INSTITUTE OF ECONOMICS
Centre for Economic and Regional Studies

Közgazdaság- és Regionális Tudományi Kutatóközpont
KÖZGAZDASÁG-TUDOMÁNYI INTÉZET

KRTK-KTI WORKING PAPERS | KRTK-KTI MŰHELYTANULMÁNYOK

Wage Structure and Inequality: The role of observed and unobserved heterogeneity

ISTVÁN BOZA

KRTK-KTI WP – 2021/31

August 2021

https://kti.krtk.hu/wp-content/uploads/2021/08/KRTKKTIWP202131.pdf

ABSTRACT

This study aims to contribute to the literature of firms and occupations as prominent drivers of wage-inequality in multiple ways. First, we synthesize novel modelling approaches of recent studies in the field and use administrative linked employer-employee panel data from an Eastern European country, Hungary, to assess the contribution of individual, firm and job heterogeneity – and their interactions – to overall wage inequality. Consistent with earlier findings from Western Europe, Scandinavia, the US and Brazil, we show that firm heterogeneity provides around 22%, individual heterogeneity 50%, and occupational heterogeneity 8% of overall wage dispersion, with wage sorting between firms and individuals in itself explaining around 9%. Notably, around half of this contribution is accountable to observable sub-components of individual and firm wage effects. Also, the same magnitude of assortativity can be found between individuals and occupations. Utilizing unique features of our data, we compare mathematics and literature test score records of 10th grade students to their future labor market outcomes, finding a positive correlation between test scores and future firm value added, a direct evidence for assortative matching in productivity. Finally we assess sorting along observable characteristics, such as gender, education, occupation or age of workers, and the ownership of employers.

JEL codes: J31, D63, I24

Keywords: wage inequality, wage decomposition, fixed effects, linked employer-employee data, sorting, assortative matching

István Boza
Department of Economics and Business at Central European University,
Nádor street 13., 1051 Budapest, Hungary.
e-mail: bozaistvan@gmail.com

Bérstruktúra és béregyenlőtlenség: a megfigyelt és nem megfigyelt különbözőségek szerepe

BOZA ISTVÁN

ÖSSZEFOGLALÓ

E tanulmány célja azon munkagazdaságtani irodalomhoz hozzájárulni, mely fókuszában a cégek és foglalkozások állnak, mint a béregyenlőtlenségek fontos hajtóerői. A szakirodalomban a közelmúltban megjelent új modellezési irányokat egységesítve, magyar adminszitratív kapcsolt foglalkoztatott-foglalkoztató adatokat használva megvizsgáljuk az egyéni, cégek közötti, illteve foglalkozások közti különbözőségek – illetve az ezek közötti kapcsolatok – hozzájárulását a gazdaság-szintű béregyenlőtlenségekhez. Hasonlóan korábbi, nyugat-európai, skandináv, amerikai és brazil eredményekhez, azt találjuk, hogy a céges különbségek 22%-kal, az egyéni különbségek 50%-kal, míg a foglalkozások 8%-kal járulnak hozzá a teljes bérszóródáshoz, miközben az egyének és vállalatok közti bérszelekció is több mint 9%-át adja a teljes egyenlőtlenségnek. Utóbbinak közel fele tudható be megfigyelhető egyéni és vállalati tényezők közötti szelekcióknak. Az egyének és foglakozások közötti nem véletlen kiválasztódás is hasonló mértékben járul hozzá a bérszóráshoz. A magyar adat egy egyediségét kihasználva összevetjük a fiatal munkavállalók korábban, tizedik osztályban, az Országos Kompetenciamérésen elért matematika és szövegértés eredményeit a későbbi munkaerőpiaci kimeneteikkel. Az általunk talált pozitív korreláció ezen teszteredmények és a későbbi munkáltató cégek vállalati hozzáadott érték mutatói között közvetlen bizonyítékként szolgálnak a termelékenységekre vonatkozó pozitív asszortatív párosítás jelenlétére. A tanulmány végén megvizsgáljuk egyes megfigyelhető jellemzők, úgy mint a munkavállalók neme, végzettsége, kora és foglalkozása avagy a munkáltató tulajdonosi szerkezete mentén jelentkező szelekciós csatornákat.

JEL: J31, D63, I24

Kulcsszavak: béregyenlőtlenség, bérdekompozíció, fix hatások, kapcsolt foglalkoztatott-foglalkoztató adat, bérszelekció, asszortatív párosítás

Wage Structure and Inequality: The role of observed and unobserved heterogeneity

István Boza[1]

[1]Central European University, Budapest
August 3, 2021

This study aims to contribute to the literature of firms and occupations as prominent drivers of wage-inequality in multiple ways. First, we synthesize novel modelling approaches of recent studies in the field and use administrative linked employer-employee panel data from an Eastern European country, Hungary, to assess the contribution of individual, firm and job heterogeneity – and their interactions – to overall wage inequality. Consistent with earlier findings from Western Europe, Scandinavia, the US and Brazil, we show that firm heterogeneity provides around 22%, individual heterogeneity 50%, and occupational heterogeneity 8% of overall wage dispersion, with wage sorting between firms and individuals in itself explaining around 9%. Notably, around half of this contribution is accountable to observable sub-components of individual and firm wage effects. Also, the same magnitude of assortativity can be found between individuals and occupations. Utilizing unique features of our data, we compare mathematics and literature test score records of 10^{th} grade students to their future labor market outcomes, finding a positive correlation between test scores and future firm value added, a direct evidence for assortative matching in productivity. Finally, we assess sorting along observable characteristics such as gender, education, occupation or age of workers, and the ownership of employers.

Keywords: wage inequality, wage decomposition, fixed effects, linked employer-employee data, sorting, assortative matching

JEL codes: J31, D63, I24

Corresponding author: István Boza, bozaistvan@gmail.com

The author would like to thank for the technical support of the Databank of the Centre for Economic and Regional Studies (formerly of the Hungarian Academy of Sciences) and for all the comments and suggestions of Márton Csillag, Andrea Weber, Dániel Horn, Attila Gyetvai, Balázs Reizer, participants of CEU PhD research seminars, seminars at IE-CERS and the XIV. Annual Conference of the Hungarian Society of Economists.

1 Introduction

For more than two decades now, labor economists have been intrigued by whether systematically high wage (or high productivity) workers tend to work at high wage firms. The seminal work of Abowd et al. (1999) – AKM, after the authors' initials – was the first, shortly followed by Goux and Maurin (1999), to propose a model in which wages are log additive in time-invariant individual and firm characteristics and time-varying factors. Using linked employer-employee panel data, these time-invariant (partly unobservable) characteristics can be captured by worker and firm fixed effects respectively, with the latter capturing wage differences among firms, controlled for the composition of their workforce with regard to both observable and unobservable worker skills. The steadily increasing availability of such data – regarding both the number of countries, detailedness, and the length of panels – and advances in econometric concerns regarding the estimation of multi-way, high dimensional fixed effect models gave rise to a series of labor studies, which aim to decompose the overall wage dispersion into differences coming from heterogeneity in the above listed observed (and unobserved) factors. And although AKM effects have been used in studies from a wide range of fields as measures of firm and worker quality, for instance in estimating inter-industry wage differentials, rent-sharing estimations or even job referral effects (Abowd et al., 2019), they may had the most influential effect on the literature of wage and earnings inequalities. Our study contributes to this literature not only by presenting evidence for another country where the sorting of high wage workers to high wage firms is a substantial element of overall wage dispersion, but also by aiming to uncover potential channels along which this phenomenon emerges. Although most exercises in this study are descriptive in nature, the methods and results presented may further the understanding of determinants of such wage sorting.

Along the natural role of individual diversity in skills, opportunities and ambitions, the heterogeneity of firms' waging schemes – originating in differences in firm productivity or the rent sharing propensity of firms, in compensating differentials or in reliance on efficiency wages – can be an important source of wage variation in the economy in itself. Besides, it may also affect the overall wage dispersion through a sorting channel as well. If positive assortative matching with regard to worker and firm productivity is present in the labor market due to complementarity in production, we would also expect 'high wage' (high productivity) workers to be systematically over-represented in 'high wage' (high productivity) firms. That is, if the individual and firm fixed effects of the AKM model capture underlying productivity differences, then the estimated fixed effect parameters should positively correlate. Although early studies found no or negative such correlation (Abowd et al., 2002; Goux & Maurin, 1999; Gruetter & Lalive, 2009; Iranzo et al., 2008; K. L. Sørensen & Vejlin, 2011; Woodcock, 2008), it had been showed that the variance and covariance terms of the estimated worker and firm effects are affected by an incidental parameter problem, labeled "limited mobility bias" (Andrews et al., 2008, 2012). The lack of observed mobility in the panel data used – on which identification of firm effects rely – and the mechanical negative relation of sampling errors in person and firm effects cause a serious downward bias in the above correlation, especially in short panels or sub-samples, possibly driving the zero or negative results found in early studies.[1]

The view on sorting was changed by the defining study of Card et al. (2013), being the first to show a critical, positive role of wage sorting in overall wage dispersion. Besides, the authors found that the dispersion of firm effects and the correlation between workers and firm effects do not only explain a substantial part of wage variance in a given period, but their increase also critically contributed to the observed increase in wage inequality in West Germany over the period of 1985-2009. The wage decomposition approach proposed by Card et al. (2013) have been reproduced by many studies to follow, including most notably Card et al. (2016), Card et al. (2018) and Torres et al. (2018) for Portugal, and Gerard et al. (2021) for Brazil, Song et al. (2019) and Lamadon et al. (2019) for the US. The findings of these and a handful of other studies are summarized in Appendix Table A1. An important takeaway from the table is that most studies of this decade find a 10-30% contribution of firm heterogeneity, and around a 10-15% contribution of wage sorting to overall wage variance – results from Italy being the exception with near zero sorting components. Studies from the last couple of years, which develop and apply bias-correction methods for the limited mobility bias of the AKM framework, such as Kline et al. (2020) for Italy, and Bonhomme et al. (2020) for the US, Austria, Norway, Sweden and Italy find systematically larger correlations of firm and person effects, larger contribution of the sorting component and lower contribution of the variance in the firm component itself – as predicted by the nature of this bias.[2] The similarity in wage composition, even among these similarly developed, but institutionally different countries is quite fascinating. Yet, there are no published results for Eastern European / post-transition countries that we know of. [3] The results presented for Hungary in this paper, however, will be largely in line with those of the aforementioned authors, further expanding the set of countries with similar wage dispersion structure. Besides, we will adapt and build upon some of the novel extensions of the AKM framework from research of the past half decade.

Beside the aforementioned econometric issue, the common measure of wage sorting in the AKM framework has been a target of criticism from a theoretical standpoint as well. Based on a branch of studies, Torres et al. (2018) argue for the importance of differentiating wage sorting from productivity sorting or assortative matching, with the latter term having its origins in the technological

[1]We will reflect on this issue in more detail throughout the study.

[2]Appendix Table A2 present bias-corrected and standard results from the same studies. Comparing consecutive rows in the table reveals that bias-corrected estimates include, on average, 6% lower firm shares and 10% higher sorting shares, with substantially higher correlations, even in the range of 0.3-0.4.

[3]The only, unpublished exception being Gyetvai (2017), who uses an earlier iteration of our dataset, consisting of 8 years only, and replicates the ensemble decomposition of Card et al. (2018), finding a 26.5% importance share for firms, 60.3% for workers and 5.1% for occupations as a third source of heterogeneity.

complementaries between the productivity of firm and its workforce. The main motivation for the distinction is that, while the wages of workers are expected to be monotonic with regard to their skills, the same may not be true for firm productivity (Eeckhout & Kircher, 2011; Lopes De Melo, 2018). However, using data from financial reports, and estimating production functions for the firms – controlling for skill/ occupational composition – Torres et al. (2018) show that these, directly estimated measures of firm productivity also correlate with worker effects, similarly to the indirect productivity measure of AKM firm effects. The correlations are even stronger, suggesting that non-monotonicities in the productivity-wage relation of firms are not negligible.[4] Using balance sheet data of incorporated firms, we will reinforce these findings. In addition, we use data on high school participation and test scores, although for a limited sample only, to propose direct measures of productivity sorting, showing the sorting of high-achievers at teenage years to high wage employers.

Besides firm heterogeneity, the heterogeneity of occupations can be an important aspect of wage formation as well. For instance, we could observe different enumeration levels of different occupations even for workers of the same skills as different jobs can bear different outside options or due to compensating differentials for occupation-specific amenities or disamenities. Still, even firms with the same occupational composition can pay on average different premia for all of their workers, so the distinction of firm and occupation heterogeneity can be really important, as the sorting of high wage workers into specific occupations and the clustering of such occupations in high wage firms could both increase the level of inequality, while the joint presence of these phenomena could even confound standard measures of wage sorting (between workers and firms). For similar considerations, Card et al. (2016) and Torres et al. (2018) introduce a third high dimensional fixed effect in the form of "job-title effects", for decomposing the gender wage gap and the overall wage variation into person, firm and occupational components in Portugal respectively.[5] As in Portugal collective bargaining agreements cover most of the work force, these authors define job titles as an occupational category under a given collective agreement, thus allowing occupations in different sectors having different average effects. Using this design, they show that not only the type of the firm and the person matters in wage determination, but indeed the type of work done by the individual as well. Our estimations will also incorporate this approach, although only controlling for occupations as a third high dimensional effect.

[4]Card et al. (2016) and Card et al. (2018) also utilize financial report data, and find such positive correlations. Moreover, both paper, and also Alvarez et al. (2018), use the estimated AKM effects to propose a measure for rent-sharing elasticities, regressing AKM firm effects (instead of average wages) on firm productivity. This measure of elasticity removes the effects of workforce composition of firms, and hence may capture true rent-sharing behaviour better than one derived from average wage levels.

[5]Later, Cardoso et al. (2018) and Addison et al. (2018) also use the same decomposition method as Cardoso et al. (2016), building on Gelbach (2016), to decompose the union-membership wage-gap and the returns of education in Portugal into occupational, individual, firm and match effect components.For Hungary, Gyetvai (2017) presented preliminary results from a wage variance decomposition on a shorter panel.

While the firm and person effects of the main AKM equations absorb any time-invariant firm or person characteristics, these effects could be further decomposed into elements explained by observable, time-invariant characteristics and an unexplained components as shown by Abowd et al. (1999), and applied by for instance Woodcock (2008) for wage-gap and Gruetter and Lalive (2009) or Torres et al. (2018) for variance decompositions. And although the latter two papers do report the full correlation structure of observed and unobserved wage components, we are the first to directly interpret the shares of the the subcomponents of sorting covariance attributable to observable, partly-observable and fully unexplained factors.

Another important question related to observable characteristics in the AKM model is whether firm effects are stable across time or groups of workers. A detailed assessment of the former problem and a model with time-varying, firm-year fixed effects is presented by Lachowska et al. (2020). Firms, however, may also pay differing premia of workers of different observable characteristics, for instance due to differences in bargaining power and the firms' rent-sharing propensity. By introducing differing firm(-group) effects or firm effects based on race and gender categories, Card et al. (2016) and Gerard et al. (2021) propose a way to decompose the differences in the average firm effects faced by ethnic or gender groups into a bargaining (within-firm) and a sorting component. And while a sorting parameter with respect to observable characteristics could be also captured by decomposing gaps in the standard or three-way AKM model, as in Cardoso et al. (2016), these flexible models may yield more precise estimates through not assuming wage-gaps to be constant across all employers. Our finding that only half of the sorting covariance is attributable to unexplained wage components, motivates us to adapt a slightly modified version of the above models for assessing bargaining and sorting differences across workers of different gender, education, occupation, age or tenure – estimating some novel AKM specifications in the process.[6]

By adapting the above listed extensions into the models we use, we aim to contribute to the literature of wage inequalities in more than one ways, with the following main findings. We start by providing evidence on another country where positive wage sorting is strongly prevalent. Although such results are already available from a handful of countries from Western Europe, Scandinavia and also from the US and Brazil, Hungary is the first Eastern European, post-soviet country to present such estimates. Surprisingly similar wage structure patterns are found to those from the countries above, further reinforcing the emerging pattern across studies, that labor markets tend to behave similarly in a wide-range of countries with different historical and institutional backgrounds. While the overall contribution of individual heterogeneity is around 50%, of firm heterogeneity 22% and of occupational heterogeneity only 8%, sorting channels turn out to be rather important. The estimated correlation between person and firm effects is 0.18, with the underlying covariance explaining 9.3% of overall

[6]Another branch – to which we do not relate in this study – investigates the role of compensating differentials as another source of firm wage heterogeneity(Lamadon et al., 2019; Sorkin, 2018).

wage variation, while the sorting of high wage workers to high wage occupation also responsible for 10.7%. Exploiting data on firms' financial reports, we also reinforce the findings of Torres et al. (2018) about the relation between wage sorting and actual matching in productivity. Notably, we find that worker heterogeneity captured by person effects is indeed correlated with the observed value added of firms, not just the assumed productivity differences reflected in wage levels.[7] We also utilize the 10^{th} grade results of young workers on The National Assessment of Basic Competences, to assess whether individual literacy and mathematics skills – measured at around the age of 16 – correlate with future worker wage or firm productivity. We find that both absolute and relative, within-school test scores move together with the worker effects, occupation effects, firm effects and firm value added as well. This latter correlation – estimated to be around 0.12-0.14 – is a direct evidence for assortative matching, with value added capturing firm productivity and test scores proxying expected worker productivity.[8]

To better understand the origin of wage sorting, we focus on sorting related to observable characteristics – accounting for half of the overall sorting in the Hungarian labor market. First, following the methodology of Cardoso et al. (2016), we decompose some of the most prevalent wage gaps into individual, firm-specific and occupation-specific components. Doing so we show that within-firm gender-based, educational or residential wage differences can be indeed exaggerated by sorting and segregation mechanisms as well. Also, we reflect on the different selection of workers based on ownership of the hiring firm, finding that multinational employers are substantial contributors to the relatively high wage sorting in Hungary, as besides paying high wages generally, they are also able to hire the most skilled workers as well. Finally, using grouped-AKM specifications that allow for differing firm-effects for workers of different observable characteristics, we also present evidence for workers of different occupations and age sorting into firms of different average wage-premia, also amplifying the corresponding within-firm wage differences.

The paper is structured as follows. Section 2 presents our take on the wage variance and wage-gap decomposition techniques established in recent literature. Section 3 discusses the sources of data. Section 4 contains the main wage variance decomposition and discusses direct and indirect measures of assortative matching, while Section 5 pursues observable patterns of sorting, by utilizing wage-gap decompositions and alternative specifications of the AKM model. Section 6 concludes.

[7] Correlation is 0.37, but the difference is partially due to higher lever of sorting among firms with balance sheet data, as on the sample of incorporated firms the baseline correlation is also higher, 0.31.

[8] Due to data limitations, this correlation relates to the sample of young workers only, for whom wage sorting measured by AKM effects is substantially weaker.

2 Wage models and decompositions

2.1 The log-additive model of wages

Building upon Abowd et al. (1999), Card et al. (2013) and Torres et al. (2018), let us consider the following, log-additive model of wages:

$$ln\ w_{ijt} = \boldsymbol{X_{ijt}\beta} + \theta_i + \psi_j + \lambda_{k(ijt)} + \varepsilon_{ijt} \tag{1}$$

where w_{ijt} is the wage of person i working for employer j in occupation k at time t. $\boldsymbol{X_{ijt}}$ consists of observable, time-varying characteristics (age, firm size, year), and the other terms are the time-invariant worker, firm and occupation effects respectively, with a zero-mean, independent residual term added. That is, we consider occupations as a third high-dimensional fixed effect instead of $\boldsymbol{X_{ijt}}$ containing hundreds of occupation dummies.[9] If such model is estimated by OLS, the effect of time-invariant characteristics of individuals and firms are absorbed by the fixed effects, and can only be obtained by running second stage regressions on the estimated fixed effect parameters.[10] Specifically one can estimate:

$$\hat{\theta}_i = \boldsymbol{W_i\eta} + \varepsilon_i^I \tag{2}$$

$$\hat{\psi}_j = \boldsymbol{Z_j\gamma} + \varepsilon_j^J \tag{3}$$

In these second stages[11] , $\boldsymbol{W_i}$ contains time-invariant, although observable characteristics of the workers, like gender, birth cohort or highest achieved education, and an estimated $\hat{\varepsilon}_i^I$ will reflect directly unobservable individual heterogeneity. Similarly $\boldsymbol{Z_j}$ contains observable firm characteristics, like industry of operations or majority ownership, while $\hat{\varepsilon}_j^J$ will incorporate the unobservable factors defining the wages of the firm, such as reliance on specific waging schemes or rent-sharing from productivity.[12] This two-stage model will serve as the basis of most of the exercises presented in this study.

[9] Torres et al. (2018) argues that using even highly detailed occupations may not be ideal, as occupational wage standards may be different across different sectors. For instance, a secretary of an IT firm may not face the same occupational wage standard as a secretary in an assembly firm. As in Hungary sectoral collective agreements are not as prevalent, and as we also lack such data, we rely only on occupations, but include sector-occupation interactions in one of our tests for model robustness.

