



Department of Economics Discussion Paper Series

The Butcher, the Brewer, or the Baker: The Role of Occupations in Explaining Wage Inequality

Shihang Hou and Luke Milsom

Number 939

Updated: January, 2022

First published: June, 2021

The Butcher, the Brewer, or the Baker: The Role of Occupations in Explaining Wage Inequality

Shihang Hou* Luke Milsom*

January 12, 2022

Abstract

Google does not pay its canteen chefs the same premium as it does its software engineers; there is heterogeneity in pay premia within a firm. To study the implications of this observation, we estimate a two-way, worker-job, fixed effect model that builds upon the canonical model by [Abowd et al. \[1999\]](#) by allowing for within-firm heterogeneity. Estimating our model on UK administrative data, we find that even if workers, and firms, were identical, 25% of current log-wage variance would remain, due solely to heterogeneity between occupations. We further document that worker heterogeneity accounts for a relatively small proportion of log-wage variance, and that between-firm pay heterogeneity is more important in higher-wage occupations and larger labour markets.

*Hou: University of Oxford, Department of Economics, shihang.hou@economics.ox.ac.uk. Milsom: University of Oxford, Department of Economics, luke.milsom@economics.ox.ac.uk. With thanks to Margaret Stevens, Ferdinand Rauch and Gabriel Ulyssea for their supervision on this project, Johannes Abeler, Abigail Adams-Prassl, Sarah Clifford, Barbara Petrongolo, Simon Quinn, and Martin Weidner for helpful discussions, Sanghamitra Mukherjee, Verena Wiedemann and participants of the Firms reading group for listening to earlier drafts, and Binta Zahra Diop, Joel Kariel, Vatsal Khandelwal, Hannah Zillesen, and the applied microeconomics group at Oxford for helpful comments and suggestions.

1 Introduction

Google does not pay its canteen chefs the same premium as it does its software engineers. As a result, the best engineers go to Google, but the best chefs work at Michelin star restaurants: workers match to jobs, not firms. This simple observation implies that the canonical model of wage decomposition pioneered by [Abowd et al. \[1999\]](#) (hereafter AKM) – which proposes a statistical model that explains log-wages as the sum of a set of worker fixed effects, firm fixed effects, and idiosyncratic errors – is misspecified because it implies that firms pay the same wage premium to all their workers¹.

In this paper, we argue that AKM models lead to the omission of an important dimension of inequality – inequality of wages across occupations. This omission leads to the overstatement of the degree of wage heterogeneity explained by individual differences, and the understatement of significant heterogeneity of earnings across occupations within firms. This implies that even if we could eliminate all wage dispersion between firms or all wage dispersion between individuals, there would still be substantial wage dispersion due to differences in occupations.

We extend the canonical framework to allow different firms to pay occupations different wage premia, by estimating a model with individual and *job* fixed effects, where jobs are defined as occupation-firm pairs. Our model embeds the AKM model as a particular case but is more flexible by allowing firms to pay different wage premia to workers in different occupations. To estimate such a model we use a two-stage approach suggested by [Bonhomme et al. \[2019\]](#) to overcome limited mobility bias ([Andrews et al. \[2012\]](#), [Jochmans and Weidner \[2019\]](#)), first clustering jobs, and secondly employing worker job-cluster fixed effect regressions.

Estimating worker-job wage variance decompositions using matched employee-employer data from the UK², we show that the share of overall log-wage variance due to individual

¹The line of research pioneered by [Card et al. \[2013\]](#) has largely emphasised the role played by firms in explaining wage inequality [[Song et al., 2019](#), [Card et al., 2018](#), [Bonhomme et al., 2019](#), [Alvarez et al., 2018](#)]. The point that the model used is misspecified is taken up by recent theoretical work by [Haanwinkel \[2020\]](#) and [Bloesch et al. \[2021\]](#), which both argue that workers who are in some sense more important to a firm are paid more within it.

²We use the Annual Survey of Hours and Earnings (ASHE) data in this paper. The UK offers a particularly interesting setting for several reasons. First, although significant prior work has studied UK wage

heterogeneity is about 18%. This is significantly lower than what is typically found in AKM regressions. [Card et al. \[2013, 2016\]](#), [Song et al. \[2019\]](#), [Alvarez et al. \[2018\]](#), [Bonhomme et al. \[2020\]](#) all find that between 50% and 60% of log wage variance is explained by worker differences. We show this reduction relative to worker-firm decompositions is to be expected theoretically if more productive workers match to more productive occupations within firms. If the true model has varying job wage premia within a firm, then estimating an AKM model would cause this varying occupation premia to become an omitted variable, which will be loaded onto the estimated worker FE and firm FE. This results in AKM estimates of log-wage variance decompositions to attribute job premia heterogeneity to worker heterogeneity.

Consequently, our estimates of the role of job heterogeneity are larger than is typically found in the AKM context for firm heterogeneity. 38% of the log-wage variance can be explained by job heterogeneity, and 29% can be explained by positive assortative matching between workers and jobs. Both of these components are considerably larger than those found in firm-worker decompositions. In bias-corrected specifications, [Bonhomme et al. \[2020\]](#) find that firm heterogeneity accounts for between 5% and 13% across various European countries and the US, and that sorting between workers and firms accounts for at most 20%. Thus, previous work somewhat puzzlingly finds that sorting is strong even when the associated differentiation between firms is relatively less significant. Our results resolve this puzzle by showing both considerable job heterogeneity and considerable positive assortative matching.

We attempt to reconcile our results with those of the AKM literature by decomposing the variance in job fixed effects and in sorting into that which can be attributed to between-occupation variance and within-occupation variance. Variation within occupation across firms captures heterogeneity in firm-level pay premia, and a weighted average of this variation is conceptually comparable to the variance of firm fixed effects estimated in a standard AKM decomposition. When considering the variation in job fixed effects, we find that the between-occupation component is twice as important as the within-occupation component and that they account for 25% and 13% of overall log wage inequality respectively. Therefore,

inequality [[Faggio et al., 2010](#), [Mueller et al., 2017](#), [Machin, 2011](#), [Lee et al., 2016](#), [Rice et al., 2006](#), [Schaefer and Singleton, 2020](#)] to the authors' knowledge this is the first time large-scale employee-employer data has been brought to bear on the question using AKM style wage decompositions. Second, the UK typifies many debates surrounding wage inequality in western economies with spatial and occupational wage inequality squarely positioned in everyday political and social discourse.

most variation is a between- rather than within-occupation phenomenon. We find similar effects when decomposing sorting with 20% explained by sorting between occupations and 8% by sorting within occupations. These results highlight the importance of occupations in explaining log wage inequality but do not completely discount the role of firms.

Armed with worker and job fixed effects, we test and reject the assumption in the canonical model that firms pay the same wage premia over occupations. We reject the null hypothesis at traditional levels of statistical significance, but also show that differences are of a magnitude that is economically significant. We then study the distribution of match effects between occupations and firms by computing the differences between our framework and a naive framework that uses linear worker, firm, and occupation fixed effects. We find that there are substantial match effects between occupations and firms; low-skill occupations have a negative match effects at most firms, while high-skill occupations have a positive match effect at the most productive firms. This result supports the results by [Haanwinckel \[2020\]](#) and [Bloesch et al. \[2021\]](#) that firms pay disproportionate large wage premia to certain, typically high-skill, workers, and disproportionately small wage premia to typically low-skill workers.

This result naturally points to the hypothesis that since wage premia are especially large for high-wage occupations, it is likely that occupational labour markets in those occupations ought to be characterised by more heterogeneity and sorting. To investigate this, we analyse the sources of wage variance in each occupation. Looking at differences across occupations we find that sorting between workers and firms is only a substantial component of wage variance within higher-wage occupations. In relatively lower-wage occupations, the sorting component is negligible. This evidence suggests that models which emphasise positive matching between workers and firms are more appropriate for characterising high-wage occupations. In lower-wage occupations where it appears workers and jobs are not so heterogeneous, search is likely to be less important.

Turning to heterogeneity across local labor markets and using a methodology first developed in [Dauth et al. \[2018\]](#), we show that high-paying jobs and workers co-locate in larger labour markets. This finding lends evidence to a large literature in spatial and urban economics which highlights agglomeration economies and/or sorting across space leading to a city wage premium (see [D’Costa and Overman \[2014\]](#), [Glaeser and Mare \[2001\]](#), and [Rice](#)

et al. [2006] for example). We also find that the variance of wages is increasing in local labour market size and sorting, especially within-occupation, explains an increasingly large proportion of this variation with size. In the smallest labour markets, almost none of the wage variance is explained by within-occupation sorting (between individuals and firms within an occupation) whereas in the largest labour markets over 10% is. Differences in pay between firms conditional on occupation and worker are only a feature of large labour markets. To our knowledge, this result is new to the literature.

Our work relates to the large and active literature exploring the factors underlying wage inequality using the two-way fixed effects approach. This literature was initiated by Abowd et al. [1999] and Abowd et al. [2002], using French data. More recently, Card et al. [2013] used this framework to argue that increasing wage inequality in recent years was driven by increasing dispersion in firm wage premia. Their result was replicated in the American context by Song et al. [2019]. Alvarez et al. [2018] find that the implementation of the minimum wage drove reductions in wage inequality in the Brazilian context. Recent econometric developments in addressing the limited mobility bias faced by two-way fixed effect regressions in relatively sparsely connected networks (Andrews et al. [2012], Jochmans and Weidner [2019], Kline et al. [2020], Bonhomme et al. [2019] and Bonhomme et al. [2020]) have suggested that the true share of log-wage variance explained by firm effects is potentially much smaller than the 20-25% found in the earlier literature. Dauth et al. [2018] use the AKM framework to study geographical income inequality in the German context.

We contribute to this literature by emphasising that if we allow firms to pay workers in different occupations varying premia, as suggested by Haanwinckel [2020] and Bloesch et al. [2021], we obtain qualitatively very different results from the AKM approach. In particular, while results from most AKM regressions find that wage inequality is largely due to individual heterogeneity, we find that this proportion is substantially smaller when we consider between-occupation moves. Instead, a lot of wage inequality is due to disparate pay between occupations, and this is particularly important in explaining wage inequality in small geographical labour markets. While there is less within-occupation between-firm variation in pay for smaller labour markets and low-wage occupations, there are large match effects between high-wage occupations and high-premium firms, and thus within-occupation

variation is more important for higher-paying occupations. Partly due to these occupations typically being located in larger labour markets, and partly due to a market size effect, we also find that within-occupation heterogeneity between firms only matters substantially in larger geographical labor markets.

The rest of our paper proceeds as follows. Section 2 describes the specification and identifying assumptions underlying our model. Section 3 discusses the UK data that we use. We present our main results relating to the log wage variance decomposition in section 4. Section 5 considers differences across occupations. We discuss the implications of our model for wage inequality across space in section 6. Section 7 concludes.

2 Econometric model

In our econometric framework, we aim to disentangle log-wage variation into the components due to worker specific heterogeneity, job specific heterogeneity, and that due to the sorting of workers into jobs. For each individual $i \in \mathcal{I}$ in each period $t \in \mathcal{T}$ we observe the firm $F(i, t) \in \mathcal{F}$ and occupation $O(i, t) \in \mathcal{O}$ they work in as well as the wage w_{it} received. We denote by $J(i, t)$ the *job* an individual i is employed doing, in period t where j is a firm-occupation pair i.e. $J(i, t) \in \mathcal{J} = \mathcal{F} \times \mathcal{O}$. We assume that log wages $\ln(w_{it})$ can be described by the sum of worker, α_i , and job, $\psi_{J(i, t)}$, components³ as well as an idiosyncratic error as shown in equation 1.

$$\ln(w_{it}) = \alpha_i + \psi_{J(i, t)} + \varepsilon_{it} \quad (1)$$

This setup follows the literature started by Abowd et al. [1999] and continued by others such as Card et al. [2013] and Song et al. [2019]. In this model person fixed effects α_i can be interpreted as a combination of skills and other factors that are awarded equally across jobs j . Similarly, ψ_j can be interpreted as a pay premium offered to all workers in job j irrespective of unobserved worker-specific characteristics. Note that the usual worker-firm models (as opposed to worker-job) use ψ_f instead of ψ_j and interpret it as a pay premium offered to all workers at a specific firm. This means that, for example, software developers,

³Throughout we additionally condition on a set of time varying worker covariates X_{it} such as age, these have been omitted in this discussion for notational brevity.

administrators, managers, cleaners, and chefs at Google all receive the same log-additive pay premium.

Our model nests this as a special case where all occupations within a firm have the same ψ_j , but in general is significantly more flexible by allowing this premium to vary within a firm by occupation. For example, this would allow Google to pay its software developers a higher premium than its chefs and the Ritz to pay a higher premium to its chefs than its software developers. Given the nested structure we can test formally whether firms do indeed offer the same wage premium across all occupations and do so in section 4.3 finding evidence of statistically and economically significant varying premia.

The fixed effects in the AKM model are identified by *movers* between firms. In the same way, the job fixed effects are identified by the same individual moving from one job to another and noting how their wages changed due to the move, and individual fixed effects were identified by observing workers' wages between jobs relative to other workers in those jobs. Denote by y the stacked vector of individual wages, $A = [a_1, \dots]$ the design matrix of individual indicators, $J = [b_1, \dots]$ the design matrix of job indicators and ε the stacked vector of error terms. Then we can write equation 1 as $y = A\alpha + J\psi + \varepsilon$ or more succinctly as $y = Z\beta + \varepsilon$ where $Z = [A, J]$ and $\beta = [\alpha, \psi]$. We will estimate equation 1 by OLS to find $\hat{\beta} = \beta + (Z'Z)^{-1}Z'\varepsilon$. This result highlights the two main hurdles facing identification. First note that implicit in the above derivation is the invertibility of $Z'Z$. This is not the usual innocuous assumption; in general $Z'Z$ is not invertible in this set up if there are *unconnected* individuals or firms. That is, $Z'Z$ is only invertible in the subset of the data representing the largest connected set. In our data the largest connected set covers over 90% of observations and will be the sub-sample that we use for estimation (this is summarised in table 4).

