
Workers < 40 years	0.55	0.20	0.54	0.20	0.57	0.20	0.60	0.20
Education level 1 (ISCED 1-2)	0.32	0.33	0.30	0.32	0.38	0.34	0.47	0.34
Education level 2 (ISCED 3-4)	0.42	0.30	0.43	0.30	0.40	0.30	0.39	0.31
Education level 3 (ISCED 5-7)	0.25	0.27	0.26	0.27	0.22	0.26	0.14	0.19
Permanent labor contracts	0.97	0.10	0.97	0.09	0.95	0.13	0.93	0.16
High tenure (>5 years)	0.54	0.25	0.56	0.24	0.48	0.25	0.40	0.25
Occupations								
Managers	0.04	0.08	0.04	0.08	0.04	0.07	0.03	0.06
Professionals	0.10	0.19	0.10	0.19	0.09	0.18	0.04	0.12
Technical assistance professionals	0.08	0.15	0.08	0.16	0.07	0.14	0.04	0.12
Clerical occupations	0.19	0.22	0.20	0.22	0.17	0.21	0.14	0.20
Service occupations	0.07	0.20	0.06	0.19	0.09	0.23	0.15	0.30
Crafts	0.25	0.33	0.24	0.33	0.25	0.34	0.25	0.35
Machine operators	0.18	0.29	0.19	0.29	0.16	0.28	0.13	0.26
Elementary occupations	0.09	0.20	0.08	0.18	0.14	0.26	0.21	0.31
Firm characteristics								
Value added/hour [d]	55.90	435.85	58.29	494.01	48.74	166.30	43.28	230.14
Wage/hour [d]	15.60	5.65	15.85	5.92	14.87	4.67	13.10	3.17
Firm-level collective bargaining	0.19	0.40	0.20	0.40	0.17	0.38	0.14	0.35
Capital (log)	10.68	1.61	10.75	1.58	10.50	1.65	10.22	1.75
Firm size (log)	4.29	1.14	4.27	1.13	4.33	1.15	4.26	1.19
Mining and quarrying	0.01	0.08	0.01	0.08	0.00	0.06	0.00	0.00
Manufacturing	0.45	0.50	0.47	0.50	0.41	0.49	0.32	0.47
Electricity, gas, and water supply	0.00	0.07	0.00	0.07	0.00	0.07	0.01	0.09
Construction	0.11	0.32	0.11	0.31	0.12	0.33	0.09	0.29
Wholesale and retail trade	0.19	0.39	0.20	0.40	0.14	0.35	0.13	0.33
Hotels and restaurants	0.02	0.15	0.01	0.10	0.07	0.25	0.18	0.39
Transport, storage, and communication	0.07	0.26	0.07	0.26	0.08	0.28	0.09	0.28
Financial intermediation	0.01	0.11	0.01	0.11	0.01	0.11	0.01	0.07
Real estate, renting, and business services	0.13	0.33	0.12	0.32	0.15	0.36	0.17	0.38
Region								
Brussels	0.15	0.35	0.10	0.30	0.28	0.45	0.44	0.50
Wallonia	0.26	0.44	0.26	0.44	0.24	0.43	0.13	0.34
Flanders	0.60	0.49	0.64	0.48	0.48	0.50	0.43	0.49
Observations	30,355		22,767		7,588		1,517	

Notes: [a] AED: average Euclidean distance. [b] SD: standard deviation. [c] CV: coefficient of variation. [d] Constant Euros. [e] Firms with low diversity are defined as those with AEDs below percentile 75. [f] Firms with high diversity are defined as those with AEDs between percentiles 75 and 95. [g] Firms with very high diversity are defined as those with AEDs above percentile 95.