[10] For instance the Stata routine of Correia (2017), *reghdfe* implements the estimation of such multi-way high dimensional fixed effects model, based on the algorithm of Guimarães and Portugal (2010). The connected set on which firm, person and occupation effects are not only identified but are also comparable is more restricted than in the two-way fixed effect case, and has to be defined according to the algorithm of Weeks and Williams (1964), as noted by both Cardoso et al. (2016), Gyetvai (2017), and Torres et al. (2018)

[11] The concept of which is already present in Abowd et al. (1999) and later utilized, for instance, by Woodcock (2008), Gruetter and Lalive (2009), T. Sørensen and Vejlin (2013), Torres et al. (2018) and Alvarez et al. (2018).

[12] Technically some variables, such as industry can be considered and estimated as fixed effects themselves, but for the sake of simplicity we assume all observable characteristics as part of $\boldsymbol{W_i}$,$\boldsymbol{Z_j}$ or $\boldsymbol{X_{ijt}}$.

2.2 Wage variance decompositions

Through the past decade, labor economists decomposed the variance of wages in multiple different ways. In this sub-chapter, we present the established approaches and link them to the above-presented three-way fixed effects model. For the sake of notational simplicity, we omit the subscripts/indices of the wage components, with all corresponding to their respective terms defined in Equations 1-3. The most simple decomposition of the variance of wages within a two-way fixed effect framework can be found, among others, in Card et al. (2013) as follows:

$$\begin{aligned} Var(w) = Var(\theta) + Var(\psi) + Var(X\beta) + Var(\epsilon) + \\ 2Cov(\theta, \psi) + 2Cov(\theta, X\beta) + 2Cov(\psi, X\beta) \end{aligned} \tag{4}$$

Besides the variation of individual, firm and time-varying characteristics and the unexplained, residual variation, this form highlights the role of the double covariance terms. Among these, the most notable one is the covariance between individual and firm effects, a common measure of wage sorting in the labor market, signalling how commonly do better (higher wage) workers match with better (higher wage) firms. If we add additional components , the formulae expand substantially as more terms appear. For example, with the addition of occupation fixed effects, λ, we get:

$$\begin{aligned} Var(w) = Var(\theta) + Var(\psi) + Var(X\beta) + Var(\lambda) + Var(\epsilon) + \\ 2Cov(\theta, \psi) + 2Cov(\lambda, \psi) + 2Cov(\theta, \lambda) + \\ 2Cov(\theta, X\beta) + 2Cov(\lambda, X\beta) + 2Cov(\psi, X\beta) \end{aligned} \tag{5}$$

This formula now assesses not only the role of diversity in the average wages different occupations pay, but the possible sorting pattern between high productivity firms and specific occupations (that is the occupational compositions of different types of firms), and also the non-random selection of individuals into occupations. For instance, if the highest paying occupations are getting more and more dominated by those who would be high-achievers in other jobs as well, inequality will increase. Also, in the standard AKM model with two fixed effects, we may overstate the role of firm effects if 'high paying jobs tend to go hand in hand with high- paying firms' as Torres et al. (2018) finds.

To assess the evolution of wage inequality in the US, Song et al. (2019) builds upon the decomposition of Card et al. (2013), but further decomposes the variance terms into between and within firm elements. [13] Suppressing the role of time-varying components, $X\beta$, the core of their decomposition is as follows:

$$Var(w) = \underbrace{Var(\theta - \bar{\theta}) + Var(\varepsilon)}_{Within-firm} + \underbrace{Var(\psi) + 2Cov(\bar{\theta}, \psi) + Var(\bar{\theta})}_{Between-firm} \tag{6}$$

[13] A concept also presented in Abowd and Kramarz (2004).

Within-firm inequality can only originate in the difference of workers effects within the firm and the residual terms. Between-firm differences, however, incorporate three factors: firms can be different in their average wage levels as captured by the firm effects, different quality workers may be employed by different firms – wage sorting – , and finally firms can differ in the *average* quality of workers they employ. The authors label the latter term *segregation*, capturing the fact that differently qualified workers may tend to work at different employers, even among firms with similar wage premia. If we include occupations and the $X\beta$ terms, the above formula becomes[14]

$$Var(w) = \underbrace{Var(\theta-\bar{\theta}) + Var(\lambda-\bar{\lambda}) + Var(\varepsilon) + 2Cov(\theta-\bar{\theta}, \lambda-\bar{\lambda}) + W}_{Within-firm} + \underbrace{Var(\psi) + 2Cov(\bar{\theta},\psi) + 2Cov(\bar{\theta},\bar{\lambda}) + 2Cov(\bar{\lambda},\psi) + Var(\bar{\theta}) + Var(\bar{\lambda}) + B}_{Between-firm} \tag{7}$$

In this somewhat complicated setup, we can observe whether the different valuation of some occupation is generated between or within firms. Similarly to workers, occupations may be specific to high wage or low wage sectors of the labor market, creating another form of segregation. Thus, we can capture, that the sorting of individuals into differently valued occupations can happen in two ways. First, as firms that tend to use the high wage occupations also employ highly qualified workers – even compared to their occupational average –. Secondly, because even within firms, the better workers achieve higher paying occupations, such as managerial positions.

As an alternative to Equation 4, Card et al. (2018), and previously Gruetter and Lalive (2009), introduce a formula decomposing the variance of wages into only covariance terms with the additive wage components, as follows:

$$Var(w) = Cov(\theta, w) + Cov(\psi, w) + Cov(X\beta, w) + Cov(\lambda, w) + Cov(\epsilon, w) \tag{8}$$

This way, we can predict how much less wages would differ if, for instance, all firms or all workers would be extremely similar. In this setup – labeled as *ensemble* decomposition by the authors – the pair-wise covariance terms from Equation 4 are equally accounted to both of their corresponding wage components. For instance, the contribution of wage sorting – a double covariance term – will now appear partly in the contribution of firm effects and partly in the contribution of worker effects, thus we could not observe its importance directly from this decomposition. Torres et al. (2018) augments the above *ensemble* decomposition by differentiating between observable and unobservable components within the time-invariant firm and person characteristics, as shown in Equations 2 and 3. Accordingly the variance decomposition will also include multiple individual and employer related terms, specifically:

[14]Slightly important terms are suppressed. $W = Var(X\beta - \bar{X}\beta) + 2Cov((\theta-\bar{\theta}) + (\lambda - \bar{\lambda}), X\beta - \bar{X}\beta) + 2Cov((\theta-\bar{\theta}) + (\lambda-\bar{\lambda}) + (X\beta - \bar{X}\beta), \varepsilon)$ and $B = 2Cov(\bar{\theta} + \bar{\psi} + \bar{\lambda}, \bar{X}\beta)$.

$$Cov(\theta, w) + Cov(\psi, w) = \underbrace{Cov(W\eta, w) + Cov(Z\gamma, w)}_{Observable} + \underbrace{Cov(\varepsilon^I, w) + |Cov(\varepsilon^J, w)}_{Unobservable} \tag{9}$$

We would add, that by differentiating the observable and unobservable terms, the wage sorting component of previous equations could be also further decomposed into (at least) four meaningful components:

$$Cov(\theta, \psi) = Cov(W\eta, Z\gamma) + Cov(\varepsilon^I, Z\gamma) + Cov(W\eta, \varepsilon^J) + Cov(\varepsilon^I, \varepsilon^J) \tag{10}$$

Accordingly, we can differentiate between the part of wage sorting that could be fully or partially attributed to observable characteristics, such as the sector or ownership of the firms, or education of workers. And also a part which only reflects assortativity between unobservable firm wage components (premium) and individual heterogeneity (productivity and skills). This provides us with a more detailed analytical tool to assess the source of overall wage sorting, and the potential prevalence of assortative matching in worker and firm (unobserved) productivity.

2.3 Indirect and direct measures of assortative matching

To assess the role of assortative matching and wage sorting, we will first estimate the standard sample correlation coefficient between estimated firm and person effects. We have to note, that even if the model is correctly specified, AKM estimations suffer from a now well-explored incidental parameter problem, labeled as limited mobility bias by Andrews et al. (2008). As firm effects are identified only from movers switching between firms, if the mobility in the labor market is low – for instance, because of short observation periods, using subsamples, or simply having few movers in given sectors – then AKM effects will be estimated with high variance, and sample variances and covariances of the estimated effects will be biased measures of the actual moments of their respective distributions. Specifically one would overstate the variance of firm effects, and thus their importance in wage variation, and also understate the covariance between firm and person effects, due to the negative correlation between sampling errors of parameters of the same observation.

While bias-correction methods have been developed by Andrews et al. (2012) and recently by Bonhomme et al. (2019) and Kline et al. (2020), these methods are often computationally exhaustive or only work subsets of data, most authors, including Card et al. (2013) and Song et al. (2019), just acknowledge the possible presence of this bias and note that it could be safely assumed that observed trends are not affected. As our panel is only a 50% sample of the population, the limited mobility bias problem probably should not be neglected, but our computational setup do not (yet) allow us to implement the methods of Bonhomme et al. (2019) and Kline et al. (2020). However, similar to Torres et al. (2018) and relying on some reassuring examples in the literature[15] , we argue that having fifteen years of data in the same panel may help overcoming this issue. Additionally, we can also observe within-year movements as well, which may also increase the number of job switches per firm used for identification. Nevertheless, we include some additional robustness estimates in which we artificially decrease the observed mobility in the data, and find only a small, although non-negligible change in the main parameters of interest.

Besides the usually reported correlation of the firm and worker effects, we will report $Cov(\varepsilon^I, \varepsilon^J)$ as well. This term captures correlation between the unobservable (residual) firm and person specific components, and therefore may reflect complementarity in productivity better. For instance, the standard measure would incorporate segregation effects as well, if women (lower person effect) would more often match with low wage sectors (low firm effect) for reasons other than productivity, such as different taste for amenities at these firms or discrimination on the employers' side, forcing women out of better workplaces.

Additionaly, we will also rely on firm value added per person as a direct measure of productivity, and following Torres et al. (2018) we will report a correlation between person effects and observed firm productivity as well. This measure is proposed by the authors as a response to critiques of interpreting wages sorting as assortative matching (in productivity) arguing that AKM firm-effects may not be monotonous in firm productivity as not only productivity and rent-sharing may shape average wage-levels of firms. Relying on a direct measure for firm productivity overcomes this issue.

Finally, we would also utilize test scores of individuals from an assessment of mathematics and literacy skills taken at around the age of 16. First, by looking at $corr(\hat{\theta}_i, score_i)$, we can check on the individuals' level, whether a high test score predicts high worker effect. If we uphold that the tested skills measure otherwise unobserved worker skill and productivity, we can test whether worker fixed effects are indeed monotonous in worker productivity. Alternately, if we take as given that worker productivity is well reflected in wages, we could answer whether these tests measure things that are related to future labour market outcomes of students. If the latter holds, then $corr(score, \hat{\psi})$ and $corr(score, \hat{VA})$ will serve as direct measures of productivity sorting, complementing the findings relying only on the AKM framework.

2.4 Wage-gap decompositions

Due to the linearity of our wage model, the overall wage difference among groups by any observable control, C, can be decomposed the following way in a similar fashion as in Cardoso et al. (2016). [16]

[15]For instance, Lachowska et al. (2020) show (their Table 4) that in a panel of 13 years, the correction of Kline et al. (2020) alters estimated results in an almost negligible manner.

[16]As ε is by design independent of any characteristic of $C \in X$, $\frac{\partial \varepsilon}{\partial C}$ is zero. The same holds true for elements of the person and firm effects, that is for $C \in Z$ or $C \in W$.

$$\frac{\partial \ln w_{ijt}}{\partial C} = \frac{\partial \theta_i}{\partial C} + \frac{\partial \psi_j}{\partial C} + \frac{\partial \lambda_{k(it)}}{\partial C} + \frac{\partial \boldsymbol{X_{ijt}\beta}}{\partial C} \tag{11}$$

In order to provide a more detailed assessment of differences across observable and unobservable time-invariant characteristics, we can use the second stage decompositions of Equations 2 and 3, by substituting the (linear) detailed decomposition of firm and person effects for their corresponding composite terms.

$$\frac{\partial \ln w_{ijt}}{\partial C} = \frac{\boldsymbol{W_i\eta}}{\partial C} + \frac{\varepsilon_i^I}{\partial C} + \frac{\boldsymbol{Z_j\gamma}}{\partial C} + \frac{\varepsilon_j^J}{\partial C} + \frac{\partial \lambda_{k(it)}}{\partial C} + \frac{\partial \boldsymbol{X_{ijt}\beta}}{\partial C} \tag{12}$$

Alternately, we would note that the differences in person traits could be decomposed into differences generated within and across firms in the spirit of the Song et al. (2019) approach. Similarly, distinguishing whether the workers of a given type of firms are generally prone to work in high wage firms or that they only earn higher wages in given types of firms can help in understanding the segregation mechanisms at hand. Accordingly, the following decomposition also holds, with barred variables denoting the firm-level mean individual effects or the person-level mean firm effects.

$$\frac{\partial \ln w_{ijt}}{\partial C} = \frac{\partial(\theta_i - \bar{\theta}_{j(i)})}{\partial C} + \frac{\partial \bar{\theta}_{j(i)}}{\partial C} + \frac{\partial(\psi_j - \bar{\psi}_{i(j)})}{\partial C} + \frac{\partial \bar{\psi}_{i(j)}}{\partial C} + \frac{\partial \lambda_{k(it)}}{\partial C} + \frac{\partial \boldsymbol{X_{ijt}\beta}}{\partial C} \tag{13}$$

Now, if we instead of a general Z consider a time-invariant, observable personal characteristic, G, the above two approaches from Equations 12 and 13 could be combined in a tractable way, as some components are again zero by definition in such case.[17] A detailed decomposition - after controlling for differences in time-varying and occupation effects - by G would then be the following.

$$\frac{\partial \ln w_{ijt} - \boldsymbol{X_{ijt}\beta}}{\partial G} - \frac{\partial \lambda_{k(it)}}{\partial G} = \underbrace{\frac{\partial(\theta_i - \bar{\theta}_{j(i)})}{\partial G}}_{\text{Within-firm gap}} + \underbrace{\frac{\partial \bar{\theta}_{j(i)}}{\partial G} + \frac{\boldsymbol{Z_j\gamma}}{\partial G} + \frac{\varepsilon_j^J}{\partial G}}_{\text{Between-firm gap}} \tag{14}$$

For instance, if G stands for a dummy on gender, this decomposition would tell us the following. How different premium firms do male and female workers sort into comes from a part that is explainable by observable firm differences, such as sectors and ownership, and a component coming from unexplained firm premia.[18] Besides, the average person effect difference between males and females can on one hand generated within firms, due productivity differences or discrimination for instance. However, another, between-firm component is present as well if, for instance, males workers tend to work at firms that usually employ highly productive, high wage individuals. This element is naturally

[17] Specifically, ε_i^I is independent of G, and there are no within person deviations in the person effect during one's lifetime.

[18] Following Woodcock (2008) we could label these terms inter-industry sorting and intra-industry sorting respectively, despite being derived slightly differently.

related to the segregation component of Song et al. (2019), and accordingly, if males and females would be equally represented in firms (no segregation), it would be zero.

Considering a time-invariant firm characteristic, F, a similar decomposition is as follows.

$$\frac{\partial \ln w_{ijt} - \boldsymbol{X_{ijt}\beta}}{\partial H} - \frac{\partial \lambda_{k(it)}}{\partial H} = \underbrace{\frac{\boldsymbol{W_i\eta}}{\partial H} + \frac{\varepsilon_i^I}{\partial H} + \frac{\partial \bar{\psi}_{i(j)}}{\partial H}}_{\text{Between-individual gap}} + \underbrace{\frac{\partial(\psi_j - \bar{\psi}_{i(j)})}{\partial H}}_{\text{Within-individual gap}} \tag{15}$$

The interpretation of this equation is similar to the case of individual characteristics. Let us consider, for example, the case of firms with majority foreign ownership. These firms may employ workforce that would earn higher wages anywhere, either because high observable (education) or unobservable skills. Besides it is not irrelevant, whether employment spells at multinationals are especially important in workers lifetime, or these workers generally tend to enter high premia firms, foreign-owned ones not being more special than other high-quality workplaces.[19]

2.5 Alternative specifications of the AKM model

One common alternative to compare the two-way, additive AKM model to is the match model, in which all employer-employee matches can have their own wage component, providing a fully elastic representation of firm premia. Even if most firms and workers don't meet more than once, in such models the age and tenure effects are calculated within employment spells of the same employer-employee matches. The estimated match effects can be then, in a second stage decomposed into firm and person effects, with the residuals of that equation, $\tilde{\omega}_{ij}$, representing the (orthogonal) match components (Woodcock, 2015). [20]

$$ln\ w_{ijt} = \boldsymbol{X_{ijt}\beta} + \omega_{ij} + \lambda_{k(ijt)} + \varepsilon_{ijt} \tag{16}$$

$$\omega_{ij} = \tilde{\theta}_i + \tilde{\psi}_j + \tilde{\omega}_{ij} \tag{17}$$

Due to the flexible assumptions of the models on firm-worker relation, these models are expected to provide an overall better fit, and more precise assessment of firm and person effects. Most authors, however, argue that the improvements by applying such models, measured by the decrease in model RMSE for instance, are marginal, and hence the linear, additive assumptions of the AKM model are not essentially mistaken. The importance of the orthogonal match terms can

[19]The decompositions in Equations 14 and 15 are also special cases of the decomposition what Boza and Ilyés (2020) proposes and applies for assessing the effect of the presence of former coworkers on entry wages.

[20]By including occupation effects as well, we actually have four fixed effects, that can be estimated in two consecutive steps, similarly as in the decomposition exercises of Cardoso et al. (2018) and Addison et al. (2018).

be also measured by $\frac{cov(\tilde{\omega}_{ij},w)}{var(w)}$, and we will use this formulation to reflect on model robustness later in the paper.

As a middle ground between standard AKM and match models, one can also allow for the firm effects to only vary over specific observable characteristics. Examples for interacting firm effects with person characteristics appear in Card et al. (2016) and Card et al. (2018), who use these specifications to test for differential rent-sharing, their main assumptions being that firms may not pay the same premia for their male and female (or educated versus non-educated) workers. If firms share their rents differently with such groups, for instance due to differing bargaining power of individuals, we expect to find differences in the average firm-group level fixed effects / wage components across the grouping characteristic. One way to formulate such model in a simple equation is:

$$ln\ w_{ijtg} = \boldsymbol{X_{ijtg}\beta} + \theta_i + \Psi_{jg} + \lambda_{k(ij)} + \varepsilon_{ijtg} \tag{18}$$

The above formulation is somewhat different from that of Card et al. (2016), Card et al. (2018) and Gerard et al. (2021), who in practice fit separate AKM models on male and female, educated and non-educated, or white and non-white subsamples, allowing for different returns for all included observable controls for the given subgroups. When testing the robustness of the AKM model, we will pertain the setup of Equation 18, assuming the same occupation, age and tenure effects for all sub-groups. Beside groupings based on gender or education, we propose three new specifications, in which the firms pay different premia for their workers of different occupations (job model), completed tenure or age. Let us refer to the family of all such models throughout the article G-AKM – after*grouped*-AKM. [21]

Similarly as in the match model, the estimated firm-group effects can be, in a second stage decomposed into the composite of the predicted effect of the grouping variable, and the (baseline) firm effects:

$$\Psi_{jg} = G\tilde{\beta}_g + \tilde{\psi}_j + \varepsilon^G_{jg} \tag{19}$$

The residual of this step conveys how much explanatory power we gain by allowing the firm effects to vary across group members. For instance, if the gender wage-gap would be the same across all-firms then a β_g parameter and the firm effects capturing the mean firm premia would already perfectly explain the firm-gender effects. The large role of ε^G_{jg} would, however, suggests that the gap may widely differ across the range of firms. Checking the differences in the average firm-group effects and the firm-effects in the second step also provides an alternative to the approach of Card et al. (2016) for assessing bargaining differences and sorting with respect to observed characteristics.

$$\frac{\partial \Psi_{jg}}{\partial G} = \tilde{\beta}_g + \frac{\partial \tilde{\psi}_j}{\partial G} \tag{20}$$

[21] The G-AKM firm-group effects could be also used to assess differential rent-sharing behaviour of firms, using the firm-group effects as measures of wage net of skill composition effects (Card et al., 2018; Card et al., 2016). For preliminary results on differential rent-sharing on Hungarian data, see Boza (n.d.).

The LHS term in Equation 20 is the overall difference in firm effects based on an observable characteristic, say gender, while the right hand side is the composite of a term capturing the within-firm wage differences, and a term capturing the different sorting of groups in G. As Appendix B demonstrates, this method will provide an estimate which is the weighted average of the two decomposition proposed in Card et al. (2016), with the additional advantage of being easily generalisable for G-s of more than two groups.[22]

Finally we note, that in a similar fashion, one may also allow firm effects to vary over time. This allows for firms paying a different premia in different periods or even year-by-year.[23] This model has been previously proposed and used by Macis and Schivardi (2016), Lamadon et al. (2019) and by Lachowska et al. (2020), with the latter labeling the model TV-AKM. In Section 5.2 we will estimate this alternative specification as well, alongside the above outlined models with firm-group interactions.