The specification in equation 1 allows a simple decomposition of the observed variance of log wages as given in equation 2 below.

$$\begin{aligned} \mathbb{V}[\ln(w_{ijt})] = & \underbrace{\mathbb{V}[\alpha_i]}_{\text{Variance due to individual}} + \underbrace{\mathbb{V}[\psi_j]}_{\text{Variance due to job}} + \underbrace{2 \cdot \text{Cov}[\alpha_i, \psi_j]}_{\text{Variance due to workers}} + v \\ & \text{heterogeneity} \qquad \qquad \text{heterogeneity} \qquad \qquad \text{sorting into jobs} \end{aligned} \quad (2)$$

This shows how, for each period, wage variance can be decomposed into four components.

$\mathbb{V}[\alpha_i]$ captures the variance due to individual characteristics, and $\mathbb{V}[\psi_j]$ similarly captures the variance due to job characteristics. Finally, $2Cov(\alpha_i, \psi_j)$ captures the component of the variance which is due to sorting or matching of individuals to jobs, and v captures unexplained variance.

2.1 Limited mobility bias

AKM type decompositions suffer from limited mobility bias which was recently found to be qualitatively and quantitatively of importance (Bonhomme et al. [2020]). In our dataset for example, although we have almost 1.4 million observations, equation 1 requires estimating 303,000 worker fixed effects and 282,000 job fixed effects. However, the problem is significantly worse than it at first appears. This is because intuitively job fixed effects are only identified from *movers* between jobs — of which there are even fewer. While this doesn't lead to biased estimates of worker or job fixed effects, it does lead to biased second order (or higher) moments calculated from the resulting estimated fixed effects⁴.

In the main body of this paper, we follow the two-step approach suggested in Bonhomme et al. [2019] and Bonhomme et al. [2020] to overcome the limited mobility bias, where we first reduce the dimensions of the problem by classifying jobs into L job classes using a k-means algorithm⁵. This gives us a very connected network of workers and job classes, on which we can estimate a two-way linear regression on workers and job classes and related variance terms with little bias. In the first step, we cluster jobs into L classes using an unsupervised machine learning algorithm, k-means clustering based on 20 points (percentiles) of the job wage distribution. We find that these thirty clusters capture some 85% of the underlying variation in job wages; this is encouraging evidence of the success of the dimensionality reduction with minimal information loss⁶. In the second step we then estimate equations of the form given in 1 using job-class rather than job fixed effects.

⁴The econometrics of this bias are reviewed in the appendix in section A.4.

⁵K-means is an unsupervised machine learning algorithm that classifies data into K groups by minimising each within-group sum of square differences from the corresponding group mean.

⁶All results presented here are not sensitive to the number of clusters used. We have replicated very similar findings using 10, 20, 50, and 100 clusters.

2.2 Endogeneity and specification tests

To uncover unbiased estimates of the fixed effects, we require that $\mathbb{E}[Z'\varepsilon|Z] = 0$, the classic exogeneity assumption in OLS. The validity of this assumption for the classic AKM framework was discussed in detail in [Card et al. \[2013\]](#), and in our paper, we focus on what changes in our context relative to the standard AKM model. As in [Card et al. \[2013\]](#), we focus on whether the assignment of workers to jobs satisfies exogeneity, i.e. $\mathbb{E}[J'\varepsilon|Z] = 0$, and consider the sufficient condition that $\forall j, Pr(J(i, t) = j|\varepsilon) = Pr(J(i, t) = j)$.

The assumption made is not violated if the labour market is characterised by sorting on individual fixed effects and job fixed effects, i.e. if workers move to higher paying jobs or firms sort to more productive workers. Rather, there are two ways in which the assumption could be violated. First, the assumption is violated if workers sort to jobs on the basis of a worker-job match-specific characteristic not captured by α_i and $\psi_{J(i,t)}$. Relative to the AKM specification, this is more of a concern, as the skills necessary for different occupations are likely to be non-unidimensional (as would be implied in our model) (see e.g. [Lindenlaub \[2017\]](#) and [Lise and Postel-Vinay \[2020\]](#) for work arguing for the importance of multi-dimensional skills). Thus, there may be dimensions of individual skill not captured by the fixed effect that generate match-specific surplus at particular occupations which are pushed to the error term.

A second possible problem is that temporary variation in wages may be correlated to the job that workers perform. A concern of this type in [Card et al. \[2013\]](#) is that the statistical model is incompatible with models of the labour market where workers move to jobs due to high transitory wage offers not related to the firm fixed effect⁷. In our model, an additional problem of this kind is if temporary occupation-specific productivity shocks lead to substantial movement between occupations in the period.

An upshot of both concerns is that wage changes due to moves in one direction should be larger than wage changes due to moves in the other. If workers sort to jobs on the basis of elements not captured in either the worker fixed effect nor the firm fixed effect, we should expect the relative wage change of a worker moving from a low fixed effect job to a higher fixed effect job to exceed in magnitude the relative wage change of workers moving the other

⁷A prominent model with such a mechanism is [Postel-Vinay and Robin \[2002\]](#).

way. On the other hand, if the exogeneity assumptions holds, the average wage change of movers from a less productive job to a more productive job should be equal to the negative of movers from the more productive job to the less productive job.

We check for this endogeneity in appendix [A.2](#), using methods developed in Card et al. (2018) in the context of the AKM model. The main idea behind these tests is to check if wage changes in moves of workers between job clusters is symmetric. We find that wage changes between estimated job clusters are approximately symmetric.

Finally, another possible criticism of this approach is that the two-way fixed effect regression imposes an inappropriate log-additive functional form on firm and job fixed effects. The log-additive structure we impose on the data could mask important heterogeneity, mechanisms or disallow potentially relevant theoretical channels. However, we feel relatively assured that, even when taking a coarse job clustering approach⁸, most of the variation of interest is captured by our approach. This is due to two reasons. First, a relatively small part is played by unexplained variation in wage decompositions (roughly 8%). Second, we follow the approach in Card et al. (2013) and show that the mean of the residuals of the two-way FE regression are near zero on average, and by job and worker cells. This exercise and its results are summarised in appendix [A.3](#). We find that the specification performs well, except for very low-skill workers at very low wage premia jobs. This could be due to institutional factors like minimum wage that are imperfectly captured by this model. This finding is qualitatively similar to what other authors find when analysing the suitability of linearity in the context of more standard AKM models (e.g. [Card et al. \[2013\]](#), [Bonhomme et al. \[2019\]](#)).

3 Data and empirical setting

We apply the empirical specification discussed in section [2](#) to matched employee-employer data from the United Kingdom. We study the UK for two main reasons. First, to the best of our knowledge, this is one of the first studies (other than [Schaefer and Singleton \[2020\]](#)) to

⁸We check the robustness of the results to various numbers of clusters ($K = 30, 50, 100$) and clustering algorithms and find no difference.

bring employee-employer data and AKM-like decompositions to bear on the question of UK wage inequality. Second, the UK typifies contemporary debates surrounding wage inequality in aggregate as well as across space, common in developed economies.

Our main data source for this paper is the UK Annual Survey of Hours and Earnings (hereafter ASHE). The ASHE is a matched employee-employer data set collected annually by surveying employers and comprises of a 1% sample of the government tax agency’s job records. The survey, which is administered in April each year, records information about the identity of the establishment, its location and the industry classification of its primary business, the worker’s identity, sex, and home address, and details about the job itself, including earnings, hours, and occupational classifications. The survey is administered by the UK Office for National Statistics (ONS) and response to the ASHE is compulsory. The survey provides a panel data set as the same individuals are tracked over time, and as filling in the ASHE survey is a legal requirement of the employer attrition is very low. The ASHE can be compared to other data sets used in the AKM literature, such as the Integrated Employment Biographies data in Germany (used by [Card et al. \[2013\]](#)), the Master Earnings File in the USA used in [Song et al. \[2019\]](#), and a mix of microdata sets including the ANST in Sweden used in [Bonhomme et al. \[2019\]](#).

Details of our data cleaning process can be found in appendix [A.1](#). Table [1](#) below provides summary statistics on the entire data set and the largest-connected set used in our main clustered worker-job fixed effect specification. We lose data on about 50,000 workers and 15,000 firms in the largest connected set, or about 16% of all unique workers and 10% of all unique firms in the data set. This amounts to 3.5% of the sample in the second period in total. Reassuringly, the hourly wage moments in the two periods do not change significantly between the connected set and the full sample. Table [4](#) in the appendix summarises how much of our data is retained in the largest connected set for various specifications.

Table 1: Summary statistics

	02-10	11-19
All observations		
Total observations	1398098	1598919
No. unique workers	303,122	339,697
No. unique firms	127,014	156,178
No. unique firm-occ pairs	282,251	335,298
Average log hourly wage	2.21	2.18
SD log hourly wage	0.53	0.51
Largest connected set of main specification		
Total observations	1346163	1542383
No. unique workers	251,187	283,161
No. unique firms	112,817	139,533
No. unique firm-occ pairs	258,676	308,095
Average log hourly wage	2.22	2.19
SD log hourly wage	0.53	0.51

Notes: This table shows summary statistics from the ASHE data for two time periods: 2002-2010 and 2011-2019 as well as for the full data and the largest connected set.

4 Log wage decomposition

4.1 Overall wage variance decomposition

We present the main results of our decomposition in table 2 below, as well as several alternative specifications that have been produced in the literature⁹. In column (1), we produce estimates from the classic AKM two-way regression with worker and firm fixed effects. We find results that are qualitatively similar to those reported in other countries: a small and negative covariance between firm and worker fixed effects and 24% of log-wage variance due to firms' fixed effects, with most of the variance in log-wages being due to worker fixed effects. Column two shows the two-way fixed effect regression with worker and firm-cluster fixed effects, where firms are clustered into 30 firm clusters using a k-means algorithm on the basis of their wage distributions. In this specification, sorting becomes positive and increases at the expense of the contribution of firms and to a lesser extent workers. This again serves to highlight the importance of accounting for limited mobility bias.

⁹We compare these estimates against estimates in the literature in table 5 in appendix section A.6

Table 2: Comparison of different specifications (2011-2019)

Proportion of overall log-wage variance	Basic-firm	Cluster-firm	Cluster-Occ	Cluster-Job
Worker component	0.64	0.54	0.56	0.18
Firm component	0.24	0.10		
Occupation component			0.06	
Job component				0.38
Sorting component	-0.02	0.19	0.20	0.29
Residual component	0.08	0.10	0.11	0.08
Adjusted R2	0.89	0.88	0.87	0.90

Notes: This table presents log-wage variance decompositions for four specifications: (i) simple AKM, (ii) bias-corrected AKM, (iii) worker-occ twfe, and (iv) worker-job twfe (main specification). We use the ASHE data over the period 2011-19 for these results.

In column (3)¹⁰, we consider a specification where log earnings are decomposed into worker and occupation-specific effects. This regression serves two purposes. First, it is of interest in and of itself and shows a remarkable similarity to the cluster firm regression in column (2). A large share of the log-wage variance appears to be due to worker heterogeneity, with between-occupation heterogeneity accounting for only 6% of the total variance. Second, it serves to highlight that the differences between (low mobility adjusted) traditional decompositions in column (2) and those we suggest in column (4) are not just driven by the inclusion of occupation fixed effects but rather by specifically the inclusion of firm-occupation fixed effects.

Our preferred job-level analysis on the other hand in column (4) produces different results in several ways¹¹. There are three qualitative differences in these results relative to the results using the traditional AKM approach. First, the share of log-wage variance attributed to differences between workers falls to only about 18% of overall wage variance, from 64% in the standard AKM decomposition. As discussed in section 2, this result arises due to the misspecification of the wage process in traditional AKM models. Intuitively, where the job fixed effect is not included in the regression in column (2), it shows up as part of the omitted variable bias in the estimate of α_i in the misspecified model if there is residual covariance between α_i and $\psi_{J(i,t)}$ after accounting for the firm workers work at. This bias

¹⁰As there are only roughly 367 occupations, there is no limited mobility issue in individual-occupation fixed effect regressions. The results do not differ when clustering is used.

¹¹We check that the job-level estimates are related to firms' revenue in appendix A.8, suggesting that these estimates can plausibly be interpreted as job-level wage premia.

component is counted as part of the individual fixed effect in the variance decomposition of the misspecified model and must be larger than the actual variance of α_i if the covariance between the individual fixed effect and the residualised, weighted sum of job fixed effects across an individual’s career is positive. This explanation implies that, in practice, workers who work at high-wage occupations tend to work at high-wage occupations for most of their career, and vice versa¹². We discuss this analysis of the differences between our results and the canonical AKM results in appendix A.5, and show that in general, if there is positive sorting of workers to jobs conditional on firms, then in the misspecified model, the variance of worker fixed effects will be over estimated and that additionally the variance of firm fixed effects in worker-firm specifications will be incorrectly estimated (although the bias is of ambiguous sign).

These results paint a fundamentally different picture of wage inequality than those from the AKM regressions. In the AKM story, the substantial amounts of wage inequality are primarily because workers are very different, due to different degrees of ability or human capital accumulation. In our case, we find that when we make use of information from moves by workers across occupations, we see that differences between workers are much smaller than previously suggested. As a caveat, there is substantial heterogeneity in the importance of the individual worker differences in explaining wage dispersion between different occupations. As we discuss in section 4.3, wage dispersion in high-wage occupations is more readily explained by individual differences than in low-wage occupations. But, our results suggest that policies which reduce job-level differences in pay can be effective in reducing overall income inequality.

Second, we find that job heterogeneity and sorting both contribute significantly to wage inequality, specifically 38% and 29% respectively. Therefore, whereas previous specifications find low firm-level complementarities (and indeed we find low occupation level complementarities in column (3)) we find large job level complementarities in wages. We also find a larger impact of sorting of workers into jobs, which explains 29% of the variation, validating the original intuition that workers match to jobs, not firms.

A strange qualitative feature of results from AKM regressions, corrected for low-mobility bias, is that they have typically found low firm heterogeneity (10% in our sample) but a

¹²We provide some direct evidence of this in appendix A.7.

relatively larger share due to sorting between firms and workers, implying high degrees of sorting to firms when there is much smaller heterogeneity between firms. This feature is not present in our findings; we find that there is substantial job-level heterogeneity, and consequently, substantial degrees of sorting between jobs and workers. In the next section, we present evidence that suggests that a possible explanation of the differences between our specification and the standard AKM specification is that the latter approach only captures within-occupation job heterogeneity while we also capture heterogeneity between occupations.