Figure 1: Firm-level average Euclidean distance and shares of Belgians

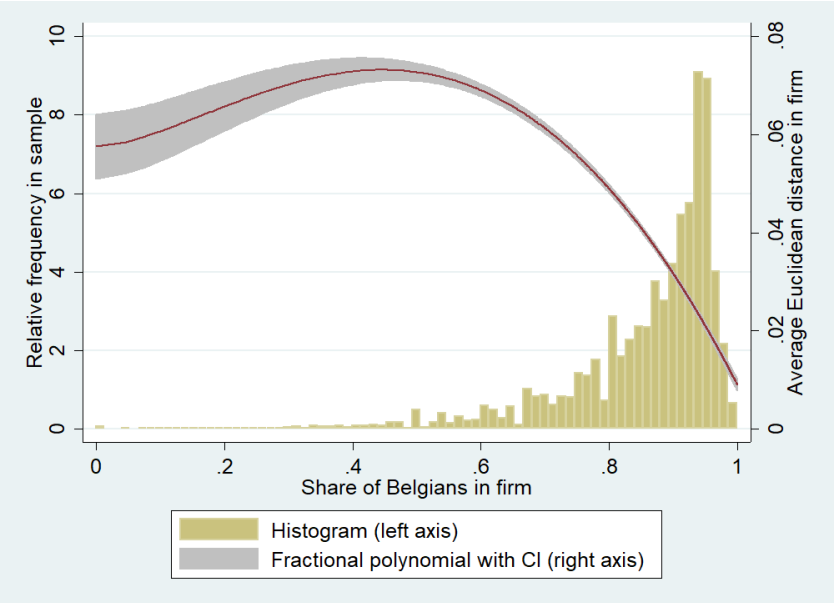


Table 2: Estimations of firm-level wage-setting equations, 1999-2010

Dependent variable: average hourly wage (log)	OLS (1)	OLS (2)	OLS (3)	FE (4)	GMM-IV (5)
AED(HDI) [a]	-1.14*** (0.06)	-0.10** (0.04)	-0.08** (0.04)	-0.15*** (0.04)	-0.37** (0.18)
Share of women [b]		-0.20*** (0.01)	-0.17*** (0.01)	-0.10*** (0.02)	-0.13*** (0.04)
Workers < 40 years		-0.16*** (0.01)	-0.18*** (0.01)	-0.14*** (0.01)	-0.12*** (0.01)
Education level 2 (ISCED 3-4)		0.36*** (0.01)	0.27*** (0.01)	0.12*** (0.01)	0.11*** (0.01)
Education level 3 (ISCED 5-7)		0.06*** (0.00)	0.04*** (0.00)	0.02*** (0.00)	0.01*** (0.00)
Managers		0.92*** (0.03)	0.79*** (0.03)	0.51*** (0.02)	0.54*** (0.03)
Professionals		0.42*** (0.02)	0.34*** (0.01)	0.21*** (0.01)	0.23*** (0.02)
Technical assistance professionals		0.26*** (0.01)	0.17*** (0.01)	0.08*** (0.01)	0.09*** (0.02)
Clerical occupations		0.04*** (0.01)	-0.03*** (0.01)	-0.06*** (0.02)	-0.07*** (0.02)
Crafts		0.07*** (0.01)	-0.03*** (0.01)	-0.06*** (0.01)	-0.07*** (0.02)
Machine operators		0.20*** (0.01)	0.15*** (0.01)	0.08*** (0.01)	0.08*** (0.02)
Elementary occupations		0.01 (0.01)	-0.05*** (0.01)	-0.06*** (0.01)	-0.07*** (0.02)
High tenure (>5 years)		0.17*** (0.01)	0.10*** (0.01)	0.05*** (0.01)	0.05*** (0.01)
Permanent labor contracts		0.04*** (0.01)	0.06*** (0.01)	0.06*** (0.01)	0.06** (0.03)
Hourly value added (log)			0.10*** (0.00)	0.01*** (0.00)	0.00 (0.00)
Capital (log)			0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Firm size (log)			0.03*** (0.00)	0.00 (0.00)	-0.00** (0.00)
Firm-level collective bargaining			0.03*** (0.00)	0.00 (0.00)	0.00 (0.00)
Brussels			0.01* (0.00)		0.01 (0.01)
Wallonia			-0.03*** (0.00)		-0.01 (0.01)
Constant	2.70*** (0.01)	2.45*** (0.02)	2.07*** (0.02)	2.64*** (0.02)	0.07*** (0.00)
Sector dummies	no	no	yes	no	yes
Year dummies	yes	yes	yes	yes	yes
Observations	30,355	30,355	30,355	30,355	12,421
R^2	0.07	0.63	0.7		0.30
R^2 within				0.35	
Under-identification test [c]					0.00
Weak-identification test [d]					56.88
Over-identification test [e]					0.11
Endogeneity test [f]					0.31