3 Source of data

In the empirical part of this paper we estimate the AKM model from Equations 1-3, and report the expanded decompositions from Equations 5, 7 and 8 through 10 to characterize wage dispersion in Hungary. Utilizing correlations between individual effects and firm effects or the value added of employers we check for wage and productivity sorting as well. By regressing firm effects on firm's value added in a simple OLS we measure cross-sectional rent-sharing elasticites as well. We will also rely on data on test scores to deliver a direct measure for assortative matching. To asses sorting mechanisms attributable to observable characteristics we first decompose some common wage-gaps across observable person or firm characteristics. Then, after testing the fit of grouped AKM models, we use such specifications to decompose differences in firm-group effects into bargaining and sorting components.

The estimations use data from the Databank of the Research Centre for Economic and Regional Studies[24]. The Panel of Administrative Data from CERS is a large, administrative, linked employer-employee panel dataset, covering a random fifty percent of the Hungarian population. The two-way panel spans from 2003 through 2017 and contains labor market data in monthly resolution, such as an ID for the employer, earnings in given month, occupation informa-

[22] As $\tilde{\psi}_j$ captures the average premium of the firm after controlling for its composition with respect to G, $\tilde{\psi}_j$ should be roughly equal to ψ_j of the original AKM specification, and therefore of the decomposition in Equation 11. The difference of the wage gap estimators, $\frac{\partial \psi}{\partial G}$ and $\frac{\partial \tilde{\psi}}{\partial G}$ signals that the assumption of a constant gap is too restrictive in decompositions using the original AKM model, such as Cardoso et al. (2016). In our estimation, while the correlation between these terms and the standard AKM firm-effects is around 0.99, we will find meaningful differences in the partial derivatives.

[23] Which assumption – not accounting for computational constraints – would in the worst case result in loss of efficiency, if the firm effects are in fact, stable over time.

[24] Formerly of the Hungarian Academy of Sciences, now of Eötvös Loránd Research Network.

tion and balance sheet data for incorporated employers.[25] We observe all taxed earnings from the given employer during the given month, but cannot differentiate between bonuses, and general wage.[26] The data does not convey any family-related information, only individual characteristics like gender, age, residence and also some variables on healthcare expenditures and specific transfers received by the individuals, which we do not utilize in this research.

The data also has some unique features regarding education. Although we do not have a common "highest education" variable available for the full panel, in the second half of the observation period we have information on the high school and university attendance of the individuals. Also we observe test scores earned on a standardized country-wide test of mathematics and literacy skills for some young cohorts in the data. The National Assessment of Basic Competences (NBAC) is conducted in each year with the participation of all students in Hungary in 6^{th}, 8^{th} and 10^{th} grades, that is around the ages of 12, 14 and 16 respectively. As we observe these scores and school identifiers only for those who have still attended one of these tests in and after 2011, the utility of this information is somewhat limited by the end of our panel. Specifically part of these cohorts – those who aim for a university degree – may be just entering the labor market after 2014 or even later, leaving only a few years of observations about labor market participation. Nevertheless, we try to make use of both the NBAC scores and high school identifiers in trying to assess the extent of assortative matching with regard to labor market entrants (without higher education). Choices about the included variables, approximations and sample restrictions are detailed in Appendix C.

Our results comprise of two larger sections. The first, Section 4, contains the results of the main decomposition techniques presented in Section 2.2 for the largest sample we had and the discussion of the role of wage components and sources of wage sorting, along the direct assessment of worker and firm productivity and tests for the validity and robustness of the model. The second set of results, in Section 5, focuses on the role of observable characteristics in wage-sorting. The section first presents decompositions of the most relevant wage-gaps in the Hungarian labor market using the three-way AKM model, then introduces the grouped-AKM approach to decompose differences in firm effects into bargaining and sorting components, building on Card et al. (2016).

[25] Unlike LEED data from many other countries we lack establishment identifiers, so we can treat only whole firms and institutions as the unit of observations.

[26] The social contribution reports which form the basis of the data have to be submitted on a monthly basis since 2012. Before that, yearly earnings from an employer were attributed to calendar months accordingly to the number of days of the employment spell belonging to the given month.

4 Results I. – Variance decompositions and sorting

4.1 Main decomposition and evidence for sorting

We start by presenting results from estimating the AKM model with additive firm, person and occupation effects, on the full sample of fifteen years of (quarterly) data pooled together, alongside the second stage regressions of estimated fixed effects on observable time-invariant components. Table 1 contains three panels corresponding to the detailed variance decomposition of Equations 8 and 9, following Torres et al. (2018), the main moments characterizing wage sorting, and also the main between-firm elements of the decomposition of Equation 6, based on Song et al. (2019).

The main decomposition provides importance shares of the wage components of similar magnitudes as previous studies shown in Appendix Table A1. Even after controlling for firm and occupational heterogeneity, individual time-invariant differences contribute to half of the overall wage variation. Of that, around one-third could be attributable to gender and skill differences – proxied by educational requirement of highest occupation –, and most part of the individual heterogeneity remains unexplained. This unexplained part, comprising, for instance, unobserved skills in itself give almost 30% of the overall wage dispersion. Occupations capture around 8% of overall variation. This component is also similar to the finding of Torres et al. (2018), who find a 15% share for the total explanatory power of additive occupation and the collective agreement of the firm.[27] The firm component accounts for a bit more than fifth of overall dispersion, with two-thirds of it accounting for factors other than sectoral differences or the type of majority ownership, while the between sector (owner type) differences in firm premia accounts for 2.4% (4.1%) of the overall dispersion. The observable elements are not negligible either. If foreign-owned, domestic private and state-owned firms and institutions would not differ systematically in their wage policies, overall wage variance would be almost 4% lower in Hungary. Finally we would note that the share of residual variation, not explained by observed factors or fixed effects, is slightly higher than in previous studies, being 14.6%.[28]

[27] The authors do not report the shares attributable to the two factors separately.

[28] As we present later, the model fit is somewhat stronger in the earlier periods of the data, and weaker for the final years of the data.

Table 1: Decomposition of wage variance, Full sample

Variance of log wages	0.338	
Ensemble decomp. (and sub-shares)		
Contribution of XB	**5.40%**	
— Year	*1.98%*	36.8%
— age*, firm size, contract, tenure*	*3.41%*	63.2%
Contribution of individual heterogeneity	**49.85%**	
— Unobserved individual heterogeneity	*29.00%*	58.2%
— Observed individual (gender, quasi ed.)	*17.62%*	35.3%
— Birth year	*0.32%*	0.6%
— Region	*2.91%*	5.8%
Contribution of firm heterogeneity	**22.21%**	
— Unobserved firm heterogeneity	*15.69%*	70.6%
— Observed firm heterogeneity (ownership)	*4.14%*	18.6%
— Sector	*2.38%*	10.7%
Contribution of occupations	**7.93%**	
Residual variation	**14.61%**	
Correlations (and contr. to overall)		
Corr(θ_i,ψ_j)	0.175	9.3%
Corr(ε_i^I,ε_j^J)	0.138	4.4%
Corr(θ_i,ψ_j) for inc. firms	0.310	15.5%
Corr(θ_i,VA_j) for inc. firms	0.364	
Corr(ψ_j,VA_j) for inc. firms	0.615	
Between-within decomposition		
Between-firm share	47.5%	
— Ind. segregation	11.3%	
— Var(ψ_j)	18.3%	
— Sorting	9.3%	
Number of Observations (1000)	66155	
Number of Firms (1000)	144	
Number of Workers (1000)	2462	

Notes: The table conveys moments relating to the components of the estimated model of Equations 1-3. The first panel comprises the ensemble decomposition based on Equation 8. Second panel contains sample correlations of estimated firm and person effects (both overall and unobserved parts) and firm value added. The third panel represents the between elements of the decomposition of Equations 6 and 7. The exact sample and variables used are defined in Appendix 6.

Considering, whether the overall wage dispersion is generated between or within firms we apply the (modified) decomposition of Song et al. (2019) from Equations 6 and 7 and present the main components in Table 1, with the full decomposition of Equation 7 presented in Appendix table A3. The figures in Table 1 highlight that around half of wage differences originates in differences between firms. This share is higher than the 22% percent share from the first panel, as it encompasses not only the fact that firms differ in their average premium (18.5%), but also the full effect of high wage workers sorting into high-wage firms (9.1%), and the fact that workers of different skills (different person effects) segregate into different firms (11.3%). The detailed decomposition also reveals that workers with higher individual wage components tend to work in higher wage occupations. This also affects the between firm differences as the occupational composition of firms and the quality of their workforce is related, accounting for another 4.1% – a pattern observed by Torres et al. (2018) as well. Even within firms, better workers get into better occupations. Specifically, two thirds of the dispersion in occupation effects happens within firms, contributing to the overall within variation by 3%.

As this decomposition already highlights, there is a positive correlation between firm and worker effects, accounting for almost one-tenth of overall dispersion in wages. The corresponding correlation is 0.17, that is not as high as in some previous studies, but definitely positive. Using the notion that this covariance term could be further decomposed according to 10, we can check in more detail the source of this sorting pattern. Unlike Torres et al. (2018), whose Table 4 reports a negative correlation between the unobserved sub-components of the fixed effects, we find a smaller, although positive correlation even for this moment as it is presented in the middle panel of Table 1, and in the detailed decomposition of Table 2. The latter table also reveals that a relatively small fraction of the covariance could be attributable to correlations between observable person and observable firm characteristics. Instead, better latent skill workers tend to sort into different sectors and ownership categories, and workers of different regions and education end up in firms with different unobserved wage components, with the co-movement of unobserved components accounting for 47% of the covariance term. That is around half of the estimated wage sorting relates, at least partially, to observable characteristics. In order to understand the channels in which wage gaps along observable characteristics shape the wage distribution, we revisit this question in Section 5.

Table 2: Sources of Covariance Between Firm and Worker Effects

	θ_i	Unobs.	Gender+q.educ	Birth	Region
ψ_j	**0.175**	74.1%	19.8%	-8.6%	14.7%
Unobs	86.0%	47.0%	29.3%	-1.8%	11.5%
Ownership	11.1%	15.9%	-2.2%	-5.1%	2.5%
Sector	3.0%	11.3%	-7.3%	-1.7%	0.7%

Notes: Column variables correspond to the second stage decomposition of worker effects into (proxied) education, gender, birth cohort, and residential components, while row variables further decompose firm effects into ownership and industry components, as proposed in Equations 2 and 3. The first row and column decompose the covariance between worker and firm effects along one dimension, and the bottom-right (main) panel of the table presents the two-dimensional covariance decomposition proposed in Equation 10.

Another way to characterize the sorting patterns is to explore which parts of the joint distribution of worker and firm effects causes the correlation. To check this we assign workers into ten quartiles both alongside their estimated worker and firm effect, then plot the joint distribution of observations in our sample with regard to these two discretized dimensions. Figure 1 suggests that while it is clear that high wage workers end up at high wage firms, it does not hold that the lowest quality workers end up in the worst firms. Instead, inferior workers match with middling firms, and the lowest premium firms employ various types of workers, which is still consistent with a positive correlation.

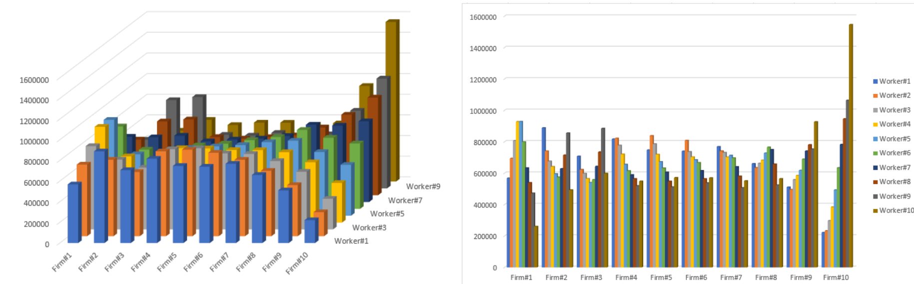


Figure 1: Joint distribution of firm and person effect deciles

Notes: The left panel presents the number of observations by cells defined along 10 deciles of estimated firm effects and 10 deciles of estimated person effects. The right panel presents the same numbers for these cells, first grouped by the firm effect deciles.

Although we lack the computational infrastructure for reproducing the bias correction methods of Kline et al. (2020) or Bonhomme et al. (2020), we would like the assess the severity of limited mobility bias in our sample. While the long panel and the ability to observe within-year mobility works in our favor, our dataset being only a 50% sample of the population decreases the level of observed mobility per firm. In Appendix Table A4 we repeat our main estimations after further decreasing our data, simulating scenarios if the dataset would be only a 20% or 10% random sample from the Hungarian population. Accordingly to the predictions of studies on limited mobility, as we artificially decrease sample size, estimated the correlation between firm and person effects decreases in columns 2 and 3, while sampling after estimating the model (columns 4 and 5) in itself does not alter the estimated moments. However, we do not see any substantial increase in the contribution of firm heterogeneity or in the share explained by sorting, even after dropping 80% of our original sample. This somewhat reassures as that our main estimations can be considered mostly reliable.[29]

[29]The last column of this table comprises results from using wage data from February, May, August and November, instead of January, April, July and October, suggesting a marginal

4.2 Firm productivity, rent-sharing and student skills

Following Torres et al. (2018), we also show that the implicit productivity measure of firm effects and the value added parameters calculated from balance sheet data are substantially correlated, as $corr(\hat{\psi}, VA)$ is around 0.6 for those firms who have such data available.[30] The correlation between this direct measure of firm productivity and AKM worker effects is 0.36. However, we have to note that on the sample of incorporated firm, which have balance sheet data available the correlation between AKM firm and person effects is also above 0.3 – see Table 1. Nevertheless, this result implies that the wage sorting inferred from AKM effects indeed reflects productivity sorting and positive assortative matching as well, reinforcing the notions of Torres et al. (2018).

These strong correlations also suggest that the wage level or premia of firms highly depends on firm productivity. This, of course, could happen due to various reasons. For instance, high wage firms may operate in dangerous sectors or demand more overtime, paying high compensating differentials. Alternately productive firms may rely more on rewarding wage schemes like efficient wages. Also, they may share the rents of being productive with their workers through higher wages. Quite importantly, the correlation of wages and productivity could be the result of more productive workers employing higher quality workforce. Following Card et al. (2018) and Card et al. (2016) we regress the estimated firm effects on the value added per worker of the firms, while controlling for 2-digit sector codes and ownership, and get an elasticity of 0.15, which is cross-sectional estimate of rent-sharing, albeit pure of workforce composition effects. This elasticity is quite similar to the findings of Card et al. (2016) – 0.16 for males, 0.14 for females – and Card et al. (2018) – 0.12 for males, controlling for broad sectors and cities. Using log sales or the lagged value of firm productivity as an instrument for productivity, we get somewhat higher estimates, similarly as the authors cited above.[31]

In the final exercise of this sub-section we focus on observations of those young workers who were in tenth grade in the academic years of 2011/2012, 2012/2013 or 2013/2014, as for these students we have data on their test scores achieved on the National Assessment of Basic Competencies – a compulsory test

role of choice among the two sampling methods.

[30] Although, unlike Torres et al. (2018) we use only the value added per worker values calculated directly from balance sheet data and do not estimate production functions controlling for capital and labor composition as the authors did. For the same, raw measure of productivity the aforementioned authors found a correlation of 0.55 and we can find similar correlation in the work of Card et al. (2016) as well – 0.42 for male and 0.38 for female workers.

[31] Appendix Table A5 presents the result for the cross-sectional rent-sharing estimations, both with using the firm-year level average wage, and the AKM firm effects as outcomes. All models control for sectors defined by the interaction of majority ownership (3 categories) and one digit industry codes (15 categories). Firms belonging to the 'no surplus' range, defined as in Card et al. (2016) are omitted from the estimations. In an accompanying paper on differential rent-sharing, Boza (n.d.), we introduce a more nuanced measure for rent-sharing, using temporal variation in firm wage levels and productivity, while still controlling for composition – by using TV-AKM effects –, and also use G-AKM models to investigate whether productivity rents are shared differently across workers of different gender, age, education, occupations or seniority within the same firms.

in mathematics and literacy skills.[32] Unfortunately, we can follow these cohorts only for 3-5 years after taking the tests, that is typically only 1-3 years on the labor market, at the age between 19 and 22, even if they did not take higher education.[33] This substantially limits the scope of conclusions to be drawn from the following results. Nevertheless, we try to explore the relation between these scores and some specific labour market outcomes of those who enter the labor market after (or instead of) high-school graduation. [34]

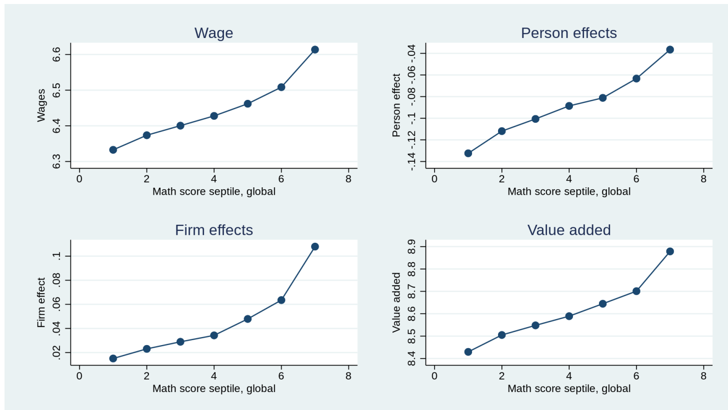
For Figure 2, we collapsed the mathematics test results of student into seven quantiles in two different ways. First, by generating the septiles across all students taking the NBAC in the same year, then only across given schools and testing years. The latter septiles therefore correspond to the within-school relative performance of students. We then plot the mean of future wage observations, estimated AKM individual effects and firm effects, and – if available – the value added measures from firms' balance sheet data in employment spells in employment spells between 2014 and 2017, by these septiles. As we can observe, students with better test results will generally earn more in their early career, have larger individual effects – which can not reflect higher education, due to the limited sample window –, and more importantly, end up with better quality firms. The latter observation holds true for both the indirect productivity measure of AKM firm effects, and the value added of firms as well. This latter pattern serves as another, direct evidence for the presence of assortative matching if we accept that these test scores are indicative of future unobserved worker productivity. Considering within-school relative test results, it seems that better students of the same cohort and school also tend to have higher wages, but a previously observable advantage of the top septile disappears, suggesting a role of between-school score differences in forming the score-wage relation. The same results for literacy test scores can be found in Appendix Figure A1. Patterns are weaker in these plots, and while literacy scores seem correlated with wage outcomes, and firm productivity, they correlate with worker-specific wage components in a less monotonous way.

Finally, in Table 3 we include the correlations between continuous test scores and the introduced wage components, now including the unexplained part of firm effects as well. These correlations are shown for both the sub-sample of all young workers with test scores and those working at firms with available balance sheet data. Also school-year level observations are generated by taking the mean of above variables in such units. All correlations in the table are positive, although most may be considered modest in size, with the main exception being the systematic sorting of students with better test scores into

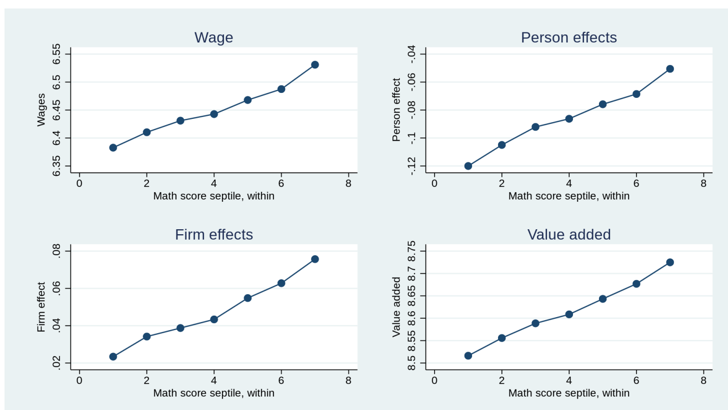
[32]This test does not serve as a basis for any further academic outcome, for instance university admission, therefore the effort put into preparation for the test may depend on student's general attitude besides their skills. For this very reason, those who are absent on the day of test do not have to retake it, so the data may bear a slight selectivity bias as well, besides not being a perfect measure of skills due to the lack of real stakes of the exam.

[33]Therefore, even those who choose a 3-year BA program, and don't work beside their university studies will not be part of the sample used for this exercise.

[34]A more direct assessment of the relation of these scores and future employment and (standard) wages can be found in Hermann et al. (2019).