4.2 Between and within-occupation log-wage decompositions

A natural question in our framework is whether heterogeneity in the job fixed effect is due primarily to differences between firms or between occupations. To answer this, we can further decompose the variation of job fixed effects, and the covariance between job and worker fixed effects, into between- and within- occupation components, using the law of total variance. The index of occupations partitions the data, and let $\mathbb{E}_o(\cdot)$, $\mathbb{V}_o(\cdot)$, and $Cov_o(\alpha_i, \cdot)$ denote the mean, variance and covariance with α_i of the random variable within the partition indexed by o . Equation 3 presents the decomposition of job variance into between and within occupation components, while equation 4 does the same for the decomposition of covariance.

$$\mathbb{V}[\psi_j] = \underbrace{\mathbb{V}[\mathbb{E}_o[\psi_j]]}_{\text{Between } o} + \underbrace{\mathbb{E}[\mathbb{V}_o[\psi_j]]}_{\text{Within } o} \quad (3)$$

$$Cov[\alpha_i, \psi_j] = \underbrace{Cov[\mathbb{E}_o[\alpha_i], \mathbb{E}_o[\psi_j]]}_{\text{Between } o} + \underbrace{\mathbb{E}[Cov_o[\alpha_i, \psi_j]]}_{\text{Within } o} \quad (4)$$

Results of this decomposition are reported in table 3. Table 3 also splits the sample into two periods and compares the decomposition across both. We find no significant differences in the drivers of inequality in the UK over time. This is in contrast to Card et al. [2013] who find, in Germany, that increases in firm heterogeneity and sorting explain changes in wage inequality.

We find that about two-thirds of the variation in job effects can be explained by between-

occupation variation and a third can be explained by within-occupation variation. The latter part, an occupation-weighted summary of wage variation between firms that is not explained by the occupation title of the worker, is about 13% in our data, which is close to the 10% estimated for the variance of firm fixed effect in the main AKM specification. This implies that the standard AKM approach misses most of the variability in job effects, which is due to differences in occupations across jobs rather than firm heterogeneity within occupations.

Similarly sorting between workers and jobs can be decomposed into within and between occupation components. The between occupation component is given by the covariance between mean worker type within an occupation and mean job fixed effect within that occupation. Whereas the within occupation component is given by the expected covariance between workers and jobs in a given occupation, across all occupations. We find that over two-thirds of the variation due to sorting can be explained by sorting between occupations, which reflects the sorting of higher-paid workers into higher-paying occupations, whereas only 9% of the log-wage variation can be explained by sorting of workers within occupations to higher-paying jobs.

Table 3: Variance decomposition from jobs model over time

Component of total log-wage variance	Period 2002-2010	Period 2011-2019
Variance of worker FE	0.20	0.18
Variance of job FE	0.34	0.38
Between-occ variation	0.23	0.25
Within-occ variation	0.11	0.13
Correlation between worker and job FE	0.32	0.29
Sorting between occupations	0.23	0.20
Sorting within occupations	0.09	0.08
Adjusted R2	0.90	0.90
RMSE	0.17	0.16

Notes: This table decomposes the log wage variance according to our job-worker two-way fixed effects model, for two period, 2002-10, and 2011-19. We further decompose the job variance and covariance components into between and within-occupation parts using the law of total variance.

Our results thus offer a different perspective on log wage inequality than most of the literature. Results from standard AKM regressions have tended to primarily emphasise the role of individual heterogeneity and secondarily the role of firms. In contrast, our results show that the occupational structure plays a large role in the story of wage inequality,

and most of the variance due to job heterogeneity and sorting of workers to jobs can be attributed to between-occupation heterogeneity and sorting of workers to occupations. Wage heterogeneity between firms within occupations is quantitatively less important, and we show in a subsequent section that it is relatively more important in high-wage occupations than in low-wage occupations.

4.3 Evidence on varying wage premia across occupations within firms

Armed with fixed effects estimates, we can turn to test the idea that motivated this paper: that firms offer different wage premia to workers in different occupations¹³. One reason for this may be that workers in some occupations are in some sense more central or important to a firm’s profit¹⁴. In this section, we present empirical evidence to suggest that firms offer varying wage premia across the occupations they employ. We take two approaches, first we analyse the estimated job fixed effects ψ_J and show that these vary within-firm across occupations. Second, we consider to what extent a naive additively linear specification in firm and occupation effects would be misspecified relative to our more general specification, which is equivalent to testing for occupation-firm specific match effects.

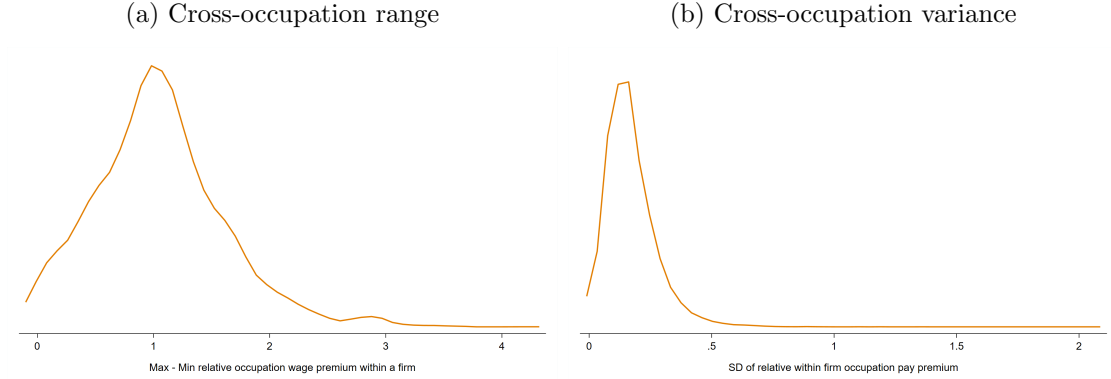
First, we plot the within-firm cross-occupation job fixed effect variance and range having removed aggregate occupation variation, in figure 1. This figure shows significant dispersion in wage premia across occupations within firms, even after accounting for aggregate occupation variation. In the left-hand side panel, the within-firm cross-occupation range is shown, and in the right-hand panel the within-firm, cross occupation variance. These statistics are calculated from the residuals of a regression of estimated job fixed effects on occupation and firm effects and therefore reflect varying wage premia within a firm accounting for aggregate

¹³In appendix A.8, we regress the estimated job fixed effects on firm characteristics to check whether the rent-sharing interpretation is appropriate and conclude that it is. We further check that the relation between the job FE and firm revenue is weaker for public firms than for private firms without collective bargaining agreements, and document heterogeneity across occupations. We leave a fuller analysis with firm profit data, which is sadly not available in the Business Structure Dataset (BSD) data set, for future work.

¹⁴In the model of Haanwinckel [2020], firms require tasks in different intensities for production, and workers in different occupations produce units of different tasks. In Bloesch et al. [2021], workers in different occupations can receive different wage premia because they have different hold-up power in the production of the firm.

occupational differences.

Figure 1: Distribution of the variance of job fixed effects across occupations within firms



Notes: This figure shows the within firm, cross occupation dispersion in wage premia removing aggregate cross-occupation effects. Both figures show density plots over firms, in the left-hand panel the cross occupation range (max-min) is displayed and in the right-hand panel, the cross occupation variance is displayed. The sample is restricted to firms with 10 or more employees. To calculate the statistics graphed estimated job fixed effects are residualised on firm and occupation effects.

Figure 1 shows significant variation, and therefore gives evidence to suggest that job wage premia do vary across occupations within firms. However, it is unclear from the figure alone whether this variation is statistically significant. Therefore, we formally test the differences graphically observed, by performing a test of the hypothesis that the variance of job fixed effects within the firm is no greater than five percent of the overall variance in job effects¹⁵. Performing this test we find that we can reject the null hypothesis for 49% of firms and therefore conclude that there is significant within-firm, cross-occupation, wage variance.

An alternative hypothesis is that while firms may offer an occupational premium to workers in different occupations, they also impose a constant firm premium on all workers within a firm. That is, we can approximate the job fixed effect estimated in our model by an additive specification given by equation 5, where $occ(i, t)$ denotes the occupation worker i is in at time t and $F(i, t)$ denotes the firm the worker is in at time t .

$$\psi_{J(i,t)} = \phi_{occ(i,t)} + \chi_{f(i,t)} \quad (5)$$

To test this interpretation, we estimate regressions of the form $\ln(w_{it}) = \alpha_i + \tau_{O(i,t)} +$

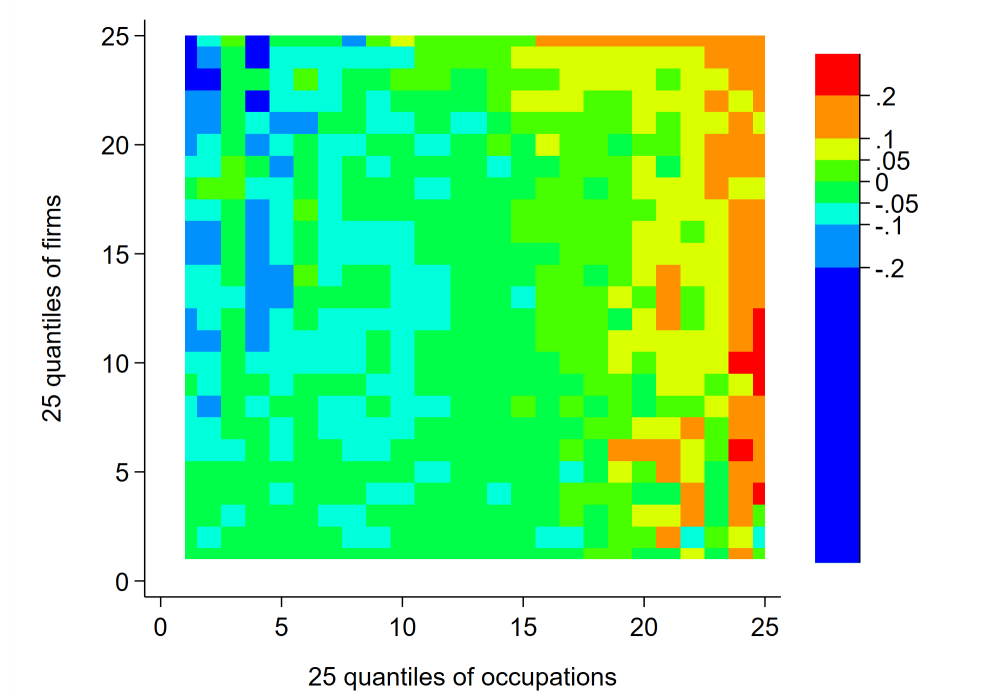
¹⁵To be precise we test whether within-firm, cross-occupation variation in job fixed effects is statistically significantly greater than 5% of the overall variance in job fixed effects across all firms and occupations using a Chi-Squared test [Snedecor and Cochran, 1989]. We find that we cannot reject the null hypothesis at the 5% level in 49% of firms with more than 10 employees in our data.

$\xi_{F(i,t)} + \varepsilon_{it}$, where τ_O are occupation fixed effects and ξ_F are firm fixed effects¹⁶. Then we can study the difference between these estimated fixed effects and our previously estimated job fixed effects $\psi_{J(i,t)}$. We interpret this difference as an occupation-job specific match effects on the offered wage premium. If firms do not pay differential wage premia across occupations, $\Delta = \psi_J - (\tau_O + \xi_F)$ will be small and not systematically related to any firm or occupation characteristics. A positive match effect of e.g. 0.1 implies that the premium offered by the particular job is roughly 10% greater than the premium implied by an additive occupation and firm fixed effect (and vice versa).

In figure 2, we produce a heat map of the difference between the job fixed effect in our main model and the sum of the occupation and firm fixed effects in the auxiliary model, that is of match effects. We plot this for 25 quantiles of occupations ranked by the average log hourly pay in the occupation and 25 quantiles of firms ranked by the average log hourly pay in the firm. Thus, a positive number (e.g. 0.05) can be interpreted as follows: if we naively replace the job fixed effect by the sum of the two fixed effects in our predicted values in our main model, we would underestimate the hourly wage by 5%. Green tiles represent quantile combinations where the error is less than 5% in either direction.

¹⁶This specification is similar to the one examined in [Torres et al. \[2018\]](#).

Figure 2: Contribution of worker-job match effects by firm and occupation mean pay quantiles



Notes: The figure shows a heatmap of the difference between the estimated job fixed effect in our main specification, and the sum of a firm fixed effect and an occupation fixed effect in an auxiliary specification, which we interpret as match effects between occupation and firms. Green squares imply that if the job fixed effect were replaced by the additive firm and occ fixed effects, the resulting predicted log wage would be within 5% of the true predicted log wage with job FE. Red implies that using the additive specification would underpredict the true wage by over 10% and blue implies that overprediction by over 10%. The x-axis indexes 25 quantiles of occupations, sorted by the mean wage within the occupation, and the y-axis indexes 25 quantiles of firms, sorted by the mean wage within the firm.

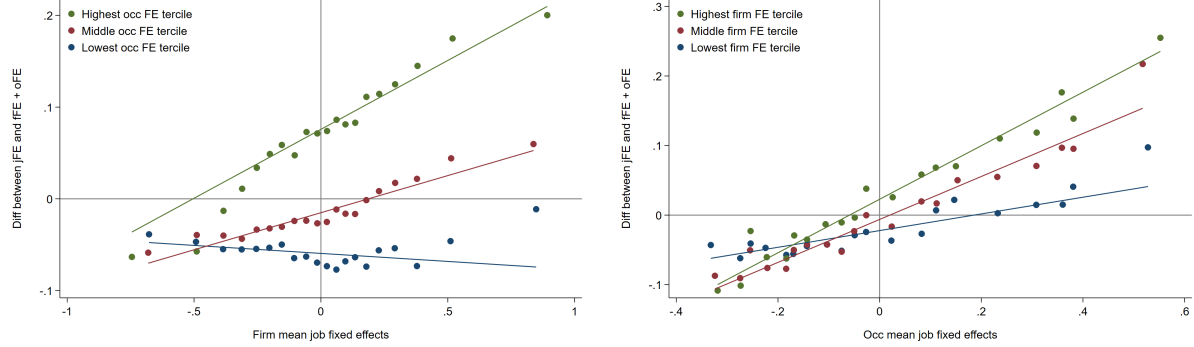
This figure shows that the linear form performs quite badly, especially at the extremes for each occupation. The linear combination would severely overestimate the wages of workers in low type occupations at all firms (as these have a negative match effect), and severely underestimate the wage of high type occupations at all types of firms. The problem seems especially acute for high wage occupations at low wage firms and low wage occupations at high wage firms, leading to mistaken predictions by over 20%. This leads us to conclude that the proportionate firm fixed effect is inappropriate for capturing the relation of pay premia to occupation and firm characteristics.