Notes: Heteroskedasticity and Autocorrelation Consistent (HAC) standard errors in parentheses. ***, **, and * denote significance at the 1, 5, and 10% levels, respectively. [a] AED(HDI): firm-level average Euclidean distance of human development index. [b] Reference categories are: share of men; education level 1 (ISCED 1-2); service occupations; share of fixed-term contracts; share of workers with more than 5 years of tenure; Flanders; manufacturing sector. [c] Under-identification test reports *p*-value of Kleibergen-Paap rk LM statistic. [d] Weak-identification test reports Kleibergen-Paap rk Wald F statistic. [e] Over-identification test reports *p*-value of Hansen J statistic. [f] Endogeneity test shows probability that endogenous regressors can be treated as exogenous.

Table 3: Separate estimations of wage-setting equations in firms with and without firm-level collective bargaining, 1999-2010

Dependent variable: average hourly wage (log)	Full sample	Firms with firm-level collective bargaining	Firms with no firm-level collective bargaining
	FE	FE	FE
	(6)	(7)	(8)
AED(HDI)	-0.15***	-0.13*	-0.17***
	(0.04)	(0.08	(0.04)
Share of women	-0.10***	-0.13***	-0.10***
	(0.02)	(0.03)	(0.03)
Workers < 40 years	-0.14***	-0.13***	-0.14***
	(0.01)	(0.02)	(0.01)
Education level 2 (ISCED 3-4)	0.12***	0.17***	0.10***
	(0.01)	(0.02)	(0.01)
Education level 3 (ISCED 5-7)	0.02***	0.01	0.01***
	(0.00)	(0.01)	(0.00)
Managers	0.51***	0.47***	0.54***
	(0.02)	(0.06)	(0.03)
Professionals	0.21***	0.24***	0.21***
	(0.01)	(0.03)	(0.02)
Technical assistance professionals	0.08***	0.08***	0.09***
	(0.01)	(0.03)	(0.02)
Clerical occupations	-0.06***	-0.06**	-0.06***
	(0.02)	(0.03)	(0.02)
Crafts	-0.06***	-0.06**	-0.06***
	(0.01)	(0.03)	(0.02)
Machine operators	0.08***	0.11***	0.08***
	(0.01)	(0.03)	(0.01)
Elementary occupations	-0.06***	-0.08***	-0.06***
	(0.01)	(0.03)	(0.02)
High tenure (>5 years)	0.05***	0.05***	0.05***
	(0.01)	(0.01)	(0.01)
Permanent labor contracts	0.06***	0.02	0.07***
	(0.01)	(0.03)	(0.01)
Hourly value added (log)	0.01***	0.01	0.01***
	(0.00)	(0.01)	(0.00)
Capital (log)	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)
Firm size (log)	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)
Constant	2.64***	2.78***	2.60***
	(0.02)	(0.06)	(0.03)
Sector dummies	no	no	no
Year dummies	yes	yes	yes
Observations	30,355	5,899	24,456
R^2 within	0.35	0.40	0.34

Notes: HAC standard errors in parentheses. ***, **, and * denote significance at the 1, 5, and 10% levels, respectively. For the rest, see the footnote of Table 2.