(a) Overall septiles



(b) Within-school septiles

Figure 2: Wage components and value added along 10^{th} grade mathematics score septiles from NABC

Notes: The seven quartiles are created along the distribution of literacy scores in year the students took the test (top panel), or within the distribution of the given school-year (bottom panel). The figures relate to those students for whom we have a test score observation no sooner than 2008 and also at least one wage observation anytime in the panel. The value added measure is available only for incorporated firms and not public institutions.

higher wage occupations. For the samples of incorporated firms, we generally find stronger correlations between test scores and firm quality, even regarding the non-sectoral firm component. Those workers who earn higher points on the NBAC test, tend to end up in firms with higher value added. This measure of assortative matching is of similar magnitude as the correlations between AKM effects in the full sample. Furthermore, we see that the correlation with the within-school relative score is substantially weaker, suggesting that the segregation of capable students at teen age indeed plays an important factor in creating sorting between high wage schools and firms (or occupations) reflected in correlations with the school-year level observations as well.[35]

Table 3: Correlation of NBAC scores and measures of productivity

	w_{ijt}	θ_i	ψ_j	ψ_j^U	VA_j	λ_k
All						
Math	0.222	0.147	0.125	0.132	.	0.279
Lit	0.165	0.056	0.079	0.115	.	0.316
Incorporated						
Math	0.241	0.164	0.156	0.142	0.173	0.307
Lit	0.191	0.080	0.137	0.126	0.133	0.298
Relative to school						
Math	0.116	0.115	0.069	0.050	.	0.079
Lit	0.063	0.025	0.037	0.039	.	0.093
Relative, Inc.						
Math	0.119	0.119	0.061	0.053	0.081	0.123
Lit	0.068	0.030	0.051	0.042	0.047	0.094
School level						
Math	0.280	0.115	0.133	0.426	.	0.246
Lit	0.261	0.084	0.100	0.233	.	0.478
School level, Inc.						
Math	0.344	0.123	0.183	0.274	0.272	0.487
Lit	0.326	0.094	0.147	0.257	0.247	0.542

Notes: sample correlations relate to around 1.02 million (0.78 million) employment observations of individuals and 9700 (9200) corresponding school-years, in the whole (incorporated) sample.

4.3 Validity and robustness of main results

As Card et al. (2013) discusses, one form of misspecification of the AKM model would occur if worker mobility, and thus the design matrices of who works where, would depend on employer-employee match effects, and hence these match ef-

[35]The topic of segregation with respect to future labor market prospects is a topic to be explored in more detail in future studies. Naturally, using a longer period of labor market outcomes would be desirable for this exercise, but the administrative data covering years after 2017 will not be available sooner than 2023.

fects could not be independent components of the error terms. To asses the relevance of match components, instead of the additive, separate worker and firm effects Card et al. (2013) estimate a model with worker-firm fixed effects to show that such a model only has a slightly larger explanatory power.[36] After estimating the same model, we also further decompose matches into separate worker, firm and (orthogonal) match effects to see, whether the importance weights of wage components in overall wage variation and sorting patterns are affected by applying such two-step model, as defined in Equations 16 and 17.

The results, presented in Appendix Table A6, suggest that, while the residual variation in this model has decreased by around five percentage points, the firm-person match effects can contribute to a similar share of overall variance. The share of other components remain quite stable, with only the contribution of occupations showing a stronger decrease. This already suggests that matches only capture residual variation unrelated to the original AKM person and firm effects. In the first panel of Appendix Figure A2 we plot the average of estimated match effects alongside the firm and individual effect deciles, and find patterns similar to that of the distribution of residuals from the original model (second panel). This plot also suggests that the included match effects were able to capture most of the residual variation which previously was coming mostly from the lowest deciles of workers and firms regarding their corresponding AKM effects. For these cells in the joint distribution, the mean residuals could be as large as 0.02-0.03 log points, indicating 2-3% average difference between predicted and actual wages for these worker-firm pairs. So, while it seems that in Hungary the match model would yield slightly superior explanatory power to the additive firm and person effect model, the assumption of exogenous mobility with respect to match effects may still hold.

We also reproduce the event study analysis presented in Card et al. (2013), investigating the wage evolution of job switchers before and after changing their employer, looking for signs of mobility depending on transitory wage-components. If the exogenous mobility assumptions of the AKM model holds, we expect to observe similar wage gains for those who move from one wage quartile to another as the losses expected for those who experience the reverse path of mobility – and no wage gains for those who remain at similar quality firms. On the other hand, no trends should be present in wages either before or after the job-switches. Appendix Figure A3 presents the mobility patterns for four wage quartiles, based on AKM firm effects, in the preceding and subsequent six months of job switches. The presented wage profiles are mostly consistent with these expectations, showing only signs of slight wage gains over time for workers leaving the bottom quartile of firms.

A main contribution of Card et al. (2013) and Song et al. (2019) is presenting the evolution of the wage decompositions over time, characterizing the sources of increase in wage inequality. As we could estimate the AKM model on the whole 15-year period, we first report the decompositions on three overlapping

[36] This (and the consecutive tests) later also appear in Card et al. (2016), Card et al. (2018), Gerard et al. (2021), and also in Macis and Schivardi (2016), Fanfani (2018), Alvarez et al. (2018) and Casarico and Lattanzio (2019).

subsamples from the overall estimation in Table 4. However, we also allow firm, individual, and occupation effects to be different in the three periods by re-estimating the AKM models on these three, shorter time periods. This may be reasonable, as assuming time-invariant firm-effects may be a source of misspecification in long panels if firms can alter their wage schemes either due to the sharing of rents from productivity changes, introducing amenities or disamenities or applying specific contracting strategies. The comparison of the two set of results also provides a way to assess how severe threat the limited mobility bias is when one has to rely on data from shorter panels – a possible drawback of using subsamples.

As the first three columns of Table 4 suggest, in Hungary overall wage dispersion did not change substantially during the 2003-2017 period, as only a slight decrease of variance is present. Accordingly, alongside the overall variation, the contribution of most wage components also remain stable over these three periods. However, the last period is slightly different, as within that period wages increase more rapidly, increasing the contribution of year effects.There is a slight increase in the total of between-firm inequality components, consistent with the comparative study of Tomaskovic-Devey et al. (2020), who show increasing trends in the between share for twelve of the fourteen countries included, Hungary being one of the two exceptions with a steady trend (of one of the highest between-firm share) during the 2003-2011 period. On the other hand, unexplained variation is quite higher in the years following 2010.[37] Compared to the decompositions from the full period AKM estimation, in the models estimated on subsamples (columns 4-6) we achieve around 3 percentage better fit in all three periods. Surprisingly, despite the concern that a shorter panel comes with more serious limited mobility bias, the estimated share of sorting and the corresponding correlations of AKM effects are not lower. On the contrary, we estimate slightly larger sorting parameters. [38]

[37]This may relate to changes in how social contributions had to be reported in the primary of source of data. Since 2011, monthly reports are required from all employers, while previously most employers had the option to report payments to workers only once per year.

[38]As Bonhomme et al. (2020) find evidence for non-negligible limited mobility bias using six years of data per country, the robustness of our results are indeed surprising. This stability may suggest, that the inclusion of within-year mobility or simply the average mobility level of the Hungarian labor market is substantially high to overcome severe limited mobility bias. Nevertheless, only applying a bias correction method would provide a clear verdict on this issue.

Table 4: Decomposition of wage variance, over time

Variance of log wages	0.348	0.331	0.325	0.346	0.331	0.322
Ensemble decomp.						
Contribution of XB	4.47%	3.56%	5.88%	4.69%	3.29%	5.77%
— Year	0.30%	0.25%	3.20%	0.50%	0.30%	3.12%
— age*, firm size, contract, tenure*	4.16%	3.31%	2.68%	4.15%	2.54%	1.95%
Contribution of individual heterogeneity	53.47%	51.67%	46.21%	54.87%	58.18%	52.65%
— Unobserved individual heterogeneity	31.71%	30.11%	26.20%	31.60%	34.45%	32.59%
— Observed individual (gender, quasi ed.)	17.77%	17.95%	17.37%	18.53%	19.88%	17.65%
— Birth year	0.86%	0.51%	-0.01%	1.32%	0.39%	-0.06%
— Region	3.12%	3.10%	2.65%	3.42%	3.45%	2.47%
Contribution of firm heterogeneity	20.96%	22.88%	23.22%	23.11%	21.39%	20.10%
— Unobserved firm heterogeneity	15.78%	15.60%	15.53%	16.46%	14.90%	13.97%
— Observed firm heterogeneity (ownership)	3.54%	4.40%	4.67%	4.37%	3.55%	3.65%
— Sector	1.64%	2.88%	3.02%	2.28%	2.93%	2.48%
Contribution of occupations	8.37%	8.04%	7.41%	7.97%	6.16%	6.72%
Residual variation	12.74%	13.85%	17.28%	9.36%	10.98%	14.76%
Correlations						
Corr(θ_i,ψ_j)	0.173	0.186	0.173	0.164	0.152	0.181
Corr(ε_i^I,ε_j^J)	0.16	0.154	0.109	0.098	0.116	0.12
Corr(θ_i,ψ_j) for inc. firms	0.311	0.33	0.305	0.286	0.308	0.308
Corr(θ_i,VA_j) for inc. firms	0.382	0.382	0.351	0.357	0.399	0.386
Corr(ψ_j,VA_j) for inc. firms	0.618	0.626	0.609	0.61	0.597	0.576
Between-within decomposition						
Between-firm share	46.3%	48.1%	48.8%	45.9%	47.6%	48.7%
— Ind. segregation	11.1%	11.5%	11.5%	12.3%	13.3%	13.8%
— $Var(\psi_j)$	17.4%	18.9%	19.4%	17.0%	17.2%	16.4%
— Sorting	9.1%	9.9%	9.5%	9.4%	10.3%	10.3%
Number of Observations (1000)	30885	30973	30869	28373	28531	28484
Number of Firms (1000)	77	86	94	66	74	83
Number of Workers (1000)	1923	1966	1932	1789	1831	1806
From	2003	2007	2011	2003	2007	2011
Until	2009	2013	2017	2009	2013	2017
Estimation sample	Full	Full	Full	Sub	Sub	Sub

Notes: See Table 1. The first three columns report decompositions of the AKM model estimated using all years, on different subsamples. The last three columns report decompositions on models re-estimated on the corresponding subsamples.

We also present our results for various subsamples of interest. Table 5 comprises results for workers employed through typical labor contracts – that is we exclude public servants from the sample –, male workers, and for workers who were below the age of 27 during the whole sample period. Important differences compared to the full sample include a stronger role of firm heterogeneity in all subsamples. Sorting is also somewhat stronger for males and those who work with typical contracts. For the youth sample, however, we find a lower explanatory power of the AKM model, mainly coming from lower contributions of individual differences, especially unobserved individual heterogeneity. As the overall variation is also lower in this sample, this reflects substantially lower variation in person effects. This may signal that career paths do not diverge enough by this age to allow individual differences alter wages substantially. The contribution of sorting, on the other hand, remains strong, despite a low estimated correlation.[39]

In this table, we also divide the sample based on our proxy of education – that is based on the highest educational requirement met by the individual in any of their observed occupations. Naturally, the role of occupations explain less variation within each of the three categories, with the largest differences being among occupation requiring higher education. In this sample overall dispersion and the contribution of unobserved worker quality is also higher, suggesting a more substantial role of soft skills for educated workers.[40] Accordingly, among workers who never work in occupations with strong educational requirements both observed and unobserved aspects of firm heterogeneity contribute to larger shares of the (within-occupation) variation. The unexplained, residual shares on these subsamples are lower, as one of the main drivers of overall variation is educational attainment in itself. Interestingly, within the educational subsamples correlations between worker and firm effects are somewhat larger than in the full sample, reinforcing that education also plays an important role in allocating workers to different sets of firms, as Table 2 suggested beforehand.

[39]The decrease is partially in line with Torres et al. (2018), who find negative sorting for workers below the age of 26.

[40]And therefore partially explaining the lower role of individual heterogeneity in the sample of young workers, of whom only a smaller share works already in occupations with graduate requirements.

Table 5: Decomposition of wage variance, subsamples

Variance of log wages	0.35	0.369	0.194	0.198	0.218	0.348
Ensemble decomp.						
Contribution of XB	5.18%	5.84%	13.75%	5.75%	6.97%	3.83%
— Year	2.07%	1.86%	9.83%	3.52%	3.11%	1.75%
— age*, firm size, contract, tenure*	3.11%	3.98%	3.92%	2.23%	3.86%	2.08%
Contribution of individual heterogeneity	46.60%	47.07%	27.58%	38.86%	37.87%	48.82%
— Unobserved individual heterogeneity	28.02%	28.31%	16.45%	33.34%	33.49%	43.21%
— Observed individual (gender, quasi ed.)	15.87%	15.67%	7.82%	4.35%	1.14%	1.77%
— Birth year	-0.28%	0.27%	1.53%	-0.69%	0.67%	1.41%
— Region	2.98%	2.81%	1.77%	1.85%	2.56%	2.43%
Contribution of firm heterogeneity	27.97%	26.27%	34.52%	34.61%	31.96%	23.91%
— Unobserved firm heterogeneity	18.47%	18.22%	22.10%	21.40%	21.01%	15.07%
— Observed firm heterogeneity (ownership)	5.98%	5.56%	8.24%	6.74%	6.50%	4.86%
— Sector	3.51%	2.49%	4.17%	6.46%	4.45%	3.98%
Contribution of occupations	6.88%	7.84%	3.79%	3.40%	2.63%	5.53%
Residual variation	13.37%	12.97%	20.37%	17.38%	20.57%	17.91%
Correlations						
Corr(θ_i,ψ_j))	0.291	0.244	0.027	0.238	0.193	0.206
Corr(ε_i^I,ε_j^J)	0.157	0.158	-0.085	0.141	0.133	0.148
Corr(θ_i,ψ_j) for inc. firms	0.31	0.322	-0.026	0.219	0.235	0.337
Corr(θ_i,VA_j) for inc. firms	0.364	0.356	0.116	0.276	0.309	0.356
Corr(ψ_j,VA_j) for inc. firms	0.615	0.582	0.572	0.563	0.593	0.643
Between-within decomposition						
Between-firm share	53.8%	49.0%	77.5%	63.0%	66.4%	46.1%
— Ind. segregation	11.2%	10.0%	14.9%	12.7%	14.0%	11.3%
— $Var(\psi_j)$	18.8%	18.8%	31.9%	28.0%	26.6%	20.2%
— Sorting	13.9%	10.6%	12.9%	9.7%	12.7%	11.5%
Number of Observations (1000)	50956	32157	1971	8243	35797	22115
Number of Firms (1000)	138	128	71	64	133	104
Number of Workers (1000)	2332	1225	306	389	1378	695
Sample	Typical.	Male	Young(≤26)	Low ed.	Mid ed.	High ed.

Notes: See Table 1. Subsamples are formed based on contract type, gender, age, and proxied education.

To check for patterns in sorting within different occupations, regions or employers, we also estimated the correlation of individual and firm effects and the contribution of sorting to wage variation in the given segment for 1-digit, broad categories of occupations, regions, and industries, as presented in Appendix Table A7.[41] Similarly to Torres et al. (2018) and Dauth et al. (2019) we find larger assortative matching for the capital, Budapest and Central Hungary, the

[41] For these estimations we use data only after 2011, as the occupation categorization system in effect changed that year, hence the categorization presented in Appendix C could not be assigned unambiguously to these broad categories.

NUTS-2 region it is embedded in. In line with Table 5, sorting is stronger in superior occupations, but the relation is non-monotonous. Sorting is absent for agricultural jobs and firms, and is the strongest in supporting sectors, logistics, transportation and energetic sectors. When collapsed to the given categories by year, cross-industry and cross-region correlations of mean firm and worker effects increase substantially, suggesting a strong role of the average difference across these units in sorting patterns, again in line with the findings from Table 2. Also these patterns are consistent with the notion that a large portion of the universities and jobs demanding professional qualifications in Hungary are centered in Budapest.

Finally, to relate to the "job-title" model of Torres et al. (2018), we tested whether a model with occupation-sector combined fixed effects would fit the model better than our main specification with only occupation fixed effects. These interacted effects would allow for different average wage levels for workers in a given occupation across different sectors, for instance allowing secretaries at IT firms and at industrial firms to receive different premia, a plausible assumption. The decomposition – including a second-stage decomposition of the combined effects – based on such model is presented in Appendix Table A8. As we can infer from this table, 95% of occupational and sectoral differences can be attributed to separate, additive wage effects of occupations and sectors, with only 5% of the variation coming from the joint effects, indicating a rather low overall role of the above inter-sectoral differences in the valuation of occupations.

5 Results II. – Wage gaps and sorting along observables

5.1 Wage gap decompositions

While the previous section of the result focused on presenting evidence for the overall presence of wage (and productivity) sorting in the Hungarian labor market, in this section we aim to uncover some of the channels in which the positive correlation between AKM person and firm effects could be generated. Specifically we focus on sorting along observable characteristics of both workers and employers, such as gender or education of individuals, the ownership of firms. First, we decompose some specific wage gaps into person, firm and occupation components, according to the main AKM model, building on Cardoso et al. (2016). Later on, we will refine and complement this set of result by models using a grouped AKM approach and a decomposition inspired by Card et al. (2016). For the initial approach, we use the specifications defined in Equations 14 and 15. That is, besides characterizing differences in occupation, firm and person effects, we further distinguish between observable and unobservable components, and also within-unit and between-unit (segregational) aspects. Table 6 contains results for the former decomposition, concerning individual characteristics.

Table 6: Person characteristics

VARIABLES	(1) Wage	(2) Ind.	(3) Firm	(4) Occ.
Male	0.160***	0.131***	0.031***	-0.002*
	(0.007)	(0.004)	(0.004)	(0.001)
Non-skilled	-0.138***	-0.109***	-0.008	-0.021***
	(0.012)	(0.006)	(0.005)	(0.002)
Higher education	0.541***	0.389***	0.045***	0.107***
	(0.007)	(0.004)	(0.004)	(0.001)
Lives in Budapest	0.176***	0.111***	0.057***	0.008***
	(0.025)	(0.011)	(0.014)	(0.001)
Observations	66,155,127	66,155,127	66,155,127	66,155,127
R-squared	0.316	0.420	0.201	0.497

VARIABLES	(1) Obs. firm	(2) Unobs. firm	(3) Within firm	(4) Between firm
Male	0.012***	0.019***	0.114***	0.017***
	(0.003)	(0.003)	(0.003)	(0.003)
Non-skilled	0.001	-0.009*	-0.107***	-0.002
	(0.003)	(0.004)	(0.006)	(0.003)
Higher education	0.005*	0.039***	0.353***	0.036***
	(0.002)	(0.003)	(0.005)	(0.003)
Lives in Budapest	0.010	0.047***	0.060***	0.051***
	(0.009)	(0.007)	(0.004)	(0.009)
Observations	66,155,127	66,155,127	66,155,127	66,155,127
R-squared	0.309	0.045	0.532	0.146

Notes: The parameters in the table are results from regression estimates of the effect of gender, education and residence (Budapest) dummies on wage components defined in Equations 11 (first panel) and 14 (bottom panel) as outcomes. The elements of X and W are included as additional controls. Such variables are quadratic age, quadratic tenure, firm size, year, contract type and birth cohort. Two-way clustered standard errors are in parentheses, with *** $p<0.001$, ** $p<0.01$, * $p<0.05$.

While naturally the individual component dominates the gender wage gap, there is also around 3% wage difference attributable of males sorting into better firms. Of this component, about one third is explained by sector or ownership, while almost 2% is due to intra-industry differences in firm premia. This may reflect exclusionary hiring, but also different preference of male and female workers for firms with different amenities (Sorkin, 2018). Somewhat surprisingly, we don't find an important role of sorting into different occupations. That does not mean, per se, that men and women work hold similar positions, but they are quite equally represented in low and high wage occupations. Differences of the individual wage component can also be generated both within-firms, or across firms. By including firm fixed effects in the regressions on estimated AKM worker effects, we can learn that the within firm wage gap, which is often the focus of studies on the gender wage gap studies, is actually 1.7% lower than what the decomposition of the first panel would suggest, due to segregation effects. For instance, either higher skill males or lower skill women may tend to work in more sex-segregated workplaces, resulting in a lower gap identified only from sex-integrated ones. Hence, alongside sorting into different premia firm, the non-random allocation of males and females of different (unobserved) skills into different firms, with respect to the quality of workforce they employ. will contribute – with a rather important magnitude – to the between-firm wage advantage of men.

Regarding education, we do not find really surprising patterns. The occupations filled by higher (proxied) educated groups are (even by definition) more prestigious occupations, paying higher wages. But as it turns out more educated workers can also get into better firms, even within the same sectors, as it is suggested by the dominance of the unobserved firm component.[42] Finally, the within-firm differences dominate the gap in individual effects, with only high skilled workers showing signs of segregation compared to baseline category. Still, this segregation component is almost as important for between firm differences, as the sorting of workers with high education into higher wage firms.

Finally, we'd like to understand why do people who live in the capital earn almost 20% more than workers living outside Budapest.[43] While certain jobs may be over or underrepresented in the capital, Budapestian workers only work in slightly better occupations on average. However, the firms Budapestians work for are considerably better, and mostly not because different ownership or sector, but due to unobserved, within-sector differences. We note that as we do not have establishment level data these firms may include cross-country chains as well. A surprising find is that Budapest residents earn higher wages for other reasons as well, both within, both between firms. The within channel may happen due to better skills, or national employers simply having to offer higher wages for their establishments in Budapest. Naturally a quite strong segregation component is present due to the spatial nature of our question, and the somewhat superior skills of those who (can) live in Budapest.