To analyse this dynamic further, we produce two complementary figures, first plotting the relation between the estimated residual against the mean log hourly pay at the firm for three terciles of occupations sorted by mean occupation wage, and the relation between the

estimated residual against the mean log hourly pay at the occupation for three terciles of firms sorted by mean firm wage.

Figure 3: Plotting estimated match effects

(a) Match effect by firm average job fixed effects for three occupation mean wage terciles (b) Match effect by occupation job fixed effects for three firm mean wage terciles



Notes: This figure plots the distribution of the difference between the estimated job fixed effect in our main model, and the sum of estimated firm and occupation fixed effects in an auxiliary regression, that is, match effects. Panel 3a plots the differences by firms, sorted by the mean job fixed effect within the firm for three terciles of occupations, while panel 3b plots the differences by occupation, sorted by mean occupation job fixed effect, for three terciles of firms. par

Figure 3 shows binned scatterplots of match effects which emphasise the importance of match effects for job-level pay premia¹⁷. Panel 3a shows that match effects increase from negative to positive as mean job fixed effects within firms increase. This relationship, however, is mediated by occupation terciles. Low-wage occupations have a negative match effect at all firms, regardless of the firm-level pay premium. On the other hand, high-wage occupations have a positive match effects at most firms, even those who offer on average below-average wages. Panel 3b shows the analogous results when plotting match effects over occupation average job fixed effects and cutting by firm occupation fixed effect terciles. We find that low occupation fixed effects occupations consistently have negative match effects and that although these increase with occupation average job fixed effects it does so faster for higher firm fixed effect firms.

Finally, we can formally test for the difference between the naive model: $\ln(w_{it}) = \alpha_i + \tau_{O(i,t)} + \xi_{F(i,t)} + \varepsilon_{it}$ and our more flexible specification $\ln(w_{it}) = \alpha_i + \psi_{J(i,t)} + \varepsilon_{it}$ using an F-test. This gives a F statistic of 3.93 and 3.24 when tested against the naive model with

¹⁷The scatterplots can be interpreted as similar to the data in figure 2, ‘sliced’ into a 2-D graph by reducing the dimensions of one of the axes.

and without including occupation fixed effects, in both cases the critical value at the 5% level is very close to 1, so we resoundingly reject equality between models.

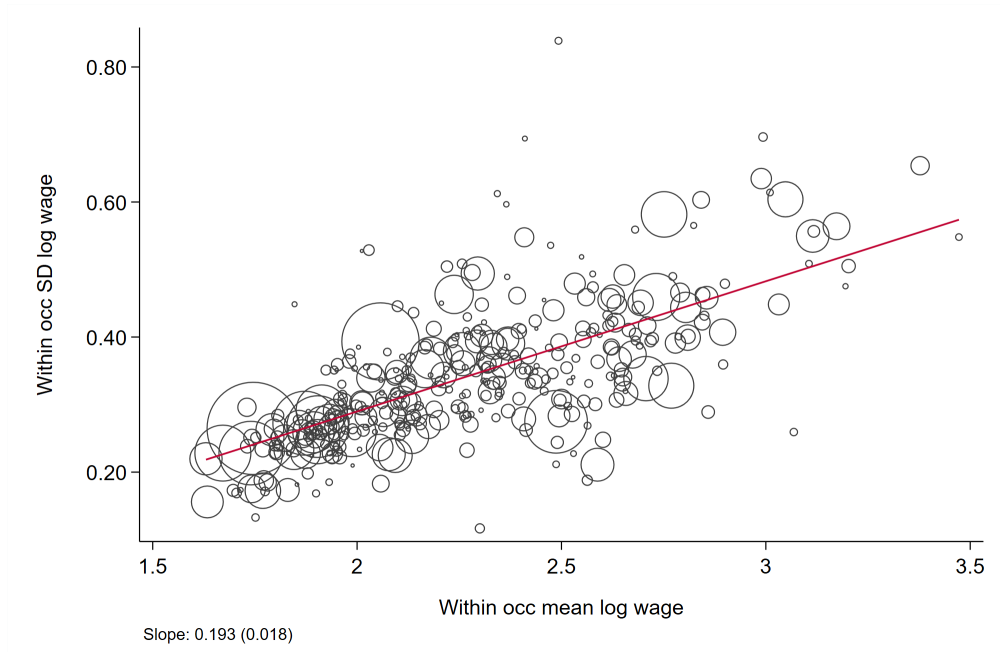
Our results support papers that argue that the standard AKM approach is misspecified, including [Haanwinckel \[2020\]](#), and [Bloesch et al. \[2021\]](#). [Haanwinckel \[2020\]](#) argues that firms should pay larger premia to workers who perform tasks which the firm use more intensively, and in a similar vein, [Bloesch et al. \[2021\]](#) show that workers which have more hold-up power within a firm should be paid more. Panel [3a](#) shows that these jobs are likely to be more highly paid and skilled occupations. Assuming a proportional firm wage premium leads to the systematic overestimation of the pay of low-wage occupations and the underestimation of the pay of high-wage occupations at high productivity firms.

5 Differences in the log-wage decomposition across occupations

An implication of the variation of wage premia across occupations within firms is that the sources of wage inequality within occupations may be quite heterogeneous. An advantage of our more flexible specification is that we can investigate the extent to which the overall wage decomposition presented is representative of the wage variance in individual occupations. Within an occupation, variation in the job fixed effect reflects differences in wage premia offered by firms, and we can more straightforwardly compute decompositions according to equation [2](#) within occupations. We find that the overall wage decomposition however disguises substantial heterogeneity in the wage decomposition in each occupation.

We start with the stylised fact that in the UK labour market, the variance of log hourly wages is co-increasing in the mean log hourly wage within an occupation. This is shown in figure [4](#), which plots a scatter plot of occupations with the mean log hourly wage on the x-axis and the standard deviation of log hourly wage within that occupation on the y-axis. The size of the bubble represents the number of workers in that occupation in the data. There is a clear approximately log-linear relationship between the mean and variance of log earnings, and dramatic outliers to this trend tend to be fairly small occupations.

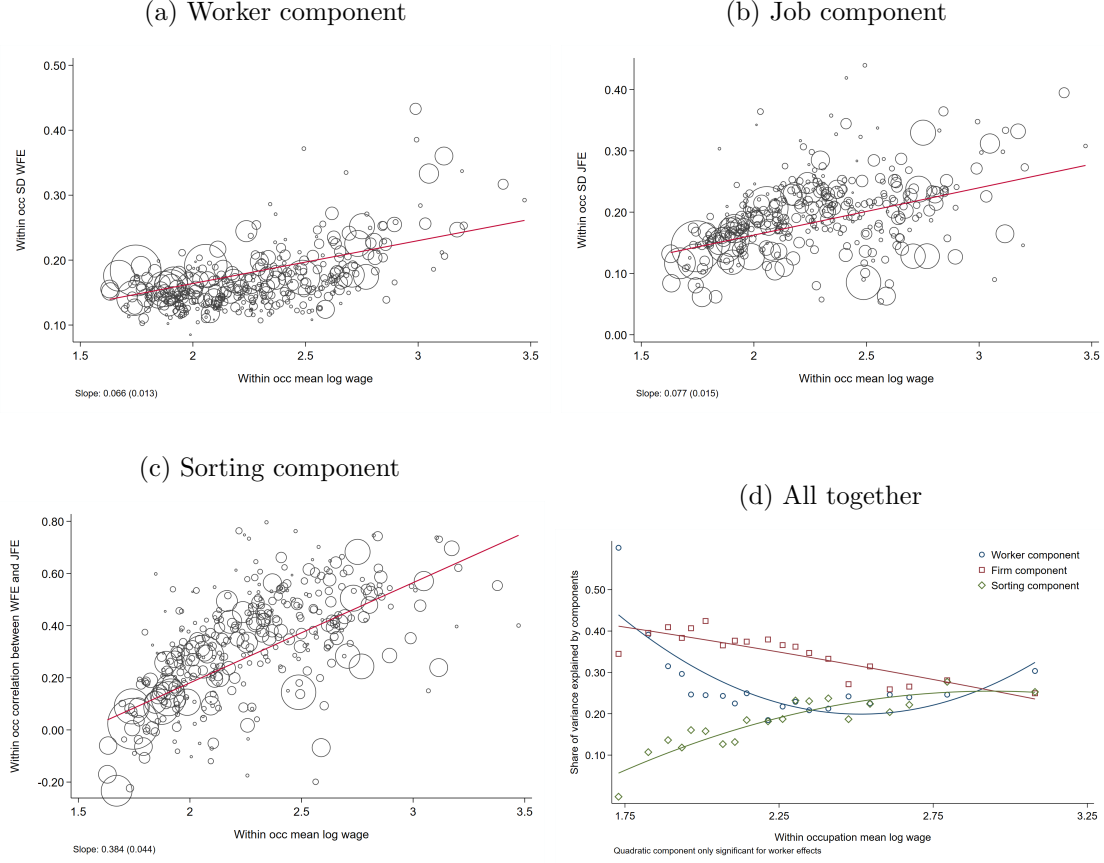
Figure 4: Within occupation mean vs SD of wages



Notes: This figure plots the mean log-wage (x-axis) and the standard deviation of log-wages (y-axis) within each occupation for all occupations in the data. Each bubble represents a particular occupation and the larger the bubble is, the more workers there are in the occupation. A red linear trend line is drawn, with slope 0.193.

What accounts for the greater wage variance in high-wage occupations relative to lower wage occupations? To understand this, we decompose the variance in each occupation into an individual component, a firm component and a sorting component. We plot the log-variance due to each component for each occupation in figure 5, panels (a) to (c). In panel 5d, we plot the variance of the components as a share of total log-wage variance within the occupation.

Figure 5: Variance of worker FE, job FE and correlation between the two components



Notes: Panel 5a plots the standard deviation of worker fixed effects in each of the 368 occupations in period 2 of the data (2011-19). Panel 5b plots the standard deviation of within-occupation job fixed effects within each occupation; since a job is identified as an occupation-firm pair, the variation is from firms offering different wages for the same occupation. Panel 5c plots the correlation between worker and firm fixed effects for each occupation. The occupations are ordered by the mean log-wage within the occupation on the x-axis.

The variances of all three components are increasing in the within-occupation mean wage, but at different rates. In particular, sorting is zero or even negative for many low-wage occupations and does not contribute to within-occupation wage variance for those occupations. However, it becomes as important as individual and firm heterogeneity for high-wage occupations. The variance of individual worker fixed effects also increases exponentially; it is relatively low for low and medium wage occupations but is quite high for the highest wage occupations. Thus, the drivers of wage variance appear to be quite different for low, medium and high wage occupations.

Our results suggest that different occupational labour markets require different modelling approaches based on the degree to which firms and individuals are heterogeneous. In

particular, the approaches emphasising firm-level wage heterogeneity and the role of matching (summarised in [Mortensen \[2003\]](#)) seem more applicable to higher-wage occupations, in which matching to a less productive firm or worker leads a larger loss of wages or profits. Lower-wage occupations exhibit almost no sorting and much smaller degrees of both worker and firm heterogeneity, and thus may be better approximated by classical or monopsonistic models of the labour market which do not emphasise matching. This conclusion resembles that reached by [Lise and Postel-Vinay \[2020\]](#), who arrive at it using a structural model of occupational mismatch with heterogeneous skills.

6 Understanding spatial wage variation through worker-job decompositions

An important dimension of wage inequality, in the UK¹⁸ but also more broadly in the world, is geographical inequality. It is well-documented that wage inequality is higher in larger labour markets (e.g. [Nord \[1980\]](#), [Baum-Snow and Pavan \[2013\]](#), and specifically in the UK context, [Lee et al. \[2016\]](#))¹⁹. Our novel²⁰ wage variance decomposition allows us to study inequality within and across geographies in the UK with a new lens, and better understand the salient features of inequality across space.

First, we find that higher type workers tend to co-locate with higher type jobs in larger labour markets. Figure 6 plots the mean worker fixed effect and mean job fixed effect within a travel-to-work area, where the size of the circle reflects the size of the local labour market. This finding supports a similar finding by [Dauth et al. \[2018\]](#) using a worker-firm two-way

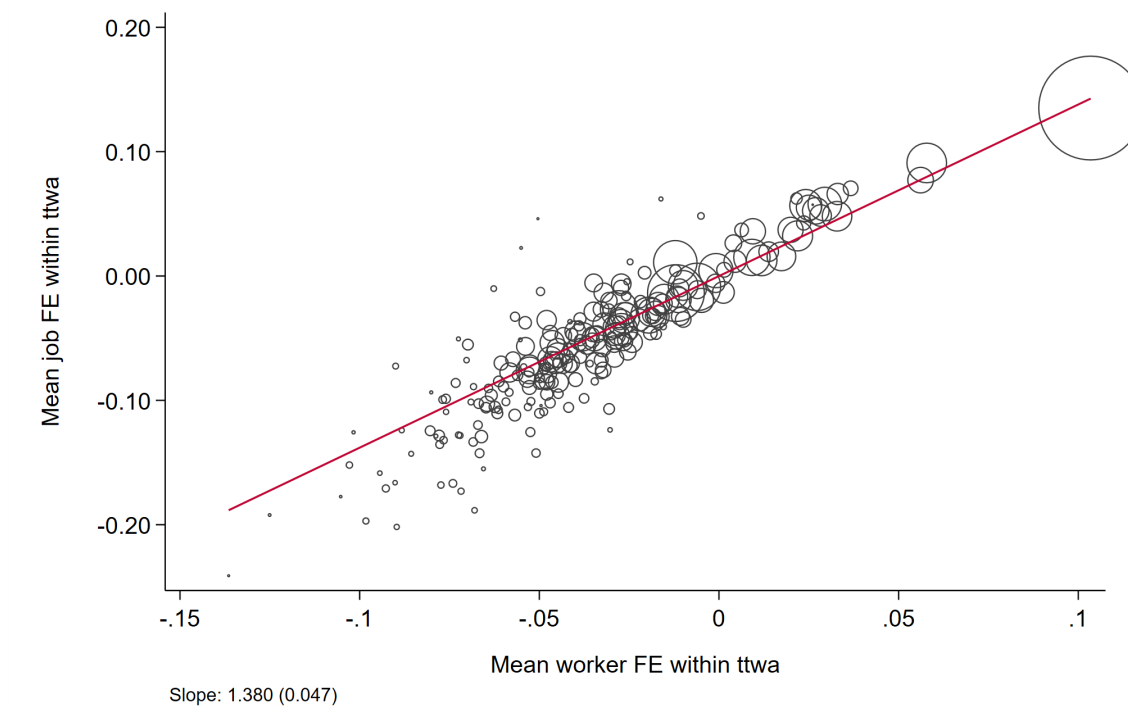
¹⁸[Agrawal and Phillips \[2020\]](#) show that nominal earnings gaps across space are large with incomes in Wales on average 40% below those in London. However, [Agrawal and Phillips \[2020\]](#) suggest that real differences are smaller and stress that within-city inequalities can be large, especially so in London. Spatial differences in income and opportunity are also increasingly dominating the political discourse in the UK, in September 2021 the department for Housing, Communities and Local Government became the Department for Levelling Up, and the new Secretary of State for Levelling Up, Michael Gove MP, called the Leveling up Agenda “the defining mission of this government”.