Table 4: Separate estimations of wage-setting equations in firms with low, high, and very high diversity, 1999-2010

Dependent variable: average hourly wage (log)	Full sample	Firms with low diversity [a]	Firms with high diversity [b]	Firms with very high diversity [c]
	FE (9)	FE (10)	FE (11)	FE (12)
AED(HDI)	-0.15*** (0.04)	-0.02 (0.11)	-0.19*** (0.06)	-0.50** (0.20)
Share of women	-0.10*** (0.02)	-0.11*** (0.03)	-0.08*** (0.02)	-0.09** (0.04)
Workers < 40 years	-0.14*** (0.01)	-0.14*** (0.01)	-0.10*** (0.02)	-0.10*** (0.04)
Education level 2 (ISCED 3-4)	0.12*** (0.01)	0.12*** (0.01)	0.09*** (0.02)	0.07 (0.04)
Education level 3 (ISCED 5-7)	0.02*** (0.00)	0.02*** (0.00)	0.01 (0.01)	0.01 (0.02)
Managers	0.51*** (0.02)	0.47*** (0.03)	0.63*** (0.06)	0.27* (0.14)
Professionals	0.21*** (0.01)	0.19*** (0.02)	0.27*** (0.04)	0.36*** (0.08)
Technical assistance professionals	0.08*** (0.01)	0.07*** (0.02)	0.11*** (0.03)	0.14*** (0.05)
Clerical occupations	-0.06*** (0.02)	-0.06*** (0.02)	-0.02 (0.02)	0.01 (0.03)
Crafts	-0.06*** (0.01)	-0.07*** (0.02)	-0.01 (0.02)	0.00 (0.04)
Machine operators	0.08*** (0.01)	0.08*** (0.02)	0.10*** (0.03)	0.12** (0.05)
Elementary occupations	-0.06*** (0.01)	-0.07*** (0.02)	-0.02 (0.02)	-0.01 (0.04)
High tenure (>5 years)	0.05*** (0.01)	0.05*** (0.01)	0.05*** (0.01)	0.06** (0.03)
Permanent labor contracts	0.06*** (0.01)	0.06*** (0.01)	0.04 (0.03)	0.01 (0.05)
Hourly value added (log)	0.01*** (0.00)	0.01** (0.00)	0.01 (0.01)	-0.01 (0.02)
Capital (log)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Firm size (log)	0.00 (0.00)	-0.00* (0.00)	0.00 (0.00)	0.02*** (0.01)
Firm-level collective bargaining	0.00 (0.00)	0.00 (0.00)	0.00 (0.01)	0.01 (0.01)
Constant	2.64*** (0.02)	2.67*** (0.03)	2.58*** (0.05)	2.54*** (0.10)
Sector dummies	no	no	no	no
Year dummies	yes	yes	yes	yes
Observations	30,355	22,767	7,588	1,517
R^2 within	0.35	0.33	0.40	0.43

Notes: HAC standard errors in parentheses. ***, **, and * denote significance at the 1, 5, and 10% levels, respectively. [a] Firms with low diversity are defined as those with AEDs below percentile 75. [b] Firms with high diversity are defined as those with AEDs between percentiles 75 and 95. [c] Firms with very high diversity are defined as those with AEDs above percentile 95. For the rest, see the footnote of Table 2.

Appendices

Appendix A: Sample selection and descriptive statistics

Table A.1: Distribution of firms by sectors of activity in the initial and restricted SES-SBS-NR samples

Type of economic activity	Initial SES-SBS-NR sample	Restricted SES-SBS-NR sample
Mining and quarrying (C)	0.00	0.01
Manufacturing (D)	0.21	0.45
Electricity, gas and water supply (E)	0.00	0.00
Construction (F)	0.15	0.11
Wholesale and retail trade, repair of motor vehicles, motorcycles and personal and household goods (G)	0.29	0.19
Hotels and restaurants (H)	0.08	0.02
Transport storage and communication (I)	0.08	0.07
Financial intermediation (J)	0.02	0.01
Real estate, renting and business activities (K)	0.17	0.13
Observations	49,944	30,355

Notes: Descriptive statistics for the period 1999-2010. The restricted sample is the one used to run the OLS and FE estimates. Notably, firms not observed during at least two consecutive periods have been dropped (for more details on the selection criteria, see Section 3.1).