[42] The findings are similar to the figures presented by Cardoso et al. (2018) for Portugal, with the highest education category differing the most from the first two. Although we lack precise educational data, except for the youngest cohorts, investigating wage differences in more detail could be possible in the future , relying on information on exact education, on peers at school, and also utilizing the test scores presented earlier in this paper.

[43] The dummy for Budapest relates only to workers who live within the administrative borders of the city of Budapest for most of our 15-year observation window. As commuters living in the agglomeration are not included in this definition, the gaps presented here are probably somewhat underestimate the gap we would get by focusing on those who actually work in Budapest.

Table 7: Ownership gaps decomposed

VARIABLES	(1) Wage	(2) Ind.	(3) Firm	(4) Occ.
Foreign-owned firm	0.385***	0.120***	0.264***	0.002
	(0.023)	(0.011)	(0.012)	(0.003)
State-owned firm	0.105*	0.042**	0.061*	0.003
	(0.042)	(0.015)	(0.024)	(0.007)
Public institution	0.014	0.051***	-0.064***	0.027***
	(0.022)	(0.010)	(0.010)	(0.003)
Observations	66,155,127	66,155,127	66,155,127	66,155,127
R-squared	0.121	0.105	0.366	0.177

VARIABLES	(1) Obs. ind	(2) Unobs. ind	(3) Within ind.	(4) Between ind.
Foreign-owned firm	0.037***	0.083***	0.166***	0.097***
	(0.006)	(0.008)	(0.006)	(0.010)
State-owned firm	0.011	0.031**	0.038***	0.023
	(0.010)	(0.010)	(0.008)	(0.021)
Public institution	0.039***	0.012*	-0.029***	-0.035***
	(0.006)	(0.006)	(0.005)	(0.008)
Observations	66,155,127	66,155,127	66,155,127	66,155,127
R-squared	0.070	0.072	0.835	0.136

Notes: The parameters in the table are results from regression estimates of the effect of majority ownership dummies on wage components defined in Equations 11 (first panel) and 15 (bottom panel) as outcomes. The benchmark category consists of domestic, private-owned firms. The elements of X and are included as additional controls. Such variables are quadratic age, quadratic tenure, firm size, year and contract type. Two-way clustered standard errors are in parentheses, with *** p<0.001, ** p<0.01, * p<0.05.

Considering the ownership of firms, in Table 7 we can see that firms with either domestic, foreign, private or state majority ownership roughly employ the same mixture of occupations, at least considering wage levels. However, compare to these market firms, which accounts for the main body of the economy, public institutions generally make use of higher paying occupations. This is probably due to many of the skill dependent occupations and the lower share of elementary/manual work present in schools or hospitals for instance.[44] Consis-

[44] For estimating this table we did not include industry dummies among the controls, as for public institutions that variable is mostly non-available, and also as many industries are mostly exclusive for the either the private or the public sector, e.g. healthcare, education or agriculture and industry. Appendix Table A9 contains the replication of Table 7 on a

tent with this notion, individuals with better observed skills (higher education) sort into these institutions. While foreign owned firms can also employ a more educated workforce, they can also poach the best workers regarding unobserved characteristics as well, substantially increasing the difference in AKM worker effects. This channel in itself can explain around a quarter of the foreign-domestic wage gap.[45] As these firms pay also higher premia even without these selection/sorting channels, they clearly contribute to overall wage sorting substantially. The strong segregation component, which accounts for 40% of the total difference in the firm component can be interpreted in this setting as a result of the differing work histories of workers who never work for multinationals versus those who tend do, with the latter group mostly getting into higher wage firms. Even within the lifetime of individuals who work in both the foreign and domestic sectors, working in the former usually includes firms with 17% higher premia. The distinction between state owned and private owned firms is not that harsh, although somewhat better wages can be also earned at these firms, by somewhat better workers in the former category.

As we have seen, the sorting of males, highly educated workers and even residents of Budapest , into high wage firms, as these workers would earn higher wages anywhere, clearly contribute to the overall wage sorting observed in Section 4. Similarly, while foreign-owned firms tend to pay higher wages, they are also able to poach the best workers in the labor market, strengthening the correlation between worker and firm productivity and corresponding wage levels. These findings are consistent with Table 2, which suggested that half of the observed wage sorting could be attributable to observable characteristics. In the remaining section of the study, we aim to further assess the level of observable sorting mechanisms, with more flexible model specifications.

5.2 Firm-group (G-AKM) specifications

In this section, we present some alternative - partly novel - specifications of the AKM model. We will use these models to assess sorting, but beforehand, we also discuss whether they provide reasonable alternatives to the standard AKM specification. In each experiment, we relax the assumption that firms have one (relative) wage premia that they pay for all they workers in all time periods. Building upon Equations 18 and 19, we first generate firm-group identifiers, then estimate our three-way AKM model with these identifiers as substitutes for the original firms. We will apply this method for six variables: calendar year, gender, education, occupation, tenure categories and age categories. Due to the different nature of these variables, the connected sets on which the AKM models

sample excluding public institutions, but including industry controls. While within-sectors the advantage of state-owned firms turns into a slight disadvantage, and foreign owned firms seem to employ more workers in high-wage occupations, the main conclusions drawn about the sign of sorting are not affected.

[45] The importance of multi-national employers in Hungary is discussed in detail in Earle et al. (2018) and Köllő et al. (2021).

can be estimated and interpreted can differ substantially.[46] The results of all six models are presented in Table 8. Besides the residual variation of models, an important new row in this table compared to Table 1 is the '*Interacting Variable*' one. This row refers to the wage variation explained by the variance of ε^G_{jg} from 19, signalling how important is the assumption that firm effects are flexible along the given observable characteristic.[47]

Time-invariance assumptions

As we noted when discussing the evolution of wage inequality, the assumption that firms have the same firm-effect over a longer period may be too restrictive, as wage policies of any firm or even whole sectors could alter during a decade, for instance. In the most extreme case, we can assume that firms may change their wage schemes in any year. The AKM model could be altered to allow for such flexibility, by including firm-year interaction effects instead of the time-invariant firm-effects. With such a model we may lose efficiency as we have a magnitude larger set of extra parameters to identify, with the year-to-year changes in firm-effects being identified mostly from the wage variation of workers staying in the given firms, and partially from the inter-firm mobility in the given year.[48] While Macis and Schivardi (2016) already use firm-year effects, Lachowska et al. (2020) proposes the detailed investigation of this alternative model, naming it time-varying or TV-AKM, which we adapt as well. In their study, the authors conclude that this misspecification issue is a 'second order concern' in the AKM framework, as allowing for time-varying firm effects does not significantly alter results on wage decompositions, neither on inference relying on AKM firm effects.[49]

[46] Specifically, we define the connected sets based on the firm-group identifiers, person IDs and occupations according to the algorithm of Weeks and Williams (1964), except for the models with gender and education as interacting variables. We will discuss the identification challenge in these models at a later point.

[47] For instance, if the gender gap would be the same in all firms a gender dummy and firm effects could perfectly predict firm-gender effects, resulting in near zero residuals in the second stage, and hence only a negligible portion of variance explained by this term.

[48] The connected set remains roughly the same as in the main model, as only those firm-year units become disconnected, where the whole workforce is replaced from the end of one year to the start of the next.

[49] As Lachowska et al. (2020) note, one possible reason for the emergence of time-varying firm effects would be the rent-sharing of firms going through productivity changes as presented in Lamadon et al. (2019). Previously, Macis and Schivardi (2016) also interprets the firm-year AKM effects as measures of the firms' rent sharing behaviour. Accordingly these effects may be used for a within-firm approach of estimating rent-sharing elasticities – a question to be considered in Boza (n.d.).

Table 8: Decomposition of wage variance, interacted effects

Variance of log wages	0.334	0.339	0.339	0.332	0.33	0.337
Ensemble decomp. (and sub-shares)						
Contribution of XB	7.2%	10.9%	10.8%	10.6%	10.3%	9.2%
Contribution of individual heterogeneity	48.8%	48.5%	47.9%	45.0%	52.7%	50.8%
— Unobserved individual heterogeneity	28.1%	28.8%	28.0%	27.3%	31.2%	29.2%
— Observed individual (educ,gender,birth,res.)	20.7%	19.8%	19.9%	17.7%	21.5%	21.6%
Contribution of firm-group heterogeneity	26.1%	23.7%	24.9%	35.6%	21.7%	22.3%
— Interacting variable (see at bottom)	2.0%	0.1%	1.1%	8.6%	0.9%	0.2%
— Unexplained	2.1%	0.5%	1.0%	2.2%	0.4%	0.3%
— Firm constant	22.1%	23.1%	22.8%	24.8%	20.4%	21.8%
— — Unobserved firm heterogeneity	15.6%	16.2%	16.0%	17.4%	14.4%	15.5%
— — Observed firm heterogeneity (sector/own.)	6.4%	6.9%	6.8%	7.4%	6.0%	6.3%
Contribution of occupations	8.6%	7.7%	7.4%		6.7%	8.1%
Residual variation	12.9%	14.6%	14.4%	14.1%	13.8%	14.2%
Correlations (and contr. to overall)						
Corr(θ_i,ψ_{jt})	0.143	0.177	0.19	0.2	0.119	0.145
Corr(ε_i^I,ε_{jt}^J)	0.113	0.124	0.152	0.141	0.131	0.151
Explained by 2Cov(θ_i,ψ_{jt})	8.4%	9.5%	9.8%	10.2%	6.7%	8.0%
Explained by 2Cov(ε_i^I,ε_{jt}^J)	4.2%	4.1%	4.3%	4.4%	3.6%	4.3%
Between-within decomposition						
Between-firm share	47.6%	46.9%	47.2%	47.8%	47.2%	47.4%
— Ind. segregation	10.9%	10.3%	10.2%	8.9%	13.9%	12.0%
— Var(ψ_j)	20.3%	18.8%	18.2%	19.2%	18.5%	18.6%
— Sorting	8.7%	9.6%	9.8%	10.2%	6.7%	8.0%
Sample						
Interacting variable	year(15)	gender(2)	educ.(3)	occup.(7)	tenure(4)	age(4)
Number of Observations (1000)	62441	69982	69304	71064	41899	63964
Number of Firms (1000)	141	142	144	163	108	142
Number of Workers (1000)	2387	2584	2554	2620	2074	2368

Notes: See Table 1. The first stages are estimated according to Equation 18, and then decomposed according to Equation 19.

Indeed, relaxing the assumptions of the AKM model in our setting provided a model fit better only by less than two percentage points. By regressing firm-year joint effects on separate firm and year effects, we can observe that the separate components, mostly the firm effects, explain 92% of the variation in the joint parameters. The remaining 8% accounts for time variation of wage premia per year within firms. So, although conclusions regarding wage sorting remain mostly unaffected, an argument could be made that we may have a drift in at least part of the firm effects. If that is the case, the rolling-AKM approach, such as the one implemented in Table 4 may provide a better characterization of the labor market – given that the limited mobility bias from using a shorter panel is corrected for.

Gender and education

A specification linking the standard AKM and the previously presented match model, in which all individual-firm match can have a different wage component, could be defined by assuming that firm-effects are heterogeneous across different groups of individuals, based either on exogenous (gender) or endogenous (educational level) person characteristics. However, the estimation of such models is not straightforward. As these variables are time-invariant for individuals in our data, there is not any mobility between the two or three distinct sets of firm-gender or firm-education units. Accordingly, we will have two or three, non-overlapping components in the mobility network, and as one normalizing condition is required in all connected components, female and male firm effects (or those of different educational groups) will not necessarily be measured on the same scale, rendering them incomparable. Hence, we follow Card et al. (2016) and re-scale the estimated effects along the assumption that the least productive firms achieve no rents, and hence compensate their workers independent of their characteristics. [50] Appendix Figure A4 illustrates the relationship between the re-scaled effects and productivity, suggesting different rent-sharing elasticities for individuals of different gender and education.[51] While the importance of this approach regarding differential rent-sharing is evident from this graph, Table 8 suggests that the model fares only slightly better than the standard AKM in explaining overall wage variation. While the explained share of observed individual heterogeneity decreases, firm-group heterogeneity increases with roughly the same extent. The component which could not be explained previously by orthogonal gender, education and firm effects accounts for 0.5% or 1% of the overall wage distribution.

Jobs as unit of interest

The main specification in our paper, and previously the findings of Torres et al. (2018), already highlighted that besides who works for whom, what people work is also an important factor in explaining wage variation. A remaining question to answer is whether it matters where someone does what she does. In a previous model (Appendix Table A8), we already assumed that working in the same occupation may be rewarded differently in different industries. As a generalization of this concept, we could also assume that the work done in a given occupation differs not only across sectors, but in every firm. Some studies in wage inequality, for instance Petersen and Morgan (1995) and Petersen et al. (1997) or Avent-Holt et al. (2020) already treat 'jobs' – defined as firm-occupation interacted categories – as a relevant level of investigating wage inequalities, albeit without controlling for unobserved individual heterogeneity.

[50]Specifically, by fitting a kinked function on our data, we also identify a set of firms for whom the increase of productivity, measured by value added per worker is not reflected in an increase of AKM firm effects. Then, assuming that firm effects of gender or education groups should be equal in these 'no surplus' firms, we normalize firm effects across groups, so that their average will be the same for this set of firms.

[51]These differences are discussed in detail in Boza (n.d.) on differential rent-sharing.

In the AKM setting, this approach relaxes the assumption of having the same wage premia for all different jobs within the firm, in a similar fashion as one can allow for different premia based on gender or race. We provide the first results from estimating this specification with both person and job effects, the latter being defined by firm times occupation (seven, 1-digit categories) interaction fixed effects, revealing that firm premia indeed varies between different groups of workers within the same firm.[52] Of the variance of the estimated joint (job) effects, more than 6% originates from sources other than the baseline difference of wage levels across occupations or firms.[53] Of the overall wage variation, this can explain 2.2%, while the magnitude of person, firm and occupation effects remain roughly the same as in the main estimations. The possible applications of this design includes, for instance, the investigation of differential rent-sharing and bargaining across occupations and the detailed assessment of the gender wage gap and differential sorting of women across occupations.

Seniority of workers

Before discussing the final set of specifications, we note that with the TV-AKM, the match model, and the job model, we have derived three main alternative parametrization of the four fixed effects in Equation 1, combining firms with time, persons and occupations respectively. Along these four dimensions, higher order combinations of fixed effects are also possible. For instance, one could define gender-firm-occupation interactions for assessing specific problems.Finally, one could also combine firm effects with the elements of time-varying characteristics, X. In this exercise, we divide firms into sub-units based on the seniority of its current workers.[54]

First, we define four groups based on completed tenure at the firm, with the categories consisting of those with less than one, three or five years of tenure, and those who had spent more than 60 months already at the given firm – although not necessarily in the same job. Accordingly, in this setup we can only use the last ten years of data, as the completed tenure category can not be defined properly for preceding observations. In a second model we use the age of the workers to form four categories, the cutoffs being at 26, 41 and 56 years. The results, presented in the last two columns of Table 8 suggest a very limited role of between-firm differences in how firms treat their workers of different seniority, as firm-tenure interactions and firm-age interaction can both explain less than 0.5% of overall wage variation.

[52] Identification of the job cell effects relies both on between-firm mobility and within-firm promotions or re-specializations. Also, as people can move between different occupations, we don't face a scaling problem as before, and estimated firm effects are directly comparable within the largest connected set.

[53] An alternative specification, with 37 distinct (2-digit) occupation categories yielded generally similar results, with 9% of the job effects attributable to the second stage residuals.

[54] The gender-firm, education-firm and occupation-sector models could be considered as special cases, in which one variable of the observable part of the person or firm characteristic is removed from its corresponding fixed effect, and is interacted with one of the high-dimensional variables of the first-stage AKM equation.

5.3 Bargaining and sorting in G-AKM

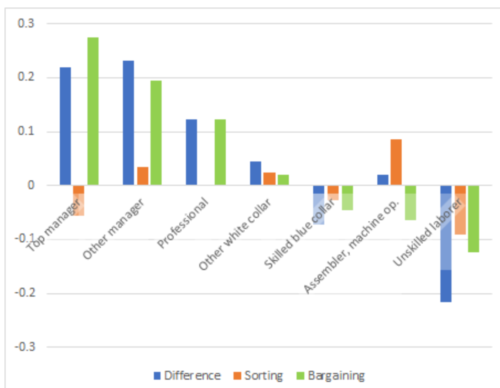
Finally, we present a simple application of the previously introduced grouped AKM specifications, building on Card et al. (2016). Namely, we will use the alternative estimation of the bargaining-sorting decomposition of firm effects, we proposed in Equations 18 through 20. We will present the average differences in Ψ_{jg} and ψ_j, the sorting effect, alongside the bargaining effect β_g, using the fact that this parameter from Equation 20 can be rewritten as

$$\tilde{\beta}_g = \frac{\partial(\Psi_{jg} - \psi_j)}{\partial G} \tag{21}$$

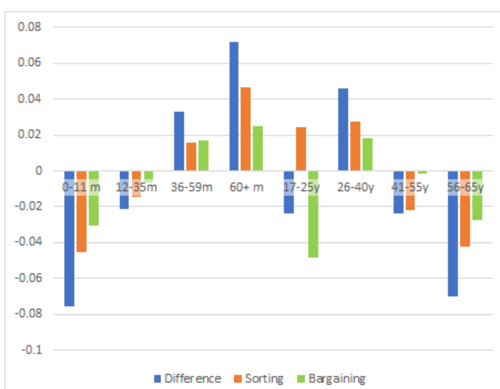
We present the results graphically in Figure 3, with bars representing deviations from the sample-level mean firm-group effects (scaled to zero), instead of arbitrarily chosen reference categories. Hence, to directly compare any two groups, the difference between their respective firm effect components should be considered.[55] The graphs clearly suggest that sorting can quite substantially form wage-level patterns across different observed characteristics. Similarly as Card et al. (2016) – and later by Casarico and Lattanzio (2019) and Lamadon et al. (2019) (in their Appendix) –, the difference between average male and female firm effects are largely attributable to the sorting channel, with bargaining comprising only around 20% of the overall difference. Regarding the educational attainment of individuals, we observe a much weaker role, as most of the wage differences are generated within and not across firms, like Table 6 also suggested earlier. The same holds also for broad job categories within firms, regarding which, sorting can even counteract the within firm differences. For instance, while assemblers and machine operators are slightly underpaid in their firms, they usually work at employers with the highest wage premia (such as multinational car manufacturers), leading to an average lower job effect than – more skilled – blue-collar workers. Sorting patterns with regards to tenure probably reflect high wage firms having lower turnover rates. Therefore those with long employment history in a given firm earn more not only because the within-firm wage-path (reflected in the increasing bargaining component), but because such spells occur more probably in higher premia firms. Finally, while the standard age-earnings profile are reflected in the overall and bargaining components, it seems that younger workers sort more frequently into high wage firms, either due to supply or demand-side factors.

[55] We also note, that these parameters are not controlled for other X variables, so are not directly comparable to Table 6. Appendix B reflects on this issue in the context of gender differences. The minor differences between Appendix Table B1 and Figure 3 are due to the slightly different sub-sample used, as the Appendix exercise is constrained to the dual-connected sample (for comparative reasons), whereas here we include all firms for which firm-gender effects are identified, even if they are gender-segregated.

(a) Gender and proxied education



(b) Occupations



(c) Completed tenure and age

Figure 3: Bargaining and sorting effects based on G-AKM models

Notes: The bars represent the mean values of Ψ_{jg} (*difference*), $\tilde{\psi}_j$ (*sorting*) and ε^G_{jg} (*bargaining*) from Equation 19 across the given categories. The firm effects with gender or education interactions are normalized to have zero as their mean in the full sample.

6 Discussion

Beside providing evidence on a substantial level of wage sorting in Hungary, throughout the different exercises presented in this paper, we were able to identify some of the observable channels in which wage sorting could emerge. For instance, it turns out that firms of foreign majority ownership are quite important contributors to inequality as they both pay high wage premia, even within the sectors they operate in, and are able to hire high wage workers. Second, the sorting problem clearly has some spatial aspect as well, with highest earners and high-paying firms both being over-represented in Budapest, with sorting itself also being stronger within (and around) the capital. Whether these descriptive patterns are driven by the residential mobility of skilled individuals, inter-generational inheritance of residence and positions or the endogenous location choice of (new) firms remains an open question. Not surprisingly, education is also a main driver of wage differences, but our results suggest that these differences are not only due to the accumulation of human capital, as higher education levels clearly open the door into high wage firms as well. Gender differences persist both within and across firms, with male workers being over-represented in firms which both employ higher wage workers on average and offer a higher wage premia as well. However, how much of the between firm differences are due to discriminatory hiring and what fraction could be explained by different preferences of male and female workers for certain amenities or disamenities of workplaces also remains an issue to be investigated in detail. As besides the sorting channels, within-firm differences in the remuneration of workers of different types turned out to be important in all observable aspects we investigate, the question of the extent to which differential rent sharing could account for such differences arises naturally.