¹⁹We replicate this fact in the UK in appendix A.9, and show that the mean and variance of earnings is increasing in log travel-to-work area population. Our spatial unit is the UK travel-to-work area, as defined using data from the 2011 census.

²⁰For an analogous study that we build on based on using the canonical AKM model to study geographic income inequality, see [Dauth et al. \[2018\]](#).

fixed effect regression, and papers based on the observation that higher productivity firms sort to larger labour markets, like [Gaubert \[2018\]](#) and [Bilal \[2021\]](#). Analogously, it coheres with a large body of literature, including [Venables \[2011\]](#), [Davis and Dingel \[2019\]](#), and [Behrens et al. \[2014\]](#), which has found evidence to suggest and propose explanations for why more productive individuals sort to larger cities.

Figure 6: Within TTWA mean of worker FE and job FE (2011-19)

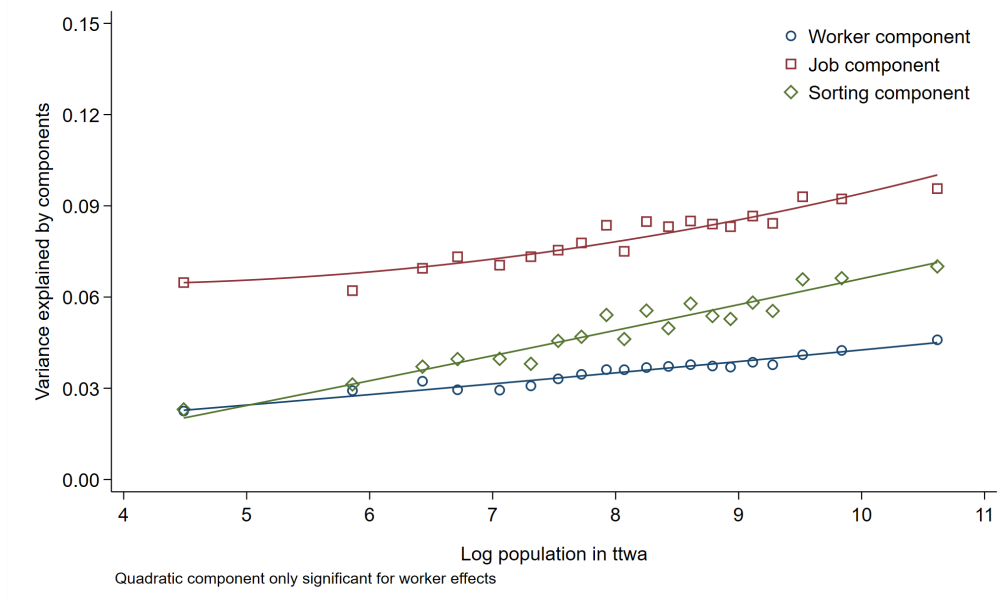


Notes: This figure plots the mean worker fixed effect (x-axis) and the mean job class fixed-effect (y-axis) within each travel to work area for each of 218 travel to work areas for 2011-19. Each bubble represents a particular travel-to-work area and the larger the bubble is, the more workers there are in the travel-to-work area observed in the data. A red linear trend line is drawn.

To study differences across labor markets in the constituent parts of within-labor market wage variation we apply our standard decomposition. Splitting log wage variance into a component due to the variance of individual fixed effects, a component due to the variance of job fixed effects, and a component due to the covariance between the two. We relate each of these components to (log) travel-to-work area size in figure 7 to explain the increasing within labor-market variation. All three constituent parts are increasing in local labour market size, but at different rates. It's clear from figure 7 that sorting becomes relatively more important in larger labor markets. This result coheres with a previous result in [Dauth](#)

et al. [2018] which finds that sorting to firms accounts for more variance in larger labour markets in Germany. This finding could be explained by the value of search being greater in larger cities due to scale effects, as was found in Petrongolo and Pissarides [2006].

Figure 7: Within TTWA variance components vs local LM size

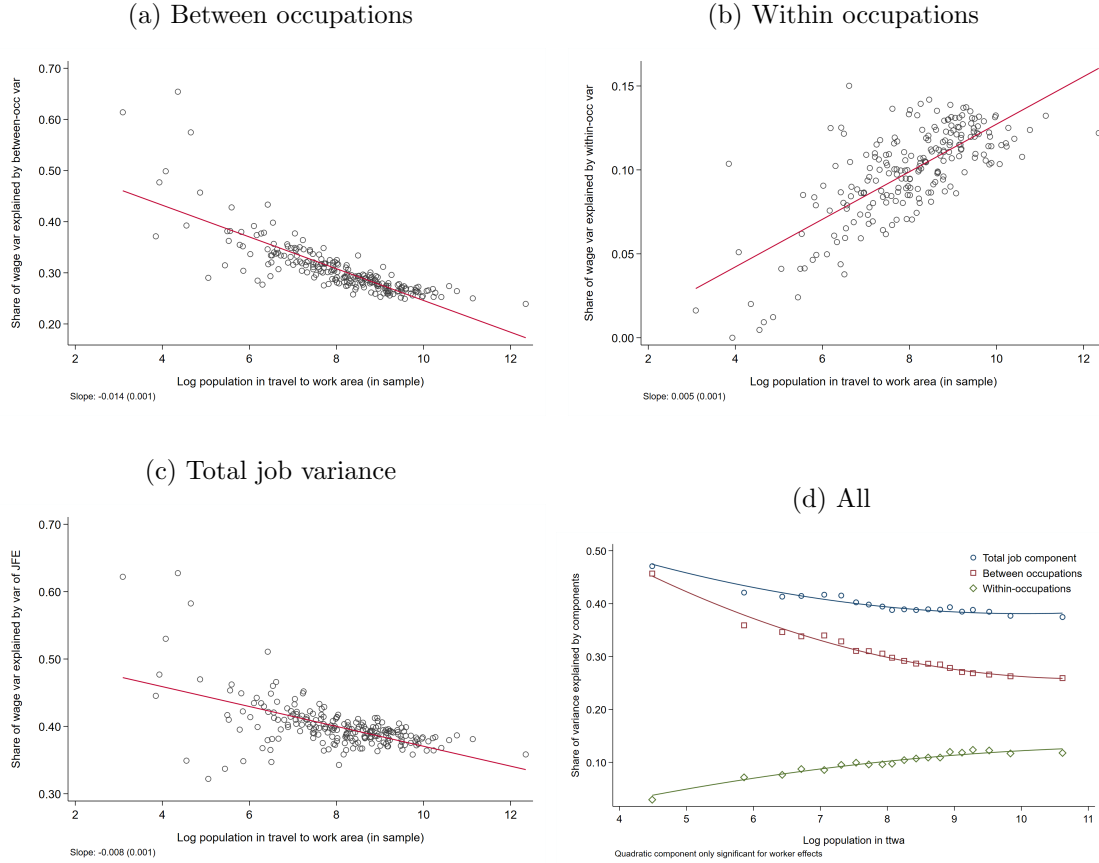


Notes: We compute the variance of individual fixed effects (blue line), the variance of job fixed effects (red line), and the covariance between the two (green) line for each travel-to-work area. We then plot a binned scatterplot against log-population within the ttwa and draw quadratic lines of best fit for all three components.

To delve deeper into what explains the patterns in figure 7, we decompose the local labor market job-class variance and covariance components into a between-occupations component and a within-occupations component using equations 3 and 4. First, we examine whether the decline in the relative importance of job heterogeneity is due to within or between occupations variance. We plot the relative importance of each factor in each TTWA in figure 8 below. Panels (a) and (b) plot the between- and within- occupation part of total log-wage variance within the TTWA as a function of log population in the TTWA. Panel (c) plots the total job variance as a share of total log-wage variance, and panel (d) combines all the other three panels in a binned scatterplot for clarity of comparison. The fall in the relative importance of job heterogeneity is primarily due to the falling relevance of between-occupation job variance as the local labour market gets larger. In larger local labour markets, wage dispersion between firms for the same occupation becomes relatively more important

from a low base, eventually accounting for about 10% of the job component of log-wage variance. Absolute, rather than relative, trends are plotted in figure 17 in appendix A.9.

Figure 8: Decomposing the within TTWA job variance component into within and between occupations

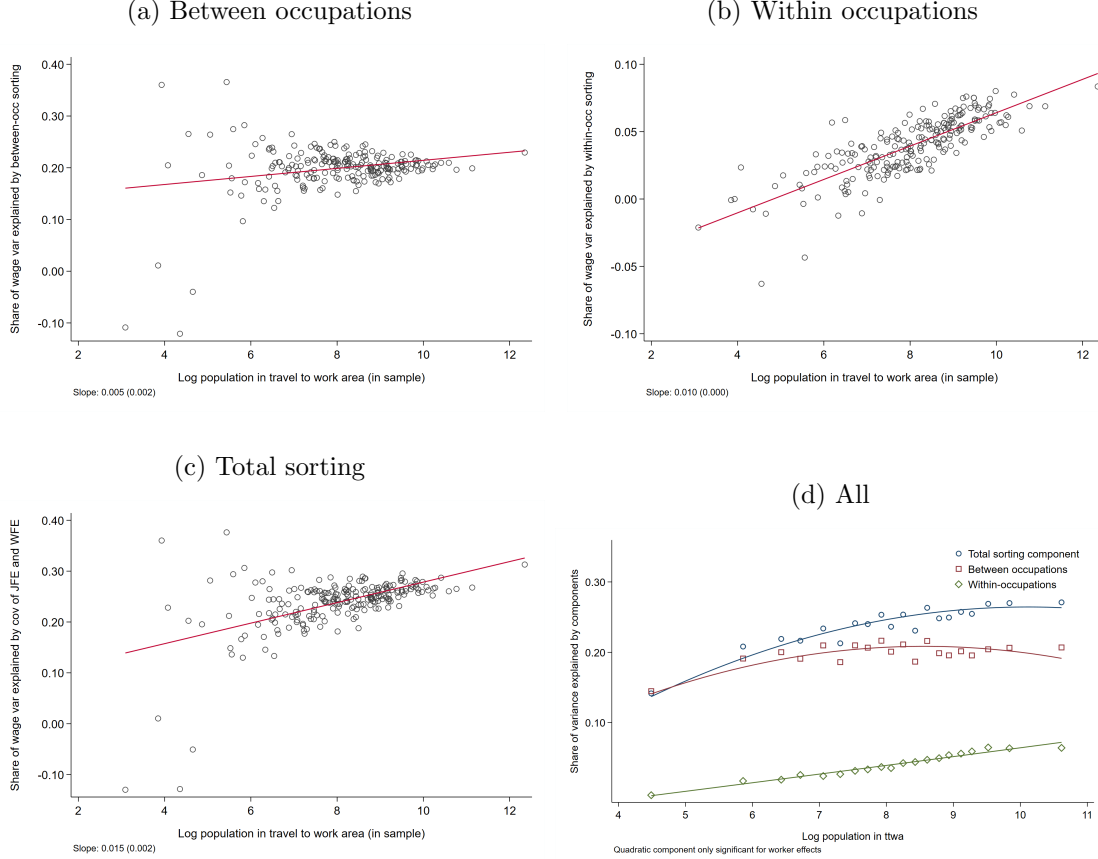


Notes: This figure presents scatterplots with log-population within the TTWA (x-axis) against between-occupation job-class variance in panel (a), within-occupation job-class variance in panel (b), total job-class variance in panel (c), and all components in bin-scatter plots in panel (d).

Similarly, figure 9 decomposes the share of within local labour market wage variance due to sorting between workers and jobs into the component between occupations and the component within occupations. (The between- and within- occupation covariances as raw values are included in figure 18 in appendix A.9.) What is immediately clear from this figure is that the majority of the previously documented increased importance of sorting in larger labour markets can be attributed to increased occupation sorting. That is, in larger labour markets better workers sort into better firms within occupation classes to a much greater

degree than they do in smaller labour markets²¹.

Figure 9: Decomposing the within TTWA wage variance sorting component into within and between occupations



Notes: This figure presents scatterplots with log-population within the TTWA (x-axis) against between-occupation sorting between worker and job-class in panel (a), within-occupation sorting between worker and job-class in panel (b), total worker and job-class sorting in panel (c), and all components in bin-scatter plots in panel (d).

In sum, although overall wage variation in the UK can mainly be described by a between-occupation story, within-occupation variation becomes a relevant feature of larger labor markets. Colloquially, wage inequality in villages and small towns is largely due to occupations, because of the differences in pay of the plumber and the shopkeeper rather than to differences between pay of plumbers at particular firms. As the local labour market gets larger, wage dispersion between firms within an occupation becomes a quantitatively more

²¹The increase in within-occupation sorting could reflect the tendency for jobs in high-wage occupations to locate in larger labour markets, given that we know that the degree of sorting is higher than in higher wage occupations. In appendix A.10, we show that while a share of this relation is explained by occupational composition effects, not all of it is explainable by composition.

important element of wage variance. The fact that better software engineers work at Google instead of smaller software firms is more relevant to wage inequality in London than in a provincial town.