Table A.2: Diversity indicators in the initial and restricted SES-SBS-NR samples

Variables	Initial SES-SBS-NR sample		Restricted SES-SBS-NR sample	
	Mean	Std. Dev.	Mean	Std. Dev.
Human Development Index (HDI)	0.87	0.029	0.87	0.028
AED(HDI) [a]	0.026	0.036	0.025	0.035
SD(HDI) [b]	0.034	0.040	0.034	0.040
CV(HDI) [c]	0.041	0.050	0.041	0.049
Observations	49,944		30,355	

Notes: Descriptive statistics for the period 1999-2010. The restricted sample is the one used to run the OLS and FE estimates. Notably, firms not observed during at least two consecutive periods have been dropped (for more details on the selection criteria, see Section 3.1). [a] AED: average Euclidean distance. [b] SD: standard deviation. [c] CV: coefficient of variation.

Appendix B: Additional descriptive statistics

This section provides additional descriptive statistics on the composition of the workforce by origin within firms (and relative HDIs). Given that the number of nationalities at birth in the sample is huge, we have regrouped them in eight broad geographical regions (i.e., Western Europe, North Africa, Eastern Europe, Sub-Saharan Africa, Asia, Latin America, North America, and South Pacific), plus Belgium as a separate category.

Table B.1 shows the number of individual observations and the average HDIs for these regions in the sample. The average HDI ranges from 0.92 for workers born with a nationality from South Pacific (a result driven by Australia and New Zealand) to 0.54 for those originating from Sub-Saharan Africa.

Table B.1: Average and standard deviation of HDIs by nationality at birth (broad geographical regions), 1999-2010

Nationalities at birth grouped by regions	Observations	Average HDI	St. Dev. HDI
Belgium	75,7584	0.88	0.00
Western Europe (except Belgium)	89,976	0.88	0.02
North Africa	42,426	0.66	0.06
Eastern Europe (including former Soviet Union)	10,219	0.79	0.03
Sub-Saharan Africa	5,602	0.54	0.08
Asia	4,863	0.65	0.11
Latin America	1,761	0.73	0.07
North America	717	0.91	0.00
South Pacific	83	0.92	0.01

Notes: HDI stands for "human development index".

The same regional grouping can be used to shed light on the intra-firm composition of the workforce and its link with the average HDIs of firms. Table B.2 shows the average composition of firms (in terms of workers' nationalities at birth grouped by regions) according to the HDI quintile to which they belong. As one might expect, firms with the lowest HDIs (i.e., belonging to the first HDI quintile) employ relatively few Belgian-born workers (60%) and more foreigners. A striking outcome is the overrepresentation of immigrants from North Africa, Eastern Europe, and Sub-Saharan Africa in this quintile; all the three groups appear to be clustered in firms with low average HDIs. The highest share of Belgians is found in the fourth quintile, in which the average firm employs only 3% of foreigners. The top quintile is marked by a strong presence of Western Europeans and the absence of non-EU15 immigrants.

Table B.2: Composition of the workforce by nationality at birth (broad geographical regions) and firm-level HDI quintiles, 1999-2010

	Firm-level HDI quintiles				
Nationalities at birth grouped by regions	1st	2nd	3rd	4th	5th
Belgium	0.60	0.80	0.83	0.97	0.89
Western Europe (except Belgium)	0.14	0.12	0.14	0.03	0.11
North Africa	0.17	0.05	0.01	0.00	0.00
Eastern Europe (including former Soviet Union)	0.03	0.02	0.01	0.00	0.00
Sub-Saharan Africa	0.03	0.00	0.00	0.00	0.00
Asia	0.02	0.01	0.00	0.00	0.00
Latin America	0.01	0.00	0.00	0.00	0.00
North America	0.00	0.00	0.00	0.00	0.00
South Pacific	0.00	0.00	0.00	0.00	0.00

Notes: HDI stands for "human development index".