Even after accounting for these observable channels, around half of the overall wage sorting remains as an unexplained, intra-industry sorting component of workers with better unobserved qualities, responsible for almost 5% of overall wage variation. Even if this share largely depends on the availability of worker or firm specific features in the dataset used by the econometrician, the further understanding of the unexplained component may be also a research aim to pursue, by collecting or linking additional data sources for instance. One such feature we were able to utilize was data on the (high-school) test scores of young workers. We have observed that within the same high-schools better students will end up not only in better occupations in the future, but also at better firms as well. However, our results also suggested that being the student of given schools is already indicative of future prospects, and hence high schools may play a strong allocative role as well. An aim to pursue by future studies can be to investigate whether some high schools are indeed able to funnel their students into better firms through providing more human capital, good labor market signals or access to superior social networks, or the sorting between high wage schools and high wage firms is just an empirical artifact caused by selection already present at the time of high-school admission.

References

Abowd, J. M., Creecy, R. H., & Kramarz, F. (2002). Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data. *U.S. Census Bureau*, *2002-06*, 1–15. https://www2.census.gov/ces/tp/tp-2002-06.pdf

Abowd, J. M., & Kramarz, F. (2004). Are Good Workers Employed by Good Firms? A Simple Test of Positive Assortative Matching Models. https://econpapers.repec.org/RePEc:ecm:nawm04:385

Abowd, J. M., Kramarz, F., & Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, *67*(2), 251–333. https://doi.org/10.1111/1468-0262.00020

Abowd, J. M., McKinney, K. L., & Schmutte, I. M. (2019). Modeling Endogenous Mobility in Earnings Determination. *Journal of Business and Economic Statistics*, *37*(3), 405–418. https://doi.org/10.1080/07350015.2017.1356727

Addison, J., Portugal, P., & Vilares, H. (2018). The Sources of the Union Wage Gap: The Role of Worker, Firm, Match, and Job-title Heterogeneity. *CESifo Working Paper*, (7392). https://papers.ssrn.com/sol3/papers.cfm?abstract%7B%5C_%7Did=3338746

Alvarez, J., Benguria, F., Engbom, N., & Moser, C. (2018). Firms and the decline in earnings inequality in Brazil. *American Economic Journal: Macroeconomics*, *10*(1), 149–189. https://doi.org/10.1257/mac.20150355

Andrews, M. J., Gill, L., Schank, T., & Upward, R. (2008). High wage workers and low wage firms: negative assortative matching or limited mobility bias? *171*(3), 673–697. https://doi.org/10.1111/j.1467-985X.2007.00533.x

Andrews, M. J., Gill, L., Schank, T., & Upward, R. (2012). High wage workers match with high wage firms: Clear evidence of the effects of limited mobility bias. *Economics Letters*, *117*(3), 824–827. https://doi.org/10.1016/j.econlet.2012.07.033

Avent-Holt, D., Henriksen, L. F., Hägglund, A. E., Jung, J., Kodama, N., Melzer, S. M., Mun, E., Rainey, A., & Tomaskovic-Devey, D. (2020). Occupations, workplaces or jobs?: An exploration of stratification contexts using administrative data. *Research in Social Stratification and Mobility*, *70*, 100456. https://doi.org/10.1016/j.rssm.2019.100456

Bagger, J., & Lentz, R. (2019). An Empirical Model of Wage Dispersion with Sorting. *Review of Economic Studies*, *86*(1), 153–190. https://doi.org/10.1093/restud/rdy022

Bonhomme, S., Holzheu, K., Lamadon, T., Manresa, E., Mogstad, M., & Setzler, B. (2020). How Much Should We Trust Estimates of Firm Effects and Worker Sorting? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3625745

Bonhomme, S., Lamadon, T., Manresa, E., Kline, P., Hansen, C., Sølvsten, M., Altonji, J., Graham, B., Lise, J., Lopes De Melo, R., Meghir, C., Mogstad, M., Neal, D., Pistaferri, L., Robin, J.-M., Rosen, A., Saggio, R., Schoenberg, U., & Shimer, R. (2019). A Distributional Framework for Matched Employer Employee Data. *Econometrica*, *87*(3), 699–739. https://doi.org/10.3982/ecta15722

Boza, I. (n.d.). *Heterogenous and differential rent-sharing, a fixed-effects approach* [Work in progress.].

Boza, I., & Ilyés, V. (2020). Decomposition of co-worker wage gains. *IZA Journal of Labor Economics*, *9*(1). https://doi.org/10.2478/izajole-2020-0008

Card, D., Cardoso, A. R., Heining, J., & Kline, P. (2018). Firms and Labor Market Inequality: Evidence and Some Theory. *Journal of Labor Economics*, *36*(S1), S13–S70. https://doi.org/10.1086/694153

Card, D., Cardoso, A. R., & Kline, P. (2016). Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women. *Quarterly Journal of Economics*, *131*(2), 633–686. https://doi.org/10.1093/qje/qjv038

Card, D., Heining, J., & Kline, P. (2013). Workplace Heterogeneity and the Rise of West German Wage Inequality. *The Quarterly Journal of Economics*, *128*(3), 967–1015. https://doi.org/10.1093/qje/qjt006

Cardoso, A. R., Guimarães, P., & Portugal, P. (2016). What drives the gender wage gap? A look at the role of firm and job-title heterogeneity. *Oxford Economic Papers*, *68*(2), 506–524. https://doi.org/10.1093/oep/gpv069

Cardoso, A. R., Guimarães, P., Portugal, P., & Reis, H. (2018). The Returns to Schooling Unveiled. *IZA Discussion Papers*, (11419), 1–45. https://ideas.repec.org/p/iza/izadps/dp11419.html

Casarico, A., & Lattanzio, S. (2019). What Firms Do: Gender Inequality in Linked Employer-Employee Data. https://www.repository.cam.ac.uk/handle/1810/294660

Correia, S. (2017). Linear Models with High-dimensional Fixed Effects: An Efficient and Feasible Estimator. *Working Paper*.

Dauth, W., Findeisen, S., Moretti, E., & Suedekum, J. (2019). *Matching in Cities* (Working Paper No. 25227). National Bureau of Economic Research. https://doi.org/10.3386/w25227

Devicienti, F., Fanfani, B., & Maida, A. (2016). Collective Bargaining and the Evolution of Wage Premia Dispersion in Italy. (10293), 1–44. http://ftp.iza.org/dp10293.pdf

Earle, J. S., Telegdy, Á., & Antal, G. (2018). Foreign Ownership and Wages: Evidence from Hungary, 1986–2008. *ILR Review*, *71*(2), 458–491. https://doi.org/10.1177/0019793917700087

Eeckhout, J., & Kircher, P. (2011). Identifying Sorting—In Theory. *The Review of Economic Studies*, *78*(3), 872–906. https://doi.org/10.1093/RESTUD/RDQ034

Engbom, N., & Moser, C. (2021). *Earnings Inequality and the Minimum Wage: Evidence from Brazil* (tech. rep.). National Bureau of Economic Research. Cambridge, MA. https://doi.org/10.3386/w28831

Fanfani, B. (2018). Tastes for Discrimination in Monopsonistic Labour Markets.

Gelbach, J. B. (2016). When Do Covariates Matter? And Which Ones, and How Much? *https://doi.org/10.1086/683668*, *34*(2), 509–543. https://doi.org/10.1086/683668

Gerard, F., Lagos, L., Severnini, E., & Card, D. (2021). Assortative Matching or Exclusionary Hiring? The Impact of Employment and Pay Policies on Racial Wage Differences in Brazil. *American Economic Review*. https://doi.org/10.1257/AER.20181596

Goldschmidt, D., & Schmieder, J. F. (2017). The rise of domestic outsourcing and the evolution of the german wage structure. *Quarterly Journal of Economics*, *132*(3), 1165–1217. https://doi.org/10.1093/qje/qjx008

Goux, D., & Maurin, E. (1999). Persistence of interindustry wage differentials: A reexamination using matched worker-firm panel data. *Journal of Labor Economics*, *17*(3), 492–533. https://doi.org/10.1086/209929

Gruetter, M., & Lalive, R. (2009). The importance of firms in wage determination. *Labour Economics*, *16*(2), 149–160. https://doi.org/10.1016/j.labeco.2008.09.001

Guimarães, P., & Portugal, P. (2010). A simple feasible procedure to fit models with high-dimensional fixed effects. *Stata Journal*, *10*(4), 628–649. https://doi.org/10.1177/1536867x1101000406

Gyetvai, A. (2017). *High Wage Workers, High Wage Firms, and High Wage Occupations* (Doctoral dissertation No. *2*).

Hermann, Z., Horn, D., Köllő, J., Sebők, A., Semjén, A., & Varga, J. (2019). Szövegértési és matematikai kompetencia hatása akeresetre és foglalkoztatási esélyekre. *Munkaerőpiaci Tükör 2018*, 45–52.

Iranzo, S., Schivardi, F., & Tosetti, E. (2008). Skill dispersion and firm productivity: An analysis with employer-employee matched data. *Journal of Labor Economics*, *26*(2), 247–285. https://doi.org/10.1086/587091

Kline, P., Saggio, R., & Sølvsten, M. (2020). Leave-Out Estimation of Variance Components. *Econometrica*, *88*(5), 1859–1898. https://doi.org/10.3982/ecta16410

Köllő, J., Boza, I., & Balázsi, L. (2021). Wage gains from foreign ownership: evidence from linked employer–employee data. *Journal for Labour Market Research 2021 55:1*, *55*(1), 1–21. https://doi.org/10.1186/S12651-021-00286-0

Lachowska, M., Mas, A., Saggio, R., & Woodbury, S. (2020). Do Firm Effects Drift? Evidence from Washington Administrative Data. https://doi.org/10.3386/w26653

Lamadon, T., Mogstad, M., & Setzler, B. (2019). Imperfect Competition , Compensating Differentials and. https://doi.org/10.3386/w25954

Lopes De Melo, R. (2018). Firm wage differentials and labor market sorting: Reconciling theory and evidence. *Journal of Political Economy*, *126*(1), 313–346. https://doi.org/10.1086/695505

Macis, M., & Schivardi, F. (2016). *Exports and Wages: Rent Sharing, Workforce Composition, or Returns to Skills?* (Tech. rep.). http://www.journals.uchicago.edu/t-and-c

Petersen, T., & Morgan, L. A. (1995). Separate and unequal: occupation-establishment sex segregation and the gender wage gap. *American Journal of Sociology*, *101*(2), 329–365. https://doi.org/10.1086/230727

Petersen, T., Snartland, V., Becken, L.-E., & Olsen, K. M. (1997). Within-Job Wage Discrimination and the Gender Wage Gap: The Case of Norway. *European Sociological Review*, *13*(2), 199–213. https://doi.org/10.1093/OXFORDJOURNALS.ESR.A018212

Song, J., Price, D. J., Guvenen, F., Bloom, N., & Von Wachter, T. (2019). Firming Up Inequality. *The Quarterly Journal of Economics*, *134*(1), 1–50. https://doi.org/10.1093/qje/qjy025

Sørensen, K. L., & Vejlin, R. (2011). Worker and Firm Heterogeneity in Wage Growth: An AKM Approach. *Labour*, *25*(4), 485–507. https://doi.org/10.1111/j.1467-9914.2011.00527.x

Sørensen, T., & Vejlin, R. (2013). The importance of worker, firm and match effects in the formation of wages. *Empirical Economics*, *45*(1), 435–464. https://doi.org/10.1007/s00181-012-0603-3

Sorkin, I. (2018). Ranking firms using revealed preference. *Quarterly Journal of Economics*, *133*(3), 1331–1393. https://doi.org/10.1093/QJE/QJY001

Tomaskovic-Devey, D., Rainey, A., Avent-Holt, D., Bandelj, N., Boza, I., Cort, D., Godechot, O., Hajdu, G., Hällsten, M., Henriksen, L. F., Hermansen, A. S., Hou, F., Jung, J., Kanjuo-Mrčela, A., King, J., Kodama, N., Kristal, T., Krízková, A., Lippényi, Z., … Tufail, Z. (2020). Rising between-workplace inequalities in high-income countries. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(17), 9277–9283. https://doi.org/10.1073/pnas.1918249117

Torres, S., Portugal, P., Addison, J. T., & Guimarães, P. (2018). The sources of wage variation and the direction of assortative matching: Evidence from a three-way high-dimensional fixed effects regression model. *Labour Economics*, *54*, 47–60. https://doi.org/10.1016/j.labeco.2018.06.004

Weeks, D. L., & Williams, D. R. (1964). A Note on the Determination of Connectedness in an N-Way Cross Classification. *Technometrics*, *6*(3), 319–324. https://doi.org/10.1080/00401706.1964.10490188

Woodcock, S. D. (2008). Wage differentials in the presence of unobserved worker, firm, and match heterogeneity. *Labour Economics*, *15*(4), 771–793. https://doi.org/10.1016/j.labeco.2007.06.003

Woodcock, S. D. (2015). Match effects. *Research in Economics*, *69*(1), 100–121. https://doi.org/10.1016/j.rie.2014.12.001

Appendices

Appendix A - Additional tables and graphs

Table A1: Wage variance decompositions based on AKM methods

Article	Country	Period (years)	Sample	Method	Bias-c.	Var(w)	Var(θ)	Var(ψ)	Cov(θ,ψ)	R(θ,ψ)
Gruetter and Lalive (2009)	Austria	1990-1997 (8)		AKM		22.4%	66.3%	37.0%	-22.5%	-0.27
Bonhomme et al. (2020)	Austria	2010-2015 (6)	BLM corrected	AKM	BLM	18.7%	n.a.	11.7%	19.6%	n.a.
Lopes De Melo (2018)	Brazil	1995-2005 (11)		AKM		60.1%	66.6%	30.0%	3.6%	0.04
Alvarez et al. (2018)	Brazil	2008-2012 (5)		AKM		47.0%	57.4%	14.9%	19.1%	0.33
Engbom and Moser (2021)	Brazil	2010-2014 (5)	KSS corrected	AKM	KSS	45.3%	29.4%	17.0%	15.2%	0.34
Gerard et al. (2021)	Brazil	2002-2014 (13)	White male	G-AKM	KSS	44.9%	36.3%	16.4%	22.8%	0.47
Gerard et al. (2021)	Brazil	2002-2014 (13)	Non-white male	G-AKM	KSS	33.2%	30.0%	16.4%	17.6%	0.40
Gerard et al. (2021)	Brazil	2002-2014 (13)	White female	G-AKM	KSS	49.8%	43.9%	15.0%	24.6%	0.48
Gerard et al. (2021)	Brazil	2002-2014 (13)	Non-white female	G-AKM	KSS	32.4%	44.4%	14.0%	18.4%	0.37
Bagger and Lentz (2019)	Denmark	1985-2003 (19)		AKM		9.7%	72.2%	14.4%	-2.1%	-0.03
T. Sørensen and Vejlin (2013)	Denmark	1985-2003 (19)	AKM	AKM		11.1%	33.1%	4.6%	-1.5%	-0.06
T. Sørensen and Vejlin (2013)	Denmark	1985-2003 (19)	Match	M-AKM		11.3%	50.3%	5.5%	-1.8%	-0.05
K. L. Sørensen and Vejlin (2011)	Denmark	1980-2006 (27)		AKM		9.4%	57.0%	13.0%	1.6%	0.03
Abowd et al. (2002)	France	1976-1987 (12)		AKM		26.9%	76.9%	30.2%	-27.2%	-0.28
Abowd et al. (1999)	France	1976-1987 (12)		AKM		26.9%	69.8%	87.0%	46.2%	0.30
Goux and Maurin (1999)	France	1993-1995 (3)		AKM		15.1%	79.3%	19.6%	1.3%	0.01
Abowd and Kramarz (2004)	France	1976-1996 (21)		AKM		35.4%	70.3%	61.5%	-31.8%	-0.24
Andrews et al. (2008)	Germany	1993-1997 (5)	Bias corrected	AKM	AGSU	5.5%	92.0%	21.6%	-13.2%	-0.15
Goldschmidt and Schmieder (2017)	Germany	2008-2008 (1)		AKM		20.5%	n.a.	26.7%	20.8%	n.a.
Card et al. (2013)	Germany	2002-2009 (8)		AKM		24.9%	51.0%	21.3%	16.5%	0.25
This study: Boza (2021)	Hungary	2003-2017 (15)		3W-AKM		33.8%	38.5%	18.3%	9.3%	0.18
Macis and Schivardi (2016)	Italy	1982-1997 (16)		TV-AKM		11.6%	49.8%	14.6%	-1.1%	-0.02
Iranzo et al. (2008)	Italy	1981-1997 (17)		AKM		11.0%	43.9%	13.1%	2.1%	0.04
Kline et al. (2020)	Italy	1999-2001 (3)	KSS	AKM	KSS	18.4%	60.8%	13.0%	16.0%	0.28
Fanfani (2018)	Italy	1996-2001 (6)	Female	G-AKM		7.9%	62.0%	19.0%	-6.3%	-0.09
Fanfani (2018)	Italy	1996-2001 (6)	Male	G-AKM		14.1%	78.7%	9.9%	5.0%	0.09
Devicienti et al. (2016)	Italy	1996-2001 (6)		AKM		13.1%	99.2%	18.3%	-4.6%	-0.05
Bonhomme et al. (2020)	Italy	1996-2001 (6)	BLM corrected	AKM	BLM	16.7%	n.a.	12.7%	20.0%	n.a.
Casarico and Lattanzio (2019)	Italy	1995-2015 (21)	Male	G-AKM		23.6%	185.0%	18.5%	-5.0%	-0.04
Casarico and Lattanzio (2019)	Italy	1995-2015 (21)	Female	G-AKM		17.2%	187.3%	22.1%	0.0%	0.00
Bonhomme et al. (2020)	Norway	2009-2014 (6)	BLM corrected	AKM	BLM	23.9%	n.a.	11.8%	16.8%	n.a.
Card et al. (2018)	Portugal	2005-2009 (5)		AKM		27.5%	64.1%	22.7%	13.0%	0.17
Card et al. (2016)	Portugal	2002-2009 (8)	Male	G-AKM		30.7%	57.5%	19.9%	11.3%	0.17
Card et al. (2016)	Portugal	2002-2009 (8)	Female	G-AKM		26.3%	60.8%	17.2%	9.8%	0.15
Torres et al. (2018)	Portugal	1986-2009 (24)	No job effects	AKM		32.3%	75.0%	18.0%	16.8%	0.23
Torres et al. (2018)	Portugal	1986-2009 (24)	With job effects	3W-AKM		32.3%	38.0%	16.1%	9.7%	0.20
Bonhomme et al. (2020)	Sweden	2000-2005 (6)	BLM corrected	AKM	BLM	16.4%	n.a.	5.0%	10.3%	n.a.
Sorkin (2018)	US	1990-1999 (10)		AKM		67.0%	51.0%	14.0%	10.0%	0.19
Abowd and Kramarz (2004)	US	2000-2008 (9)		AKM		80.0%	78.7%	16.3%	1.5%	0.02
Song et al. (2019)	US	2007-2013 (7)		AKM		92.4%	51.5%	8.8%	11.7%	0.28
Woodcock (2015)	US	2007-2013 (7)	No match	AKM		41.0%	71.0%	19.5%	-1.0%	-0.01
Woodcock (2015)	US	2007-2013 (7)	Orth. match	M-AKM		41.0%	70.7%	19.8%	-0.6%	-0.01
Bonhomme et al. (2020)	US	2010-2015 (6)	BLM corrected	AKM	BLM	41.4%	n.a.	6.2%	15.0%	n.a.
Lamadon et al. (2019)	US	2001-2015 (15)	AKM, BLM	AKM	BLM	45.0%	72.4%	3.2%	13.1%	0.43
Lamadon et al. (2019)	US	2001-2015 (15)	TV-AKM, BLM	TV-AKM	BLM	45.0%	73.5%	3.3%	13.4%	0.43
Abowd et al. (2002)	US	1990-2000 (11)		AKM		27.8%	81.6%	19.2%	-2.0%	-0.03
Lachowska et al. (2020)	US	2002-2014 (13)	AKM , KSS	AKM	KSS	40.7%	61.4%	11.6%	16.9%	0.32
Lachowska et al. (2020)	US	2002-2014 (13)	TV-AKM, KSS	TV-AKM	KSS	40.7%	62.2%	13.5%	14.8%	0.26

Notes: G-AKM refers to estimating AKM on different groups, 3W-AKM refers to AKM estimations with high dimensional occupation effects included, TV-AKM refers to time-varying AKM which utilizes firm-year effects, M-AKM to models with match effects. KSS bias-correction refers to the method of Kline et al. (2020), BLM to Bonhomme et al. (2019) and AGSU to Andrews et al. (2008).