7 Conclusion

In this paper, we propose that the classic two-way fixed effect model with worker and firm fixed effects is misspecified and that this misspecification produces a misleading decomposition of the components of log-wage variance. Instead of matching to firms, workers match to jobs which may differ within firms — e.g. a software engineer and a chef at Google are different positions. Further, it is likely that the firm would want to offer differing wage premia to workers in different occupations at the firm.

We estimate a two-way fixed effects model with individual and job fixed effects. We find that the component of variance due to individual heterogeneity, usually found to be 50-60%, accounts for only 18% of log wage inequality in our model. Conversely, we find that job heterogeneity and the sorting of workers into jobs account for 38% and 29% of the log-wage variance respectively. The majority of variation in these components can be explained by cross rather than within-occupation differences, highlighting the importance of occupations and jobs rather than firms, in explaining UK wage inequality.

Our approach also allows us to analyse how wage inequality within occupations differs across occupations. We find that wage variance in higher mean wage occupations is explained by sorting to a greater degree. We thus conclude that occupational labour markets should be analysed using different approaches, depending on the degree to which individual and firm heterogeneity affect productivity. Occupations in which heterogeneity is more important are better explained by search and matching models which produce more wage dispersion, while occupations, where workers and firms are more interchangeable, are better modelled by spot models of the labour market.

Finally, we consider the implications of analysis for understanding wage inequality in local labour markets. We find that sorting is increasingly important in larger labour markets, which could be due to scale effects in search markets. We also find that sorting to firms within

occupations are relatively much more important in larger labour markets than in smaller ones; while wage inequality in small towns is likely to be due to differences in occupation, wage inequality in large cities are as much due to workers being at different firms within the same occupation as being in differing occupations.

References

- John M Abowd, Francis Kramarz, and David N Margolis. High wage workers and high wage firms. *Econometrica*, 67(2):251–333, 1999.
- John M. Abowd, Robert H. Creecy, and Francis Kramarz. Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data. Longitudinal Employer-Household Dynamics Technical Papers 2002-06, Center for Economic Studies, U.S. Census Bureau, March 2002.
- Sarthak Agrawal and David Phillips. Catching up or falling behind? geographical inequalities in the uk and how they have changed in recent years. *The Institute for Fiscal Studies*, page p2, 2020.
- Jorge Alvarez, Felipe Benguria, Niklas Engbom, and Christian Moser. Firms and the decline in earnings inequality in brazil. *American Economic Journal: Macroeconomics*, 10(1): 149–89, 2018.
- Martyn J Andrews, Leonard Gill, Thorsten Schank, and Richard Upward. High wage workers match with high wage firms: Clear evidence of the effects of limited mobility bias. *Economics Letters*, 117(3):824–827, 2012.
- Nathaniel Baum-Snow and Ronni Pavan. Inequality and city size. *Review of Economics and Statistics*, 95(5):1535–1548, 2013.
- Kristian Behrens, Gilles Duranton, and Frédéric Robert-Nicoud. Productive cities: Sorting, selection, and agglomeration. *Journal of Political Economy*, 122(3):507–553, 2014.
- Adrien Bilal. The geography of unemployment. Technical report, National Bureau of Economic Research, 2021.
- Justin Bloesch, Birthe Larsen, and Bledi Taska. Which workers earn more at productive firms? position specific skills and individual worker hold-up power. 2021.
- Stéphane Bonhomme, Thibaut Lamadon, and Elena Manresa. A distributional framework for matched employer employee data. *Econometrica*, 87(3):699–739, 2019.

- Stéphane Bonhomme, Kerstin Holzheu, Thibaut Lamadon, Elena Manresa, Magne Mogstad, and Bradley Setzler. How much should we trust estimates of firm effects and worker sorting? Technical report, National Bureau of Economic Research, 2020.
- David Card, Jörg Heining, and Patrick Kline. Workplace heterogeneity and the rise of west german wage inequality. *The Quarterly journal of economics*, 128(3):967–1015, 2013.
- David Card, Ana Rute Cardoso, and Patrick Kline. Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *The Quarterly Journal of Economics*, 131(2):633–686, 2016.
- David Card, Ana Rute Cardoso, Joerg Heining, and Patrick Kline. Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics*, 36(S1):S13–S70, 2018.
- Wolfgang Dauth, Sebastian Findeisen, Enrico Moretti, and Jens Suedekum. Matching in cities. Technical report, National Bureau of Economic Research, 2018.
- Donald R Davis and Jonathan I Dingel. A spatial knowledge economy. *American Economic Review*, 109(1):153–70, 2019.
- Sabine D’Costa and Henry G Overman. The urban wage growth premium: Sorting or learning? *Regional Science and Urban Economics*, 48:168–179, 2014.
- Giulia Faggio, Kjell G Salvanes, and John Van Reenen. The evolution of inequality in productivity and wages: panel data evidence. *Industrial and Corporate Change*, 19(6): 1919–1951, 2010.
- Cecile Gaubert. Firm sorting and agglomeration. *American Economic Review*, 108(11): 3117–53, 2018.
- Edward L Glaeser and David C Mare. Cities and skills. *Journal of labor economics*, 19(2): 316–342, 2001.
- Daniel Haanwinckel. Supply, demand, institutions, and firms: A theory of labor market sorting and the wage distribution. *Unpublished manuscript*, 2020.

- Koen Jochmans and Martin Weidner. Fixed-effect regressions on network data. *Econometrica*, 87(5):1543–1560, 2019.
- Patrick Kline, Raffaele Saggio, and Mikkel Sølvsten. Leave-out estimation of variance components. *Econometrica*, 88(5):1859–1898, 2020.
- Neil Lee, Paul Sissons, and Katy Jones. The geography of wage inequality in british cities. *Regional Studies*, 50(10):1714–1727, 2016.
- Ilse Lindenlaub. Sorting multidimensional types: Theory and application. *The Review of Economic Studies*, 84(2):718–789, 2017.
- Jeremy Lise and Fabien Postel-Vinay. Multidimensional skills, sorting, and human capital accumulation. *American Economic Review*, 110(8):2328–76, 2020.
- Stephen Machin. Changes in uk wage inequality over the last forty years. *The Labour Market in Winter-The State of Working Britain*, 2011.
- Dale Mortensen. *Wage dispersion: why are similar workers paid differently?* MIT press, 2003.
- Holger M Mueller, Paige P Ouimet, and Elena Simintzi. Wage inequality and firm growth. *American Economic Review*, 107(5):379–83, 2017.
- Stephen Nord. Income inequality and city size: An examination of alternative hypotheses for large and small cities. *The Review of Economics and Statistics*, pages 502–508, 1980.
- Barbara Petrongolo and Christopher Pissarides. Scale effects in markets with search. *The Economic Journal*, 116(508):21–44, 2006.
- Fabien Postel-Vinay and Jean-Marc Robin. Equilibrium wage dispersion with worker and employer heterogeneity. *Econometrica*, 70(6):2295–2350, 2002.
- Patricia Rice, Anthony J Venables, and Eleonora Patacchini. Spatial determinants of productivity: analysis for the regions of great britain. *Regional science and urban economics*, 36(6):727–752, 2006.

Daniel Schaefer and Carl Singleton. Recent changes in british wage inequality: Evidence from large firms and occupations. *Scottish Journal of Political Economy*, 67(1):100–125, 2020.

George W Snedecor and William G Cochran. Statistical methods, eight edition. *Iowa state University press, Ames, Iowa*, 1191, 1989.

Jae Song, David J Price, Fatih Guvenen, Nicholas Bloom, and Till Von Wachter. Firing up inequality. *The Quarterly journal of economics*, 134(1):1–50, 2019.

Sónia Torres, Pedro Portugal, John T Addison, and Paulo Guimaraes. 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, 2018.

Anthony J Venables. Productivity in cities: self-selection and sorting. *Journal of Economic Geography*, 11(2):241–251, 2011.

Office for National Statistics. (2020). Annual Survey of Hours and Earnings, 1997-2020: Secure Access. [data collection]. 17th Edition. UK Data Service. SN: 6689, <http://doi.org/10.5255/UKDA-SN-6689-16>

Office for National Statistics. (2021). Business Structure Database, 1997-2021: Secure Access. [data collection]. 14th Edition. UK Data Service. SN: 6697, <http://doi.org/10.5255/UKDA-SN-6697-14>

A Appendix

A.1 Details on data cleaning

We focus our decomposition analysis on the most recent period from 2011-19, during which the UK standard occupation classification 2010 is used. We observe the most fine-grained 4-digit occupational classifications, giving us 368 occupations in total²². Our preferred wage

²²One of these, [2321], has fewer than 9 observations in the entire period, and is excluded for statistical disclosure purposes.

variable, used throughout as the focus of our analysis, is the hourly wage, which is computed as the gross pay divided by the total number of hours. We trim the top and bottom 1% of the sample to remove outliers. Our main unit of geographical analysis is the travel to work area (hereafter TTWA), a non-overlapping geographical unit provided by the Office for National Statistics. The TTWA is defined as an area where at least 75% of the resident workforce work in the area and 75% of the people who work in the area also live in the area and is thought of as a local labour market. TTWAs are the UK equivalent of commuter zones in the US which have become the standard unit for measuring local labour markets. TTWAs differ in size, and urban areas such as London can be very large. We use the 2011 classification of the TTWA in our analysis, giving us 218 travel-to-work areas in the data in total. Observations with incomplete information about their home address are used in the estimation of the fixed effects model but are not included in the descriptive analysis on geographical dimensions.

In the main decomposition analysis we can only use the largest connected set of workers and jobs to estimate fixed effects. Table 4 shows the overall size of the data set and the proportion thereof that is retained when considering the largest connected set over various decompositions reported in the main text. As one can see, the majority of the data is retained in most specifications with over 96% of observations used in our main decomposition reported in the final row.

Table 4: Largest connected set of common specifications (2011-19)

	No. obs	No. workers	No. firms	No. occs	No. jobs
All	1598919	339697	156178	368	335298
LCS for firm FE	93.57	81.01	63.57	100.00	79.30
LCS for clustered firm FE	96.46	83.36	89.34	100.00	91.89
LCS for job FE	88.50	77.98	59.14	100.00	56.93
LCS for clustered job FE	96.46	83.36	89.34	100.00	91.89

Notes: This table documents the proportion of the data retained in the largest connected set over various decompositions performed in the main text.

A.2 Testing the identifying assumption

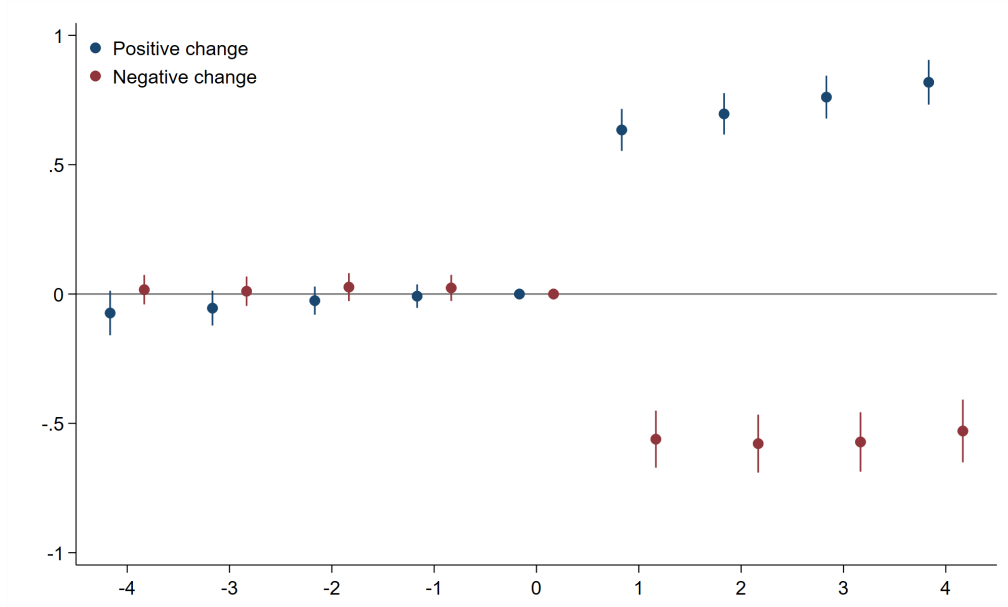
In section 2, we described the exogeneity assumption we need for our estimated fixed effects to not be biased, and listed some examples under which exogeneity would fail. In this

section, we perform a number of tests to show that this assumption is likely to hold. We follow analogous strategies in our context to the approaches that [Card et al. \[2018\]](#) takes to establish the suitability of the base AKM model.

The two-way fixed effect framework implies that when workers change jobs, they would experience a log wage change equal to the difference in the job fixed effect at the old job and at the new job. One implication of this is that for similar workers, the log wage change of moving from a former job to a latter job should be equal to the negative of log wage change of moving from that latter job to the former job. This may not be the case if for example job transitions occur due to workers matching to more suitable jobs over time. In that case, the mean wage changes associated with moves from less to more productive jobs should exceed that of moves in the other direction.

This idea is the basis of both tests presented in this section. First, we adopt the event study approach in [Card et al. \(2013\)](#) when we consider the wage effects of a worker moving from a less productive job to a more productive job and vice versa. [Figure 10](#) considers the wage rate in years relative to the year before a job switch occurs of workers who will (or have) switched jobs from the 1st to 4th (or 4th to 1st) quartile of jobs in terms of job fixed effects. This figure estimates wage trajectories using only within job switch pair variation and taking into account individual fixed effects. Wage trajectories are normalised to the year before the move (0'th year).