The diversity indicator used in this paper is based on the average Euclidean distance between HDIs of all the individual workers in each firm. To show how this indicator relates to the composition of the workforce within each firm, we have divided firms by AED(HDI) quintiles. As shown in Table B.3, the bottom quintile displays no diversity at all, that is, all the members of the workforce are born with a Belgian nationality and are, therefore, associated with the same HDI. In the second and third quintiles, the share of Western Europeans increases but only a few non-EU15 immigrants are employed, which explains why the overall HDI-diversity remains low. In the fourth and especially the fifth quintile, we see a strong presence of workers born with a non-EU15 nationality, especially from North Africa, Eastern Europe, and Sub-Saharan Africa.

Table B.3: Composition of the workforce by nationality at birth (broad geographical regions) and firm-level AED(HDI) quintiles, 1999-2010

	Firm-level AED(HDI) quintiles				
Nationalities at birth grouped by regions	1st	2nd	3rd	4th	5th
Belgium	1.00	0.91	0.82	0.79	0.61
Western Europe (except Belgium)	0.00	0.09	0.16	0.12	0.14
North Africa	0.00	0.00	0.01	0.05	0.16
Eastern Europe (including former Soviet Union)	0.00	0.00	0.01	0.02	0.03
Sub-Saharan Africa	0.00	0.00	0.00	0.00	0.03
Asia	0.00	0.00	0.00	0.01	0.02
Latin America	0.00	0.00	0.00	0.00	0.01
North America	0.00	0.00	0.00	0.00	0.00
South Pacific	0.00	0.00	0.00	0.00	0.00

Notes: AED(HDI) is the firm-level average Euclidian distance of HDIs.

Appendix C: Estimates using the standard deviation of HDIs within firms

Table C: Estimations of firm-level wage-setting equations using the standard deviation of HDIs instead of the average Euclidean distance

Dependent variable: average hourly wage (log)	OLS (1)	OLS (2)	OLS (3)	FE (4)	GMM-IV (5)
SD(HDI) [a]	-0.71***	-0.00	-0.05*	-0.11***	-0.40**
	(0.06)	(0.03)	(0.03)	(0.03)	(0.18)
Control variables:					
Year dummies	yes	yes	yes	yes	yes
Worker and job characteristics	no	yes	yes	yes	yes
Firm characteristics	no	no	yes	yes	yes
Value added	no	no	yes	yes	yes
Observations	30,355	30,355	30,355	30,355	12,421
R^2	0.06	0.63	0.70		0.30
R^2 within				0.35	
Under-identification test					0.00
Weak-identification test					96.04
Over-identification test					0.11
Endogeneity test					0.16

Notes: HAC standard errors in parentheses. ***, **, and * denote significance at the 1, 5, and 10% levels, respectively. [a] SD(HDI): firm-level standard deviation of human development index. For the rest, including the various types of controls in the different specifications, see the footnote of Table 2.

Appendix D: Estimates using a binary explanatory variable

Table D: Estimations of firm-level wage-setting equations using the share of immigrants instead of the AED(HDI)