Table A2: Comparison of bias-corrected and standard results from the literature

Article	Country	Period (years)	Sample	Method	Bias-c.	Var(w)	Var(θ)	Var(ψ)	Cov(θ, ψ)	R(θ, ψ)
Bonhomme et al. (2020)	Austria	2010-2015 (6)		AKM	BLM	18.7%	n.a.	11.7%	19.6%	n.a.
Bonhomme et al. (2020)	Austria	2010-2015 (6)		AKM		18.7%	n.a.	18.7%	4.7%	n.a.
Gerard et al. (2021)	Brazil	2002-2014 (13)	Non-white F	G-AKM	KSS	32.4%	44.4%	14.0%	18.4%	0.37
Gerard et al. (2021)	Brazil	2002-2014 (13)	Non-white F	G-AKM		30.7%	62.2%	23.1%	7.7%	0.10
Gerard et al. (2021)	Brazil	2002-2014 (13)	Non-white M	G-AKM	KSS	33.2%	3n.a.	16.4%	17.6%	0.40
Gerard et al. (2021)	Brazil	2002-2014 (13)	Non-white M	G-AKM		33.9%	50.8%	23.0%	11.4%	0.17
Gerard et al. (2021)	Brazil	2002-2014 (13)	White F	G-AKM	KSS	49.8%	43.9%	15.0%	24.6%	0.48
Gerard et al. (2021)	Brazil	2002-2014 (13)	White F	G-AKM		46.9%	59.2%	19.7%	18.0%	0.26
Gerard et al. (2021)	Brazil	2002-2014 (13)	White M	G-AKM	KSS	44.9%	36.3%	16.4%	22.8%	0.47
Gerard et al. (2021)	Brazil	2002-2014 (13)	White M	G-AKM		44.9%	52.2%	20.6%	18.0%	0.28
Engbom and Moser (2021)	Brazil	2010-2014 (5)		AKM	KSS	45.3%	29.4%	17.0%	15.2%	0.34
Engbom and Moser (2021)	Brazil	2010-2014 (5)		AKM		45.3%	34.0%	18.1%	13.5%	0.27
Andrews et al. (2008)	Germany	1993-1997 (5)		AKM	AGSU	5.5%	92.0%	21.6%	-13.2%	-0.15
Andrews et al. (2008)	Germany	1993-1997 (5)		AKM		5.7%	94.4%	23.5%	-18.0%	-0.19
Bonhomme et al. (2020)	Italy	1996-2001 (6)		AKM	BLM	16.7%	n.a.	12.7%	20.0%	n.a.
Bonhomme et al. (2020)	Italy	1996-2001 (6)		AKM		16.7%	n.a.	23.1%	-1.3%	n.a.
Kline et al. (2020)	Italy	1999-2001 (3)		AKM	KSS	18.4%	60.8%	13.0%	16.0%	0.28
Kline et al. (2020)	Italy	1999-2001 (3)		AKM		18.4%	71.8%	19.5%	4.2%	0.06
Bonhomme et al. (2020)	Norway	2009-2014 (6)		AKM	BLM	23.9%	n.a.	11.8%	16.8%	n.a.
Bonhomme et al. (2020)	Norway	2009-2014 (6)		AKM		23.9%	n.a.	24.4%	-7.7%	n.a.
Bonhomme et al. (2020)	Sweden	2000-2005 (6)		AKM	BLM	16.4%	n.a.	5.0%	10.3%	n.a.
Bonhomme et al. (2020)	Sweden	2000-2005 (6)		AKM		16.4%	n.a.	14.6%	-8.1%	n.a.
Lachowska et al. (2020)	US	2002-2014 (13)		AKM	KSS	40.7%	61.4%	11.6%	16.9%	0.32
Lachowska et al. (2020)	US	2002-2014 (13)		AKM		40.7%	63.0%	11.8%	16.7%	0.31
Lachowska et al. (2020)	US	2002-2014 (13)		TV-AKM	KSS	40.7%	62.2%	13.5%	14.8%	0.26
Lachowska et al. (2020)	US	2002-2014 (13)		TV-AKM		40.7%	63.7%	14.0%	14.5%	0.24
Lamadon et al. (2019)	US	2001-2015 (15)		AKM	BLM	45.0%	72.4%	3.2%	13.4%	0.44
Lamadon et al. (2019)	US	2001-2015 (15)		AKM		45.0%	75.0%	9.0%	5.2%	0.10
Bonhomme et al. (2020)	US	2010-2015 (6)		AKM	BLM	41.4%	n.a.	6.2%	15.0%	n.a.
Bonhomme et al. (2020)	US	2010-2015 (6)		AKM		41.4%	n.a.	12.2%	1.1%	n.a.

Notes: G-AKM refers to estimating AKM on different groups, 3W-AKM refers to AKM estimations with high dimensional occupation effects included, TV-AKM refers to time-varying AKM which utilizes firm-year effects, M-AKM to models with match effects. KSS bias-correction refers to the method of Kline et al. (2020), BLM to Bonhomme et al. (2019) and AGSU to Andrews et al. (2008).

Table A3: Decomposition of wage variance, based on Song et al. (2019)

Variance of log wages	0.338	
Between-firm Variance	**0.160**	47.5%
— *Var*($\bar{\theta}$)	*0.038*	11.3%
— *Var*(ψ)	*0.062*	18.3%
— Var($\bar{\lambda}$)	*0.003*	0.8%
— Var($\bar{X}\beta$)	*0.006*	1.7%
— *2Cov*($\bar{\theta}$,ψ)	*0.031*	9.3%
— 2Cov($\bar{\lambda}$,$\bar{\theta}$)	*0.014*	4.1%
— 2Cov($\bar{\lambda}$,ψ)	*-0.003*	-0.8%
— 2Cov($\bar{X}\beta$,$\bar{\theta}$)	*0.008*	2.4%
— 2Cov($\bar{X}\beta$,ψ)	*-0.002*	-0.7%
— 2Cov($\bar{X}\beta$,$\bar{\lambda}$)	*0.004*	1.1%
Within-firm Variance	**0.177**	52.5%
— Var($(\theta - \bar{\theta})$)	*0.092*	27.3%
— Var($(\lambda - \bar{\lambda})$)	*0.005*	1.5%
— Var($(X - \bar{X})\beta$)	*0.007*	2.1%
— Var(ε)	*0.049*	14.6%
— 2Cov($(\lambda - \bar{\lambda})$,$(\theta - \bar{\theta})$)	*0.022*	6.6%
— 2Cov($(X - \bar{X})\beta$,$(\theta - \bar{\theta})$)	*0.001*	0.2%
— 2Cov($(X - \bar{X})\beta$,$(\lambda - \bar{\lambda})$)	*0.001*	0.2%
— 2Cov(ε,$(\theta - \bar{\theta})$)	*0.000*	0.0%
— 2Cov(ε,$(\lambda - \bar{\lambda})$)	*0.000*	0.0%
— 2Cov(ε,$(X - \bar{X})\beta$)	*0.000*	0.0%
Number of Observations (1000)	66155	
Number of Firms (1000)	146	
Number of Workers (1000)	2462	

Notes: Decomposition is based on 7.

Table A4: Decomposition of wage variance, estimated on random sub-samples

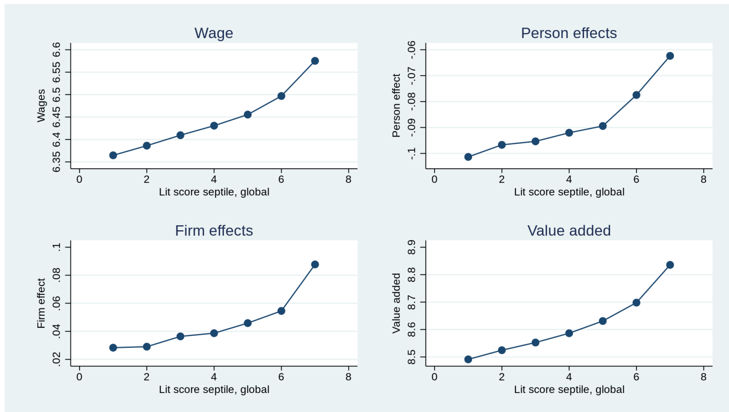
Variance of log wages	0.338	0.336	0.335	0.337	0.339	0.338
Ensemble decomp.						
Contribution of XB	5.40%	5.33%	5.14%	5.40%	5.39%	5.99%
— Year	1.98%	1.95%	1.92%	1.99%	1.97%	2.21%
— age*, firm size, contract, tenure*	3.41%	3.38%	3.23%	3.42%	3.42%	3.77%
Contribution of individual heterogeneity	49.85%	50.71%	51.50%	49.79%	49.90%	49.32%
— Unobserved individual heterogeneity	29.00%	29.22%	29.49%	28.94%	29.06%	28.58%
— Observed individual (gender, quasi ed.)	17.62%	18.16%	18.64%	17.62%	17.61%	17.61%
— Birth year	0.32%	0.38%	0.39%	0.33%	0.32%	0.23%
— Region	2.91%	2.96%	2.97%	2.90%	2.91%	2.90%
Contribution of firm heterogeneity	22.21%	21.49%	20.96%	22.25%	22.16%	22.43%
— Unobserved firm heterogeneity	15.69%	15.28%	14.98%	15.72%	15.59%	15.87%
— Observed firm heterogeneity (ownership)	4.14%	3.83%	3.59%	4.13%	4.15%	4.21%
— Sector	2.38%	2.37%	2.39%	2.40%	2.43%	2.35%
Contribution of occupations	7.93%	7.97%	8.01%	7.93%	7.97%	8.02%
Residual variation	14.61%	14.51%	14.38%	14.63%	14.57%	14.25%
Correlations						
Corr(θ_i,ψ_j))	0.175	0.150	0.130	0.175	0.175	0.170
Corr(ε_i^I,ε_j^J)	0.138	0.111	0.088	0.136	0.138	0.137
Corr(θ_i,ψ_j) for inc. firms	0.31	0.293	0.275	0.311	0.310	0.307
Corr(θ_i,VA_j) for inc. firms	0.364	0.369	0.376	0.364	0.362	0.361
Corr(ψ_j,VA_j) for inc. firms	0.614	0.606	0.600	0.614	0.615	0.616
Between-within decomposition						
Between-firm share	47.7%	46.9%	46.0%	47.9%	47.7%	48.0%
— Ind. segregation	11.6%	11.5%	11.3%	11.2%	11.2%	11.2%
— $Var(\psi_j)$)	18.2%	18.0%	17.8%	18.5%	18.5%	18.7%
— Sorting	9.1%	9.0%	8.9%	9.2%	9.1%	9.1%
Number of Observations (1000)	66155	24809	11526	26462	6616	66354
Number of Firms (1000)	144	115	84	133	93	144
Number of Workers (1000)	2462	932	438	985	246	2468
Reporting sample	50%	20%	10%	20%	10%	50%
Estimation sample	Full	Sub	Sub	Full	Full	February

Notes: See Table 1. Column 1 is the main result of 1. Columns 2 and 3 are from AKM models re-estimated on randomly drawn 20% and 10% samples of the population of workers (without replacement). Columns 4 and 5 use the wage components estimated as in Column 1, but reported on random subsamples. Column 6 represents an AKM model estimated on data using a different sampling of monthly observations, using wage data from February, May, August and November instead of January, April, July and October.

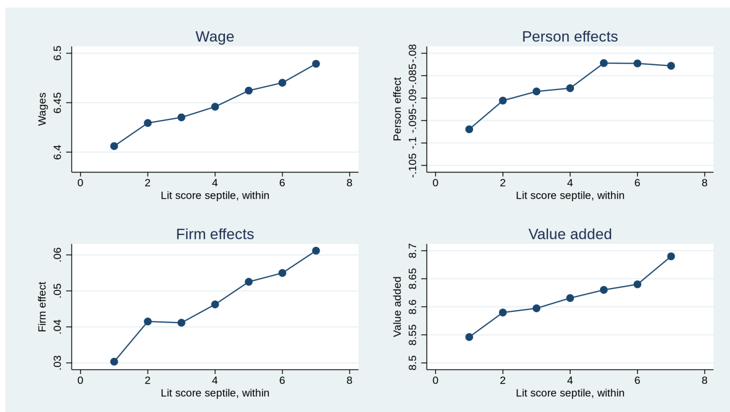
Table A5: Cross-sectional rent-sharing elasticity estimates

	(1) OLS Wi sector	(2) OLS Wi sector	(3) IV - Lag Wi sector	(4) IV - Lag Wi sector	(5) IV - Sales Wi sector	(6) IV - Sales Wi sector
Outcome:	lnW	ψ_j	lnW	ψ_j	lnW	ψ_j
Ln(VA/L)	0.346***	0.153***	0.391***	0.172***	0.401***	0.173***
	(0.010)	(0.005)	(0.011)	(0.006)	(0.013)	(0.007)
Firm-years	394,585	363,196	280,761	263,104	394,531	363,147
R-squared	0.618	0.525	0.455	0.320	0.444	0.316
Number of units	45	44	45	44	45	44

Notes: *** significant at the 0.1% level. Cluster-robust standard errors in parentheses. Outcome is the logarithm of value added per (reported) number of workers of the firm in the given year. Controls include firm size and fixed effects for 45 sectors defined as the interaction of 3 ownership and 15 industry categories. The instruments used are the one-year lags of productivity and the log sales per worker of the firms in the same year.



(a) Overall septiles



(b) Within-school septiles

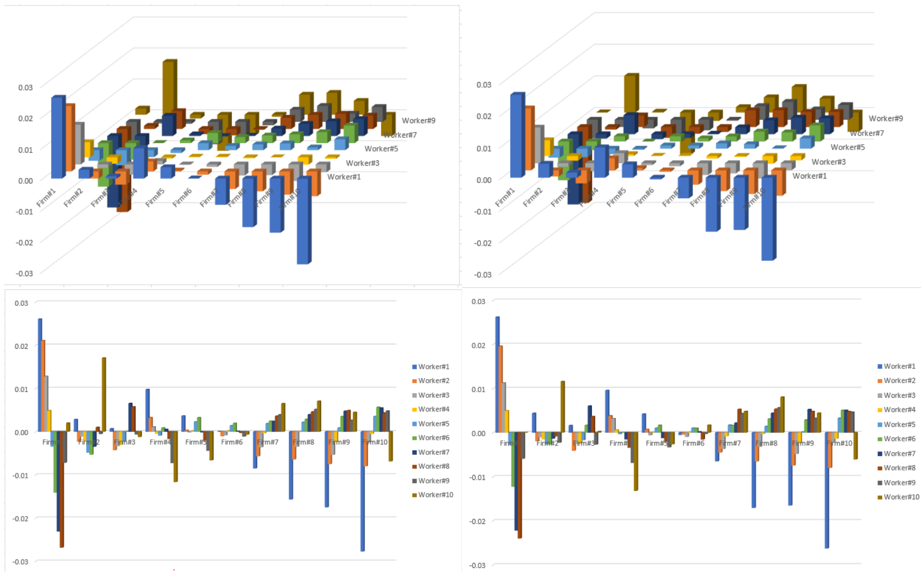
Figure A1: Wage components and value added along 10^{th} grade literacy score septiles from NABC

Notes: The seven quartiles are created along the distribution of literacy scores in year the students took the test (top panel), or within the distribution of the given school-year (bottom panel). The figures relate to those students for whom we have a test score observation no sooner than 2008 and also at least one wage observation anytime in the panel. The value added measure is available only for incorporated firms and not public institutions.

Table A6: Decomposition of wage variance, with match effects

Variance of log wages	0.338	
Ensemble decomp. (and sub-shares)		
Contribution of XB	**5.05%**	
— Year	*2.04%*	40.4%
— age*, firm size, contract, tenure*	*3.01%*	59.6%
Contribution of match heterogeneity	**5.06%**	
Contribution of individual heterogeneity	**51.79%**	
— Unobserved individual heterogeneity	*30.38%*	58.7%
— Observed individual (gender, quasi ed.)	*18.15%*	35.0%
— Birth year	*0.38%*	0.7%
— Region	*2.88%*	5.6%
Contribution of firm heterogeneity	**23.70%**	
— Unobserved firm heterogeneity	*16.68%*	70.4%
— Observed firm heterogeneity (ownership)	*4.55%*	19.2%
— Sector	*2.47%*	10.4%
Contribution of occupations	**4.50%**	
Residual variation	**9.90%**	
Correlations (and contr. to overall)		
Corr(θ_i,ψ_j)	0.176	10.0%
Corr(ε_i^I,ε_j^J)	0.131	4.5%
Corr(θ_i,ψ_j) for inc. firms	0.288	15.5%
Corr(θ_i,VA_j) for inc. firms	0.346	
Corr(ψ_j,VA_j) for inc. firms	0.608	
Between-within decomposition		
Between-firm share	47.3%	
— Ind. segregation	12.0%	
— Var(ψ_j)	19.0%	
— Sorting	10.0%	
Number of Observations (1000)	71914	
Number of Firms (1000)	161	
Number of Workers (1000)	2660	

Notes: See Table 1. The first stage is estimated with match and occupation fixed effects as in Equation 16, with match effects decomposed according to Equation 17, and the resulting firm and person effects decomposed according to 2 and 3.



(a) Match effects (b) Residuals

Figure A2: The estimated match effects and residuals along firm and worker effect deciles

Notes: The left panel presents the mean value of $\tilde{\omega}_{ij}$ from Equation 17 by cells defined along 10 deciles of estimated firm effects and 10 deciles of estimated person effects. The right panel contains the mean values of ε_{ijt} from Equation 1 along the same distribution.

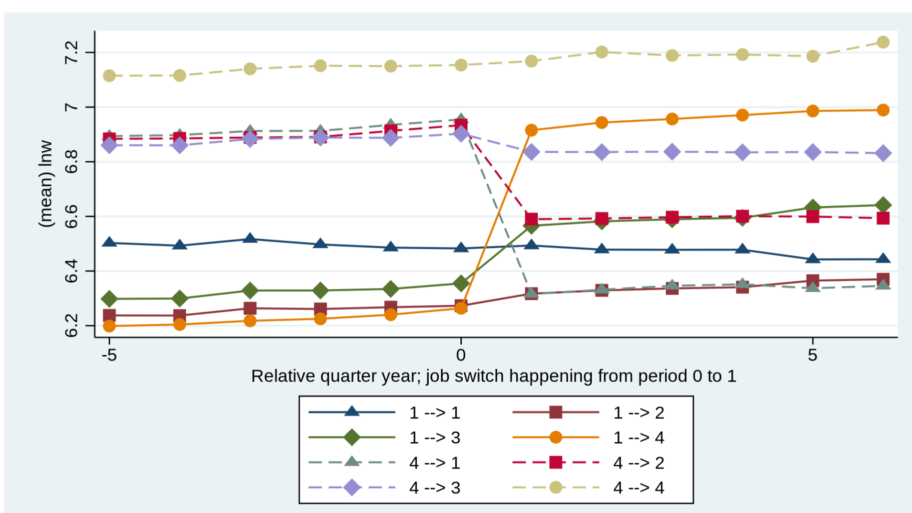


Figure A3: Event study of Card et al. (2013)

Notes: Data points represent mean log wages of job-switchers in the 18 months before, and 18 months following a job-to-job transition (on a quarterly basis), categorized by the firm effect quartile the worker belonged to before and after the switch. Only switches originating or arriving in the bottom or the top quartiles are included in the graph.

Table A7: Contribution of wage sorting, by sectors, regions and occupations

	N	M(w)	Var(w)	Corr.	Contr.
Occupation					
Managers	1692694	7.3	0.390	0.234	12.2
Professionals	5609251	7.0	0.305	0.243	14.1
Technicians and associate professionals	5520028	6.7	0.295	0.292	16.4
Office and management occupations	1930135	6.6	0.221	0.190	11.2
Commercial and services occupations	3176894	6.3	0.116	0.046	2.9
Agricultural and forestry occupations	223064	6.3	0.119	-0.007	-0.4
Industry and construction occupations	3540587	6.5	0.206	0.232	12.7
Machine operators, assembly workers, drivers	4655858	6.5	0.184	0.110	6.0
Elementary occ. requiring no qualification	4513468	6.2	0.109	0.061	4.3
Collapsed to occupation-years	70	6.6	0.120	0.107	2.6
Sector					
A -Agriculture, forestry and fishing	611855	6.5	0.182	-0.037	-1.8
D -Electricity, gas, steam and air conditioning	7317011	6.6	0.310	0.289	13.5
E -Water supply, sewerage, waste management	305221	7.1	0.345	0.198	7.2
F -Construction	492562	6.6	0.190	0.153	6.1
G -Wholesale and retail trade; repair of vehicles	873188	6.5	0.281	0.238	12.7
H -Transporting and storage	3092502	6.6	0.289	0.305	15.8
I -Accommodation and food service activities	2225119	6.6	0.246	0.220	9.7
J -Information and communication	679052	6.3	0.146	0.087	4.7
K -Financial and insurance activities	757030	7.2	0.396	0.136	7.1
L -Real estate activities	824052	7.2	0.389	0.172	6.7
M -Professional, scientific and technical activities	274265	6.5	0.259	0.184	10.3
N -Administrative and support service activities	670799	7.0	0.468	0.318	16.1
O -Public administration, defence, social security	1547679	6.4	0.224	0.224	12.5
Q -Human health and social work activities	174167	6.7	0.250	0.069	3.4
R -Arts, entertainment and recreation	417977	6.4	0.221	0.136	5.7
S -Other services, activities	175165	6.7	0.334	0.072	3.9
T -Activities of households as employers	240482	6.5	0.229	0.235	11.7
Collapsed to sector-years	148	6.7	0.084	0.716	31.2
Region					
Budapest	4891390	6.8	0.425	0.277	14.0
Central Hungary	3873678	6.7	0.362	0.236	12.1
Central Transdanubia	3929306	6.6	0.293	0.103	5.8
Western Transdanubia	3370532	6.6	0.289	0.111	6.1
Southern Transdanubia	2716328	6.5	0.287	0.125	6.4
Northern Hungary	3647330	6.6	0.280	0.098	5.3
Northern Great Plain	4442112	6.5	0.264	0.087	4.6
Southern Great Plain	3817505	6.5	0.264	0.077	4.2
Unknown	181013	6.4	0.245	0.152	8.2
Collapsed to region-years	63	6.6	0.024	0.557	14.3
All categories	30869194	6.6	0.325	0.173	9.1

Notes: The table reports the size (# of observations), mean log wage, the variance of log wages in given occupations, sectors or regions, alongside the correlation of estimated firm and worker effects in these cells with their respective contribution to $Var(w)$ as well.