Figure 10: 1 to 4 (and 4 to 1) job switches



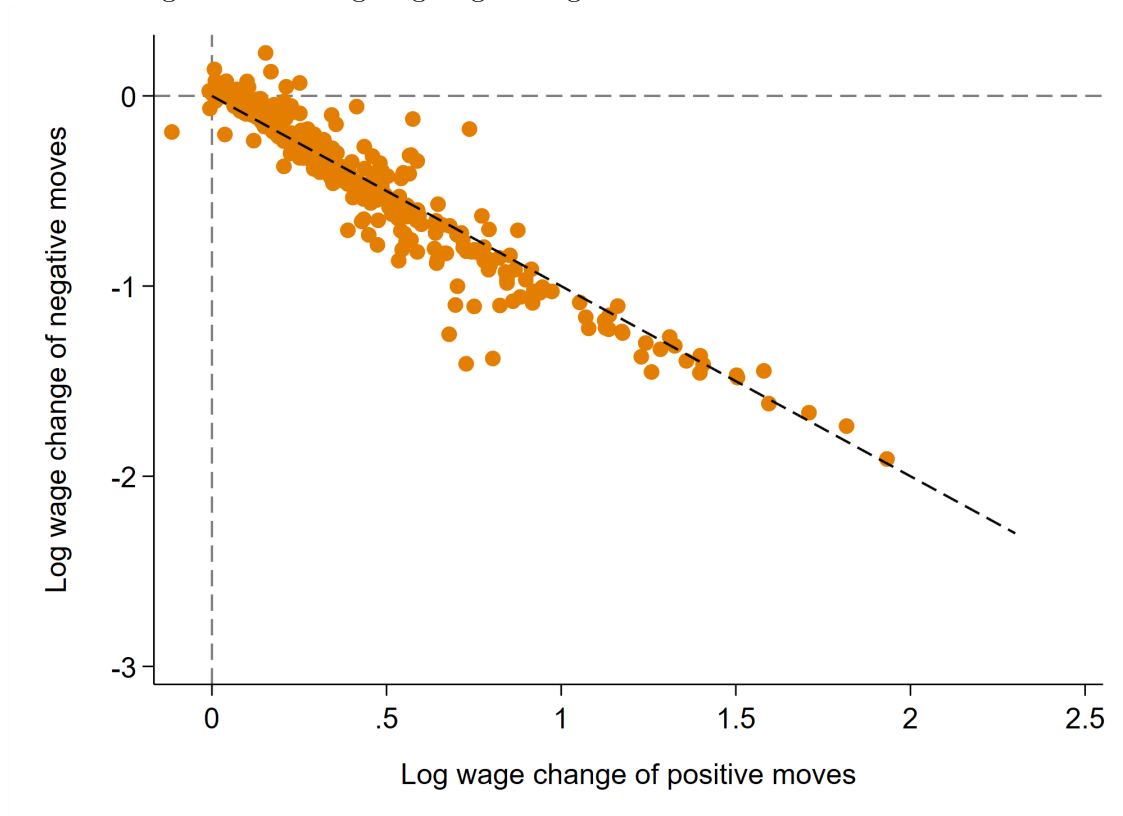
Notes: This figure shows the estimated log wage over time for workers who switch between jobs in the first and fourth quartile of the job fixed effect distribution. By including job pair fixed effects and individual fixed effects this figure shows the relative wages of workers who make a positive vs negative switch where wages are normalised to the year before the switch (0th year).

Two things are apparent from figure 10. First, there are no clear increases or decreases in pay in the year prior to, or proceeding, job changes, supporting what Card et al. [2018] describe as the step-like pattern of wage changes following job changes. Second, the degree to which wages change is symmetric; positive switches are associated with a similar increase in wages as negative switches are associated with a decrease. This gives evidence to suggest that there is not a large worker-job match-specific component of wages.

Second, the implied symmetry of wage changes in between job pair moves can be tested more directly by considering the average log wage change when workers move from a job to another. Because the set of moves between jobs is fairly sparse, we consider the average wage changes of moves between the wage clusters which we use to correct for limited mobility bias. We plot the changes in figure 11, restricting our analysis to firm-cluster pairs where we have observed at least 10 moves. If the exogeneity assumption is satisfied, then all the points should lie on the 45 degree line. We find that wage changes are roughly asymmetric as the model predicts, although there are a few outliers. The severity of the problem seems more acute for small moves, while larger moves between clusters seem largely symmetric as the model predicts. Many of the points also show that the wage loss associated with a

move to a lower productivity cluster exceeds the wage gain of moving to a more productive cluster, which similarly contradicts prominent narratives of between job transitions that do not cohere with the implications of the two-way fixed effect framework.

Figure 11: Average log wage changes of moves between firm clusters



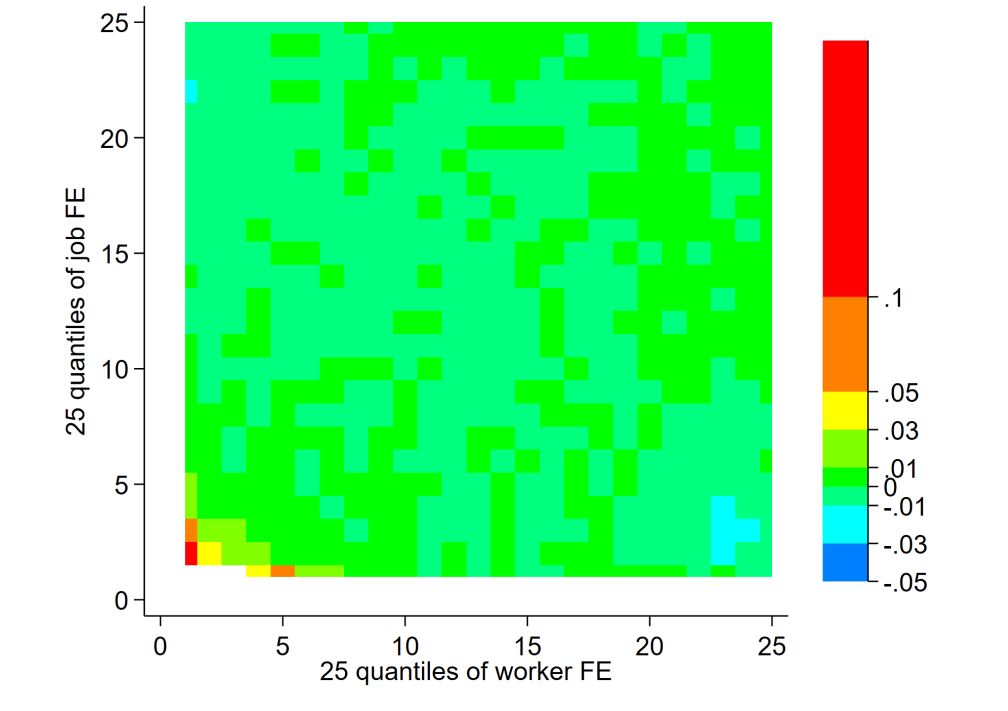
Notes: This figure plots the within-cluster average wage changes of moves between firm clusters. We use 30 firm-clusters which you use to correct for limited mobility bias. On the x-axis, we plot the wage change associated with moves from less productive to more productive clusters, while on the y-axis we plot the average wage change of moves from a more productive cluster to a less productive cluster. Each point represents a firm cluster pair, and there are 435 such pairs. We drop 94 such pairs where there are fewer than 10 moves in either direction. The dashed line is the 45 degree line which represents perfect symmetry.

A.3 Is the log-additive functional form a good specification for wages?

In this section, we check whether the log additive functional form in fixed effects is a good approximation to the data. To do this, we follow the test in [Card et al. \[2013\]](#), and aggregate our estimated fixed effects into 50 quantile groups. For each worker quantile and job quantile, we take the mean of the residual of the two-way fixed effect regression, reasoning that if the specification is a good fit to the data, then the mean of the residuals should be 0 or close

to zero. We plot a heat map in figure 12 with the quantiles of worker FE on the x-axis, the quantiles of job FE on the x-axis, and the size of the mean residuals as the colour within the cell. Warmer colours (yellow, orange, red) mean that the model underpredicts the true wage in the data while cooler colours (cyan, blue) means that the model overpredicts the actual wage data. Green implies that the mean residual in the cell is within 1% on each side of 0.

Figure 12: Mean of residuals by worker FE and job FE quantiles (25 groups each)



Notes: This figure plots the mean of the two-way fixed effect residuals for 25 worker FE quantiles by 25 job FE quantile cells for the two-way fixed effect regression with worker and job fixed effects without clustering.

The figure shows that for the vast majority of cells the difference between the predicted value and the true data is within 1% of the actual log wage. The model is particularly bad for a small number of low-skill workers at low-wage premium jobs, severely underestimating the pay of such workers by more than 10% and up to 30% at worst. This could be because of institutional factors, like minimum wage which are not explicitly considered in the model specification.

A.4 Review of limited mobility bias

Limited mobility bias occurs when second (or higher) order moments are estimated for fixed effects, which are estimated off few observations. In this case, the estimated fixed effects themselves are unbiased, but the high-order moments are biased. [Andrews et al. \[2012\]](#) pointed out this potential threat to identification and [Jochmans and Weidner \[2019\]](#) show how to understand the likely size of the bias as a function of the number of moves between workers and firms, or of the connectedness of the network of workers and firms. To see this more formally consider the variance of the fixed effects β , denoted as $V_Q = \beta'Q\beta$ for some matrix Q . Then our estimate of the variance would be given by $\hat{V}_Q = \hat{\beta}'Q\hat{\beta}$ the bias of which can be characterized as follows:

$$\mathbb{E}[\hat{V}_Q] = V_Q + \mathbb{E}[\varepsilon'Z(Z'Z)^{-1}Q(Z'Z)^{-1}Z'\varepsilon] \quad (6)$$

This bias arises from an insufficient number of movers between jobs (or firms in the classic AKM literature) and is therefore termed the *limited mobility bias*. As a result, the variance of job (or firm) fixed effects is overstated at the expense of understating the co-variance between workers and jobs. This is a problem the whole literature attempting AKM decompositions faces but the scale of the issue has only recently become apparent with [Bonhomme et al. \[2020\]](#) for example showing that the limited mobility bias is severe and that bias correction is very important. In our case, relative to more standard worker-firm decompositions, by introducing jobs we increase mobility intuitively by capturing within-firm, cross-occupation moves, but simultaneously increase the number of parameters to be estimated. Therefore it's unclear ex-ante whether limited mobility bias represents a smaller or larger problem in our framework.

We solve the limited mobility bias by following [Bonhomme et al. \[2019\]](#) in first clustering job using a k-means algorithm, into K clusters. By performing this first step we greatly increase movers between clusters whilst retaining much of the underlying variation (30 clusters capture over 85% of the underlying job level wage variation).

A.5 Analysing the result of misspecifying the model with worker-firm FE when the true DGP is with worker-job FE

In the paper, we noted that when the true data-generating process underlying the data involves worker-job (firm-occupation pair) fixed effects but a worker-firm FE two-way fixed effect regression is specified, the model leads to the overestimation of variance attributed to individuals under certain conditions. We show this in more detail in this section. This result is conceptually similar to the work in [Abowd et al. \[1999\]](#) to show the effects of misspecifying the model with worker-industry FE when the true model is worker-firm.

We can rewrite the true data-generating process as follows.

$$\begin{aligned}\ln(w_{it}) &= \alpha_i + \phi_{F(i,t)} + (\psi_{J(i,t)} - \phi_{F(i,t)}) + \varepsilon_{it} \\ \phi_f &= \frac{1}{n_f} \sum_{j \in \bar{f}} \psi_j\end{aligned}$$

It seems from this rewriting that what misspecification does is to add a term to the part that varies across individuals and periods which is now an unobserved omitted variable, and which will therefore cause the regression to estimate the individual fixed effects and the mean firm fixed effects with bias. Let W denote a $NT \times 1$ vector of wages, L a $NT \times N$ matrix of indicators for individuals $i = 1, \dots, N$, and J a $NT \times H$ matrix of indicators for jobs $j = 1, \dots, H$. Let F denote a $NT \times G$ matrix of indicators for the firm the job is at (for $f = 1, \dots, G$). Let ε denote a $NT \times 1$ vector of ε_{it} errors. Finally, let K be a $H \times G$ matrix where each row $j = 1, \dots, H$ indexes which firm the job is at. Let α denote a $N \times 1$ vector of personal fixed effects, Ψ a $H \times 1$ vector of job fixed effects and Φ a $G \times 1$ vector of firm-specific mean job fixed effects.

$$\begin{aligned}W &= L\alpha + J\Psi + \varepsilon \\ &= L\alpha + F\Phi + J(\Psi - K\Phi) + \varepsilon\end{aligned}$$

When J is not observed, estimating an OLS regression on L and F leads to omitted

variable bias. Suppose we estimate the misspecified model. Then the parameters estimated have omitted variable bias as follows, where $M_A = I - A(A'A)^{-1}A'$ for an arbitrary matrix A .

$$\alpha^* = \alpha + (L'M_F L)^{-1} L' M_F J(\Psi - K\Phi) \quad (7)$$

$$= \alpha + (L'M_F L)^{-1} L' M_F J\Psi \quad (8)$$

$$\Phi^* = \Phi + (F'M_L F)^{-1} F' M_L J(\Psi - K\Phi) \quad (9)$$

$$= (F'M_L F)^{-1} F' M_L J\Psi \quad (10)$$

The M_F and M_L matrices are residualising matrices, where $M_F L$ refers to the residual of a linear projection of F on L as in the Frisch-Waugh-Lovell theorem. It is instructive to re-write the expressions above as follows.

$$\alpha^* = \alpha + \beta \quad (11)$$

$$\Phi^* = \iota \quad (12)$$

Where β is the regression coefficient found from estimating the following by OLS: $J\Psi = (M_F L)\beta + \xi$ and ι is similarly the regression coefficient from estimating $J\Psi = (M_L F)\iota + \xi$. Thus, the bias in variance estimation is easily characterisable.

$$\mathbb{V}[\alpha^*] - \mathbb{V}[\alpha] = \mathbb{V}[\beta] + 2Cov(\alpha, \beta) \quad (13)$$

$$\mathbb{V}[\Phi^*] - \mathbb{V}[\Phi] = \mathbb{V}[\iota] - \mathbb{V}[\Phi] \quad (14)$$

From this formula we can see that the individual fixed effects will be over estimated if $Cov(\alpha, \beta)$ is not sufficiently negative. To gain intuition as to the signs and magnitudes of $\mathbb{V}[\beta]$ and $2Cov(\alpha, \beta)$, we must understand β . β is an $N \times 1$ vector of coefficients, an individual element of β is large if for that individual the correlation between the firm-residualised individual and job effect is large. If the strength of this relationship varies over individuals $\mathbb{V}[\beta]$ will be large. If this relationship is stronger for higher-FE individuals

$Cov(\alpha, \beta)$ will be positive, this is the most likely scenario.