Dependent variable: average hourly wage (log)	OLS (1)	OLS (2)	OLS (3)	FE (4)	GMM-IV (5)	FE (6)	GMM-IV (7)
Share of hours worked by:							
EU15 workers [a]	*Reference*	*Reference*	*Reference*	*Reference*	*Reference*		
Non-EU15 workers	-0.24*** (0.01)	-0.05*** (0.01)	-0.03*** (0.01)	-0.04*** (0.01)	-0.06*** (0.01)		
Belgian workers [b]						*Reference*	*Reference*
Non-Belgian workers						0.02 (0.02)	-0.11** (0.05)
Control variables:							
Year dummies	yes	yes	yes	yes	yes	yes	yes
Worker and job characteristics	no	yes	yes	yes	yes	yes	yes
Firm characteristics	no	yes	yes	yes	yes	yes	yes
Value added	no	no	yes	yes	yes	yes	yes
Observations	30,355	30,355	30,355	30,355	12,421	30,355	12,421
R^2	0.06	0.62	0.67		0.65		0.30
R^2 within				0.66		0.35	
Under-identification test					0.00		0.00
Weak-identification test					30.10		191.78
Over-identification test					0.33		0.14
Endogeneity test					0.16		0.00

Notes: HAC standard errors in parentheses. ***, **, and * denote significance at the 1, 5, and 10% levels, respectively. [a] Share of workers born in the EU15. [b] Share of workers born with a Belgian nationality. For the rest, including the various types of controls in the different specifications, see the footnote of Table 2.

Appendix E: Estimates based on a polarization index

The concept of polarization might provide some additional insights into the dynamics underlying wage discrimination against immigrants. Based on the concept of polarization introduced by Esteban and Ray (1994), Montalvo and Reynal-Querol (2005) developed an index of ethnic polarization, which is broadly used in applied research on ethnic diversity (Alesina and La Ferrara, 2005). The polarization index they propose is the following:

$$PI = 1 - \sum_{i=1}^{N} \left(\frac{1/2 - s_i}{1/2}\right)^2 s_i, \quad (2)$$

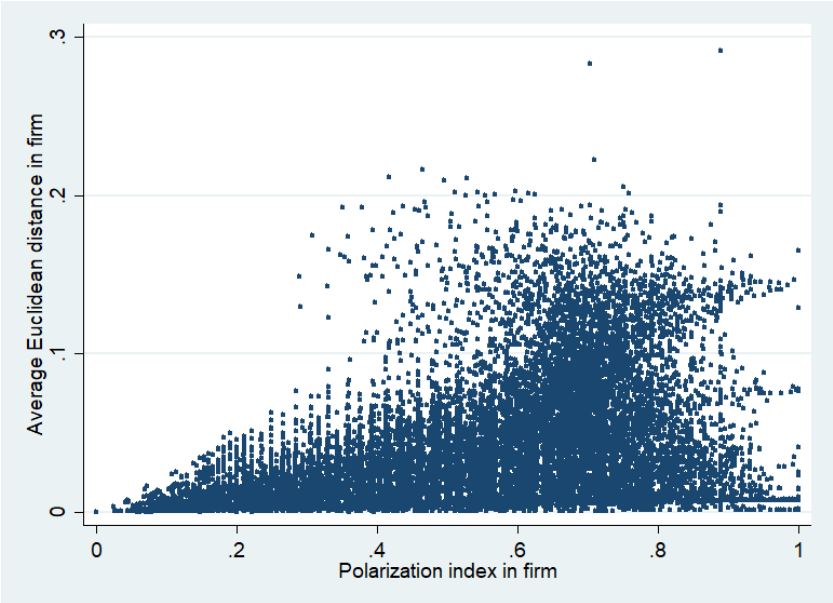
where s_i is the proportion of a given ethnic group over the total population, and N is the total number of ethnic groups in the population. In this paper, we compute such polarization index at the firm-year level. A given PI characterizes each firm in each year, s_i is simply the share of a given ethnic group (operationalized as the nationality at birth) over the firm's total workforce, and N is the total number of ethnic groups in the firm.

As explained in Montalvo and Reynal-Querol (2005), this polarization index reaches its maximum when two equally-sized ethnic groups face each other (for instance, when the firm's workforce is composed of 50% of Belgians and 50% of Italians), whereas it declines as the configuration of ethnic groups differs more and more from such half and half split. Therefore, polarization is low when the firm's workforce is mainly composed of Belgian workers, and it is similarly low when the firm's workforce is much diversified, that is, when many different ethnic groups coexist.