Table A8: Decomposition of wage variance, occupation-sector effects

Variance of log wages	0.336	
Ensemble decomp. (and sub-shares)		
Contribution of XB	**5.05%**	
— Year	*1.92%*	38.0%
— age*, firm size, contract, tenure*	*3.13%*	62.0%
Contribution of individual heterogeneity	**49.69%**	
— Unobserved individual heterogeneity	*28.80%*	58.0%
— Observed individual (gender, quasi ed.)	*17.62%*	35.5%
— Birth year	*0.35%*	0.7%
— Region	*2.90%*	5.8%
Contribution of firm heterogeneity	**22.03%**	
— Unobserved firm heterogeneity	*14.92%*	67.7%
— Observed firm heterogeneity (ownership)	*3.94%*	17.9%
— Sector	*3.17%*	14.4%
Contribution of sector-occupations	**8.69%**	
— Occupation	*8.85%*	101.9%
— Sector	*-0.72%*	-8.2%
— Unexplained	*0.55%*	6.4%
Residual variation	**14.54%**	
Correlations (and contr. to overall)		
Corr(θ_i,ψ_j)	0.225	11.0%
Corr(ε_i^I,ε_j^J)	0.141	4.3%
Corr(θ_i,ψ_j) for inc. firms	0.326	15.9%
Corr(θ_i,VA_j) for inc. firms	0.365	
Corr(ψ_j,VA_j) for inc. firms	0.626	
Between-within decomposition		
Between-firm share	46.4%	
— Ind. segregation	10.9%	
— Var(ψ_j)	15.9%	
— Sorting	11.0%	
Number of Observations (1000)	61358	
Number of Firms (1000)	102	
Number of Workers (1000)	2362	

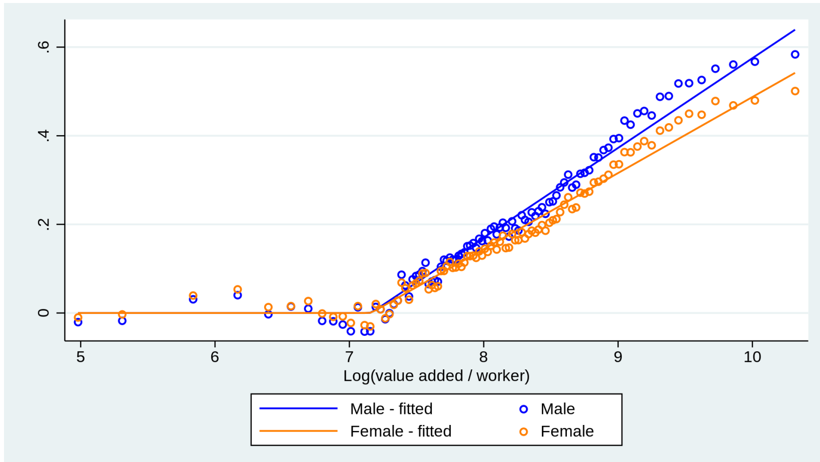
Notes: See Table 1. In this model occupation categories are interacted with firm industries to firm the third fixed effect of the first-stage estimation. These parameters are then decomposed into additive occupation and sector effects, with the residuals reflecting the importance of interaction terms.

Table A9: Ownership gaps decomposed, with sectoral controls.

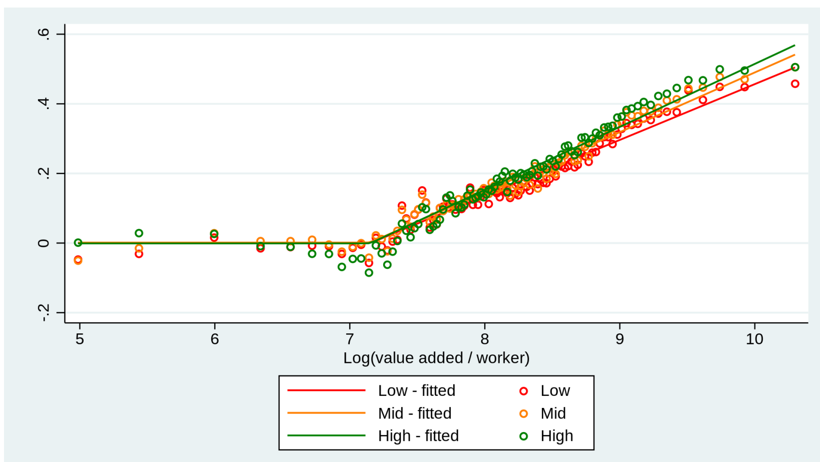
VARIABLES	(1) Wage	(2) Ind.	(3) Firm	(4) Occ.
Foreign-owned firm	0.353***	0.122***	0.223***	0.010***
	(0.020)	(0.010)	(0.011)	(0.002)
State-owned firm	-0.039	-0.011	-0.020	-0.006
	(0.041)	(0.017)	(0.023)	(0.006)
Public institution	0.000			0.000
	(0.000)			(0.000)
Observations	43,281,722	43,281,722	43,281,722	43,281,722
R-squared	0.178	0.092	0.326	0.088

VARIABLES	(1) Obs. ind	(2) Unobs. ind	(3) Within ind.	(4) Between ind.
Foreign-owned firm	0.049***	0.073***	0.131***	0.092***
	(0.005)	(0.006)	(0.006)	(0.009)
State-owned firm	-0.002	-0.008	0.014	-0.034
	(0.010)	(0.011)	(0.010)	(0.020)
Public institution	0.000			0.000
	(0.000)			(0.000)
Observations	43,281,722	43,281,722	43,281,722	43,281,722
R-squared	0.073	0.089	0.847	0.095

Notes: The parameters in the table are results from regression estimates of the effect of majority ownership dummies on wage components defined in Equations 11 (first panel) and 15 (bottom panel) as outcomes. The benchmark category consists of domestic, private-owned firms. The elements of X and Z are included as additional controls. Such variables are quadratic age, quadratic tenure, firm size, year, contract type and firm industry. Two-way clustered standard errors are in parentheses, with *** $p<0.001$, ** $p<0.01$, * $p<0.05$.



(a) Gender-firm effects (on dual connected set)



(b) Education-firm effects (on triple connected set)

Figure A4: Re-scaled group-firm effects, versus log value added per worker of firms

Notes: Data points are mean estimated firm-group fixed effects corresponding a hundred percentiles of firm-year observations along the distribution of the logarithm of value added per worker for firms with balance sheet data available. Firm-gender and firm-education effects are both normalized to have a zero mean value in all categories for observations below a log value added of 7.15 – the threshold that provided the best fit for the kinked function presented on the graphs.

Appendix B - Relation to Card et al. (2016)

In this Appendix we demonstrate, through the example of gender sorting and differences in bargaining, the relation of our approach to the framework of Card et al. (2016). While in the latter, the gender-based sorting is captured either by $E(\Psi_{j\text{Male}}|\text{Male}) - E(\Psi_{j\text{Male}}|\text{Female})$ or by $E(\Psi_{j\text{Female}}|\text{Male}) - E(\Psi_{j\text{Female}}|\text{Female})$, in our setting the corresponding term would be $\frac{\partial \tilde{\psi}_j}{\partial G} = E(\tilde{\psi}_j|\text{Male}) - E(\tilde{\psi}_j|\text{Female})$, which has to be between the two measures of Card et al. (2016). This follows from the fact that $\tilde{\psi}_j$ is actually a weighted average of female and male effects of the given firm. Specifically, let us consider the model from Equation 19:

$$\Psi_{jg} = G\tilde{\beta}_g + \tilde{\psi}_j + \varepsilon_{jg}^G \tag{22}$$

It can be shown, that in such a simple model with only firm fixed effects and G being a dummy for two gender categories, the following holds:

$$\tilde{\psi}_j = s_{Mj}(\Psi_{j\text{Male}} - \tilde{\beta}_g) + (1 - s_{Mj})\Psi_{j\text{Female}} \tag{23}$$

Where s_M is the share of male workers in the given firm, $\frac{N_M}{N_M + N_F}$, while $\Psi_{j\text{Male}}$ and $\Psi_{j\text{Female}}$ correspond to the firm effects for workers of the given gender at the firm. For simplicity, let us assume that s_M is constant across firms.[56] Then:

$$\begin{aligned} E(\tilde{\psi}_j|\text{Male}) - E(\tilde{\psi}_j|\text{Female}) = & \\ & (s_M E(\Psi_{j\text{M}}|\text{M}) - s_M\tilde{\beta}_g + (1 - s_M)E(\Psi_{j\text{F}}|\text{M})) - \\ & (s_M E(\Psi_{j\text{M}}|\text{F}) - s_M\tilde{\beta}_g + (1 - s_M)E(\Psi_{j\text{F}}|\text{F})) = \\ s_M(E(\Psi_{j\text{M}}|\text{M}) - E(\Psi_{j\text{M}}|\text{F})) + & (1 - s_M)(E(\Psi_{j\text{F}}|\text{M}) - E(\Psi_{j\text{F}}|\text{F})) \end{aligned} \tag{24}$$

That is, in this particular setting, our proposed estimator for sorting, $\frac{\partial \tilde{\psi}_j}{\partial G}$ will be the weighted average of the two alternative estimations Card et al. (2016) propose, with the weights s_M and s_F. It also follows that the estimator for bargaining will be also linear combination, as

$$\tilde{\beta}_g = \frac{\partial(\Psi_{jg} - \psi_j)}{\partial G} \tag{25}$$

To demonstrate that the above argumentation holds, and to assess how severe is the simplifying assumption of s_M being constant is, we replicate the estimators of Card et al. (2016) in Appendix Table B1.

[56] For the following arguments to hold without concern, it would be enough to assume that the expected value of s_M is the same for males and females and that it is independent of both $\tilde{\psi}_{j\text{Male}}$ and $\tilde{\psi}_{j\text{Female}}$. While any segregation by gender violates the former assumption, and the latter can be violated as well, these assumptions simplify the argumentation. Later, we will show that in our data this simplification has negligible importance.

Table B1: Relation to Card et al. (2016), dual-connected set

	Diff in Ψ_{jg}	Sorting	Bargaining	Sort. sh	Barg. sh
(1) Male distribution, female effects	0.1059	0.1004	0.0055	94.78%	5.22%
(2) Female distribution, male effects	0.1059	0.0789	0.0270	74.53%	25.47%
(3) Obs. distribution, firm mean effects	0.1059	0.0890	0.0169	84.03%	15.97%
$s_M(1)+(1-s_M)(2)$, with $s_M=0.4678$	0.1059	0.0890	0.0169	84.00%	16.00%
(3) with controls for XB	0.0628	0.0456	0.0172	72.59%	27.41%

Notes: first two rows are based on Card et al. (2016). The third row reports the decomposition of Equation 24. The fourth row is a weighted average of Rows 1 and 2, with the weight given by the in-sample average of the male-share variable in the sample. The estimation sample is restricted to firms for which both the male and female firm effects fall into their respective connected set. The estimation for the final row controls for age, tenure, calendar year and firm size.

The first two rows are replications of the estimators of Card et al. (2016), while the third row is the fixed effect approach proposed in this paper. As we can observe the decomposition in row 3 is indeed between the two previous estimates. Row 4 is defined by Equation 24, weighting rows 1 and 2 assuming a constant share of male workers across workplaces, $s_M = 46.78\%$, obtained as the mean of the within-firm share of male workers across the sample. The difference between rows 3 and 4 are of negligible magnitude, suggesting a small role of correlations between firm effects and gender ratios. We also note that the results of the first two rows are quite similar to the finding of Card et al. (2016), who present 6% or 31% share of the bargaining component, depending on the specification of choice. Therefore, our results also suggest an over-representation of male workers in firms with smaller gender gaps.[57]

An advantage of our approach, besides providing one estimator instead of two, is that it can be easily generalised to G variables of more than two categories. Also, we can easily incorporate the effect of X control variables, by estimating and subtracting $X\beta_X$ from the elements in Equation 19. The results, presented in row 5, suggests a 4.6% sorting parameter, which is comparable to the parameter in Table 6 (Panel A, Column 3, 3.0%). The source of the difference is either the slightly different sub-sample – this table refers to the dual-connected set of (integrated) firms – or, the differing assumptions of the AKM and G-AKM models, suggesting that assuming a common wage premium across firms is too restrictive compared to the model allowing for firm-specific gender gaps.

[57] Casarico and Lattanzio (2019) also reproduces the exercise of Card et al. (2016), and also report a weighted average of the male and female distribution based decompositions, with the weights of 0.5-0.5. They find around 70% importance of the sorting channel. The Online supplement to Lamadon et al. (2019) replicates these results as well, presenting almost identical sorting shares both with and without bias correction applied to the AKM estimations.

Appendix C - Variable definitions, estimation issues

Sample. Although we have monthly data, for computational convenience we use data from only every third month of the year, namely January, April, July and October.[58] We excluded partial months at the start or end of employment spells and used only months when workers were employed (insured) for all days in the given month, hence avoiding issues related to the imprecise measurement of wages in these months. We also excluded employers with less than 5 observed workers for two reasons. First, data from smaller firms is prone to be less reliable. Second, identification of the firm effects of small employers relies only on a small number of moves and thus estimations including them are more prone to limited mobility bias (Bonhomme et al., 2020). [59] We kept workers between the age of 17 and 65, as younger workers should be affected by compulsory schooling age, and by the age of 65 most Hungarians retire. We kept workers with standard contracted employment, including public servants and employees of public institutions (public workers) as well. Individual entrepreneurs, self-reliant farmers and other independent forms of employment are excluded.

Mobility. The connected set on which the estimated fixed effects are directly comparable has to be defined according to the algorithm of Weeks and Williams (1964), as noted by both Torres et al. (2018) and Gyetvai (2017). This three-way connected set for our main specification includes 91.9% of observations, 86.2% of firms, 92.1% of workers from the sample defined above. As our panel is only a 50% sample, limited mobility bias could not be neglected. However, we trust that having fifteen years of data in the same panel helps greatly in overcoming this issue. Furthermore, using quarterly data, we observe 60 time periods with within-year movements also contributing to the set of job switches used for identification of the firm effects.

Wages. Our wage variable is defined the following way. We calculated hourly wages by dividing monthly earnings by four times the reported weekly work hours. (If no value was reported, we imputed the most common value, 40 hours per week.) Then, within all calendar months wages were winsorized, that is values below the bottom and above the top percentile cut-offs were re-coded to the corresponding cut-off values. Finally, nominal wages were divided by a monthly consumer price index, and then taken the logarithm of.

Time-varying factors. Building upon the findings and specifications of Card et al. (2018) and Torres et al. (2018), we included in the main AKM estimation as time varying terms quadratic and cubic age terms, with the age profile assumed to be flat at the age of 40. We included tenure and quadratic tenure (measured in months) to capture within spell wage evolution and added dummies to control for calendar years, as even the baseline level of real wages may vary across subsamples. We also control for the (logarithmic) size of the

[58] Using February, May, August and November did not alter meaningfully the results of main estimations. The re-estimation of our main model using these months is included in Appendix Table A4.

[59] Song et al. (2019) also omit employer-year observations with fewer than 5 employees in the year. While our restriction is more strict, abandoning it did not affect results substantially.

firm. Finally, the type of contract is accounted by dummies, reflecting whether the individual has a private or a public contract of employment.

Time-invariant terms. Anonymous person identifiers are provided in the data. Occupational differences are captured by high-dimensional occupation categories, coming from the Hungarian equivalent of the ISCO occupation categorization system. The classification was substantially altered in 2008, resulting for different codes being used before and since 2011. To overcome this issue, we harmonised the two category sets by using clusters of codes in which all old categories has to correspond to exactly one of the codes in the new nomenclature. Using this crosswalk, we ended up with 332 occupation clusters/ categories. Finally, instead of the original firm identifiers, we assigned firms new ones if their ownership changed with regard to the majority of foreign or state capital in the firm, or if they changed their main reported sector of operation. This way, we allow firms to have different wage premiums during different ownership or management regimes. Therefore, ownership and sector will become truly time-invariant characteristics of firms defined this way.[60]

Firm characteristics. Time invariant firm characteristics are sector categories created from 2-digit codes of the Hungarian equivalent of the NACE system of industries, corresponding to 61 distinct categories, and dummies indicating the majority of ownership – with domestic private, foreign private, state owned firm and public institution being the possible employer categories.

Individual characteristics. Individual time-invariant characteristics in our models include gender, the year of birth capturing cohort effects and the residential districts that individuals lived in for the most years during the time span of our panel. (In the case of multiple modes, the latest residence was used.) Districts are Local Administrative Units (LAU-1), of which Hungary has a total of 175. Finally, dummies for low and high quasi-education categories are included. This education variable is implicitly inferred from the data, and corresponds to the highest educational requirement of the occupations we ever observe the given individual working in. Specifically, we define the low education category as those who only ever worked as machine operators, assembly workers, drivers or in other elementary occupations requiring no qualification (ISCO categories 8 and 9). The high category consists of those who worked at least once as a manager or as a professional in jobs, which require the autonomous application of higher educational degrees (ISCO categories 1 or 2). Everyone else forms the in-between, middling category. Appendix Table C1 comprises the distributions of key categorical variables on the largest connected sample used for the majority of estimations presented in the study.

Estimation. For estimating the AKM model we use the method of Correia (2017), implemented in *Stata* under the command *reghdfe*.

[60] In Torres et al. (2018), the authors argue that changes in these variables are not common or has no substantial effect in Portugal and treat these variables as time-invariant elements of the second-stage regressions, while in-fact some within-firm variation remains in their data. The (minor) drawback of our approach may be losing some efficiency of estimates with the addition of extra estimable firm unit parameters and the use of smaller units in cases, where similar effects would apply for the same firm even under different regimes.

Table C1: Descriptive statistics on categorical variables

Gender	Observations	Share	**Region of residence (mode)**	Observations	Share
Female	36,815,363	50.8%	Budapest	11,794,027	16.3%
Male	35,593,233	49.2%	Central Hungary	8,779,139	12.1%
Proxied education			Central Transdanubia	9,028,080	12.5%
Low ed.	8,700,436	12.0%	Western Transdanubia	7,949,449	11.0%
Mid ed.	39,054,621	54.0%	Southern Transdanubia	6,521,187	9.0 %
High ed.	24,634,902	34.0%	Northern Hungary	8,583,571	11.9%
Age category			Northern Great Plain	10,279,085	14.2%
17-25 years	6,109,979	8.4%	Southern Great Plain	8,947,871	12.4%
26-40 years	28,893,529	39.9%	Unknown	526,187	0.7 %
41-55 years	30,129,289	41.6%	**Occupation category**		
56-65 years	7,275,799	10.1%	Political/religional/ngo leader	516,957	0.7 %
Tenure category			Top manager	588,628	0.8 %
¡= 12 months	13,045,516	27.0%	Other manager	3,739,393	5.2 %
12-35 months	13,251,047	27.4%	Professional	11,732,497	16.2%
36-60 months	6,768,405	14.0%	Other white collar	17,787,811	24.6%
60+ months	15,328,702	31.7%	Skilled blue collar	18,626,922	25.7%
Ownership type			Assembler, machine op.	9,759,379	13.5%
Domestic, private	26,073,563	36.0%	Unskilled laborer	9,296,730	12.8%
Foreign owned	16,522,151	22.8%	Unknown	360,279	0.5 %
State owned	5,789,831	8.0%	**Year**		
Public inst.	24,023,051	33.2%	2003-2005	14,336,394	19.8%
Employment type			2006-2008	14,695,375	20.3%
Standard contract	56,872,100	78.5%	2009-2011	14,309,224	19.8%
Public servant	3,301,429	4.6%	2012-2014	14,104,661	19.5%
Public worker	12,235,067	16.9%	2015-2017	14,962,942	20.7%

Notes: The distributions refer to the connected sample of the main estimations in Table 1.