This feature makes the polarization index radically different from our measure of ethnic diversity – the AED(HDI) index –, which, in essence, is a fractionalization index. Differently from the polarization index, AED(HDI) gets larger as more ethnic groups (actually, socio-economic conditions of ethnic groups) face each other. Said differently, AED(HDI) increases as the workforce is more fractionalized.

That the polarization index and our (fractionalization-type) index are not the same can be seen in Figure E, which shows how those indices relate to each other. While the (linear) correlation coefficient is positive and somewhat large in magnitude (0.57), the relationship is clearly non-linear. In particular, when the polarization index is high, the AED(HDI) index is generally low. Similarly, among firms with high or very high AED(HDI) indices, the polarization index takes a variety of values ranging from around 0.3 to 1.

Figure E: The polarization index compared to the AED(HDI) index at the firm level



Given that the polarization index captures radically different information compared to AED(HDI), we do not expect to see similar estimation results in the two cases. This is even more true if one considers that our results show that discrimination only occurs in firms characterized by high fractionalization (i.e., high values of AED(HDI)), that is, when the polarization index is not correspondingly high. The first five specifications of Table E replicate the main results (i.e., those reported in Table 2), but inserting the polarization index instead of the AED(HDI) index. The estimated coefficients associated with the polarization index are mostly small in magnitude and not significant, except for two (basic) OLS specifications. The FE estimates, our preferred ones given that the GMM-IV tests do not indicate endogeneity issues, point to a very small and not significant impact of ethnic polarization on wages. In addition, Column (6) of Table E reports our preferred FE estimates for the case in which we insert both the polarization index and the AED(HDI) index. The results confirm a not significant impact of polarization, and show that ethnic fractionalization – captured by AED(HDI) – has a strongly significant impact on wages, as we found in the main estimates. These results support the idea that migrants are discriminated against when they are more fractionalized, whereas high polarization (i.e., high shares of a particular migrant ethnic group opposed to correspondingly high shares of Belgian workers) is not found to generate discrimination against migrants. Thus, these results cast further light on the fact that *l'union fait la force*: when migrants are not fractionalized, no discrimination is detected. As shown in Table B.3, high polarization is likely to reflect a situation where Belgians and other Western Europeans co-work within firms (with almost no other immigrants), while high fractionalization is most probably encountered within firms employing less Belgian-born workers and many immigrants originating from Western Europe but also from North Africa, Eastern Europe, Sub-Saharan Africa, Asia, and Latin America.

Table E: Estimations of firm-level wage-setting equations using a polarization index as the main explanatory variable instead of (or in complement with) the AED(HDI) index

Dependent variable: average hourly wage (log)	OLS	OLS	OLS	FE	GMM-IV	FE
	(1)	(2)	(3)	(4)	(5)	(6)
Polarization index [a]	-0.06***	-0.00	0.02***	-0.01	-0.08	0.00
	(0.01)	(0.01)	(0.00)	(0.01)	(0.05)	(0.01)
AED(HDI)						-0.17***
						(0.04)
Control variables:						
Year dummies	yes	yes	yes	yes	yes	yes
Worker and job characteristics	no	yes	yes	yes	yes	yes
Firm characteristics	no	no	yes	yes	yes	yes
Value added	no	no	yes	yes	yes	yes
Observations	30,355	30,355	30,355	30,355	12,421	30,355
R^2	0.05	0.63	0.70		0.30	
R^2 within				0.35		0.35
Under-identification test					0.00	
Weak-identification test					30.10	
Over-identification test					0.33	
Endogeneity test					0.16	

Notes: HAC standard errors in parentheses. ***, **, and * denote significance at the 1, 5, and 10% levels, respectively. [a] Polarization index: firm-level polarization index computed based on Equation (2). For the rest, including the various types of controls in the different specifications, see the footnote of Table 2.