It's harder to sign the difference in firm fixed effect variance. ι is the firm-analagous version of β , it is large if the firm conditional on worker is correlated with job FE. However, it's unclear whether the strength of variation in this coefficient will be less than or greater than the variation in firm effects itself and so signing the difference is difficult.

In general, from the above we can say that we expect the variance of the estimated worker and firm fixed effects in the mis-specified model to be wrong, and for it to be increasingly wrong the more residual variation there is in the relationship over workers between workers and jobs (conditional on firms) and over firms between firms and jobs (conditional on workers).

A.6 Comparing log-wage decomposition results against other results in the literature

In this section we compare results from estimating the usual worker-firm AKM model decomposition on our UK data to that from previous studies using data in other countries. Table 5 displays the main results. Panel [A] compares results using the basic AKM decomposition which does not correct for limited mobility bias whereas panel [B] shows results correcting for limited mobility bias. The main take away from this table is that in these specifications the UK data appears to behave similarly to that from other contexts. One possible exception is in the estimated negative effect found due to sorting in panel [A]. However, this is not as unusual a finding as it would appear from this table; in a recent survey, [Bonhomme et al. \[2020\]](#) show that in 10 of the 37 AKM estimates from previous studies, a negative sorting effect is found.

Table 5: Comparison of basic AKM and clustered correlated fixed effect decomposition results with select estimates from previous studies.

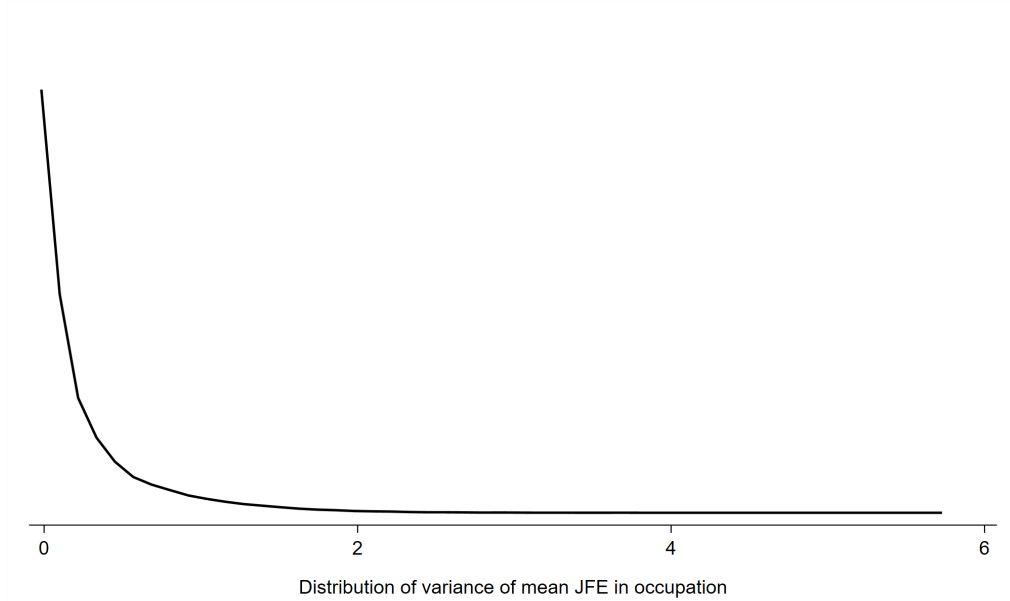
[A] Basic AKM Decomposition					
Paper	Period	Country	WFE	FFE	Sort
Card et al. [2013]	2002-09	Germany	51%	21%	16%
Card et al. [2016]	2002-09	Portugal	58%	20%	11%
Song et al. [2019]	2007-13	USA	52%	9%	12%
Alvarez et al. [2018]	2008-12	Brazil	58%	16%	18%
Our data	2011-19	UK	64%	24%	-2%
[B] Clustered Random Effects Bias Correction					
Paper	Period	Country	WFE	FFE	Sort
Bonhomme et al. [2019]	2002-04	Sweden	60%	3%	12%
Our data	2011-19	UK	54%	10%	19%

Notes: This table compares results using the UK data used in this paper when estimating classic and bias-corrected AKM models, to those found in selected papers from the previous literature.

A.7 Direct evidence that occupational changes are relatively unlikely

In section 4.1, we argued that the misspecification of the AKM leads to over estimation of the variance of the individual fixed effect especially if an individual’s occupation is persistent over time. In this section we provide some direct evidence of the persistence of occupation choice. We use mean job fixed effect within an occupation as a proxy for the relative productivity of the job. We then calculate the variance of the mean job fixed effect across occupations for each individual, and normalise it by the value of $\mathbb{V}[\mathbb{E}_o[\psi_j]]$ for the whole population. This measures an individual’s relative variance in occupational quality for their time in the panel relative to the overall population between-occupation job fixed effect variance. We plot the values for workers who have been in the panel for at least 5 years in figure 13 below. We find that this variance is very low for most individuals and the distribution is skewed towards 0. This suggests that most workers do not substantially change the quality of the occupation they are employed at over time.

Figure 13: Distribution of the variance of mean JFE within occupations over individuals



Notes: This figure plots the variance of the occupation mean fixed effect for the occupations an individual was observed in over the length of the panel, normalised over the overall variance in occupational mean job fixed effects.

A.8 Additional results on the relation between job FE as wage premia related to rent-sharing

As discussed in the introduction, estimates of the firm fixed effects in implementations of the AKM model have been interpreted as wage premia arising from firms rent-sharing with their workers (see [Card et al. \[2018\]](#) in particular for an attempt to synthesise the rent-sharing and AKM literatures). We check this story in our setting by investigating whether the job fixed effect $\psi_{J(i,t)}$ correlate with the firm's average productivity. Specifically, we consider a model where the job fixed effects can be modelled as follows, where $\phi_{F(i,t)}$ denotes industry fixed effects and $\chi_{Occ(i,t)}$ denotes occupation fixed effects.

$$\psi_{J(i,t)} = \alpha_0 + \beta \cdot \ln(\text{prod}_{F(i,t)}) + \phi_{F(i,t)} + \chi_{Occ(i,t)} + u_i \quad (15)$$

Since $\psi_{J(i,t)}$ is an estimated coefficient, we have to take into account its covariance matrix when estimating the parameters of equation 15 including β . We do this by substituting equation 15 into equation 1, and estimating the whole model²³. We use data on firm characteris-

²³An alternative approach is to use Generalised Least Squares after estimating the covariance matrix of

tics from the Business Structure Database (BSD), which contains detailed panel information on firm formation, as well as some basic characteristics like industry, employment, tax status and revenue. A major disadvantage of the BSD is that it does not contain information on firm profits, and we are unable as a result to test directly whether the estimated coefficients relate to firm productivity. We proxy for productivity using revenue per worker.

We run our analysis on three samples. Besides the full sample, we also consider splitting the sample into two groups (1) private workers not covered by a collective bargaining agreement and (2) private workers who are covered by a collective bargaining agreement and public/non-profit sector workers²⁴. If the estimated fixed effects were driven by rent-sharing, then we would expect that workers in private firms not covered by a collective bargaining agreement to have a higher value of β_o over all occupation groups. We thus estimate equation 15 on sub-samples of the two groups. Results for private, non-CA workers are presented in column (2) while results for public sector or workers under CA are presented in column (3)²⁵.

We present the results in table 6. We find that the job fixed effect is increasing in the revenue per worker of the firm, and that this correlation is greater for the sub-sample of jobs at private firms that are not covered by collective agreements. This coheres with the interpretation of these job fixed effects as wage premia paid to workers.

$\psi_{J(i,t)}$.

²⁴We use indicators in the ASHE dataset on whether the worker works at a public, private or non-profit institution. We also use indicators in the dataset on whether the worker is covered by a collective agreement. We lump workers under all kinds of collective agreements in the same category.

²⁵We also calculated the regression on public/private, and CA/non-CA sub-samples separately. Results for public sector workers are similar to those for CA workers and results for private workers are similar to those for workers not under CAs. Thus, we present only results for the two combined groups, noting that the results are not driven in particular by either the public/private or the CA/no-CA dimension.

Table 6: Estimates of β in equation 15 (2011-19)

	All workers	Private, no CA workers	Public or CA workers
Variable	(1)	(2)	(3)
Firm revenue per worker	0.018 (0.002)	0.027 (0.002)	0.005 (0.002)
Obs	1537793	740,090	715,920
R^2	0.897	0.911	0.906

Notes: This table shows the results from estimating $\psi_{J(i,t)} = \alpha_0 + \beta \cdot \ln(\text{prod}_{F(i,t)}) + \phi_{F(i,t)} + \chi_{Occ(i,t)} + u_i$ on the full sample in column one, a sample of private workers not covered by a collective bargaining agreement in column two, and on public workers or those covered by a collective bargaining agreement in column three.

We run a similar regression with different specifications to check if the relation to firm characteristics differs across occupations. Table 7 summarises the coefficients of a similar regression to equation 15, but where we interact the log worker productivity variable with the 1 digit occupation group; we also plot the coefficients in figure 14. We find that for all occupation groups, there is a positive relation between revenue per worker and the job fixed effect. We find statistically significant relationships for all occupation categories for the sample of jobs at private firms and jobs not covered by collective agreements. On the other hand, we find that the relations between the wage premium for managers, professionals, customer service and machine operators and log revenue per worker are not statistically significant. This again seems to be evidence in support of the wage premium interpretation of the job fixed effects.

An interesting aspect of the results in figure 14 is how the coefficient varies across occupations. The coefficient is broadly higher for more skilled workers than for less skilled workers for jobs in private firms, with workers in care and leisure having a particularly low correlation and customer service having a high correlation. In the public sector however, the premium paid to managers and professionals have no relation to firm revenue per worker, but the premium paid to mid and low-skill occupations seem to still exhibit some relation. This could reflect workers in high-skill occupations in these public institutions being better able to insure for wage variation.

For robustness, we check whether this pattern persists if we use a measure of the skill

requirement of the occupation²⁶. The results for these regressions are summarised in table 8 and figure 15. We similarly find that overall the relation between revenue per worker and the job fixed effect is greater for lower-skilled workers than for higher-skilled workers, but this is primarily driven by a composition effect. The relation is statistically insignificant and near zero for high skill workers at public firms and at jobs covered by collective agreements, but very large for similarly skilled workers in jobs at private firms not covered by a collective agreement. In contrast, the correlation of job fixed effect with revenue per worker is positive and statistically significant in all samples for lower skilled workers. We leave a full explication of these puzzles to future work.

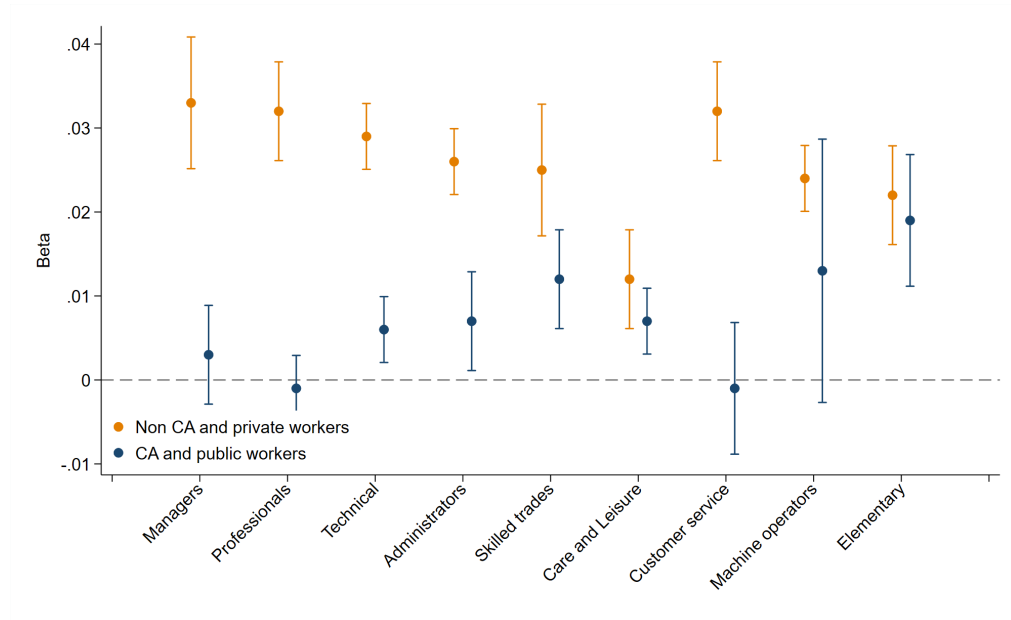
Table 7: Coefficients on log revenue per worker interacted with 1 digit occupation group

Occupation Group	All workers (1)	Private, no CA workers (2)	Public or CA workers (3)
Managers, directors and senior officials	0.023 (0.003)	0.033 (0.004)	0.003 (0.003)
Professional occupations	0.010 (0.004)	0.032 (0.003)	-0.001 (0.002)
Associate professional and technical occupations	0.019 (0.002)	0.029 (0.002)	0.006 (0.002)
Administrative and secretarial occupations	0.021 (0.002)	0.026 (0.002)	0.007 (0.003)
Skilled trades occupations	0.024 (0.003)	0.025 (0.004)	0.012 (0.003)
Caring, leisure and other service occupations	0.010 (0.002)	0.012 (0.003)	0.007 (0.002)
Sales and customer service occupations	0.022 (0.004)	0.032 (0.003)	-0.001 (0.004)
Process, plant and machine operatives	0.025 (0.003)	0.024 (0.002)	0.013 (0.008)
Elementary occupations	0.022 (0.002)	0.022 (0.003)	0.019 (0.004)
Obs	1537793	740,090	715,920
R^2	0.897	0.911	0.906

Notes: This table shows the coefficients and corresponding standard errors found when estimating equation 15, interacting log revenue per worker with occupation group.

²⁶We use the skill requirements for occupations used by the UK Home Office, and set out in their Appendix J tables. RQF 3 corresponds to a high school degree, while RQF 4 corresponds to 1 year of a bachelors degree. RQF 6 corresponds to a bachelor degree.

Figure 14: Relation between job fixed effect and firm revenue per worker, by occupation group (2011-19)



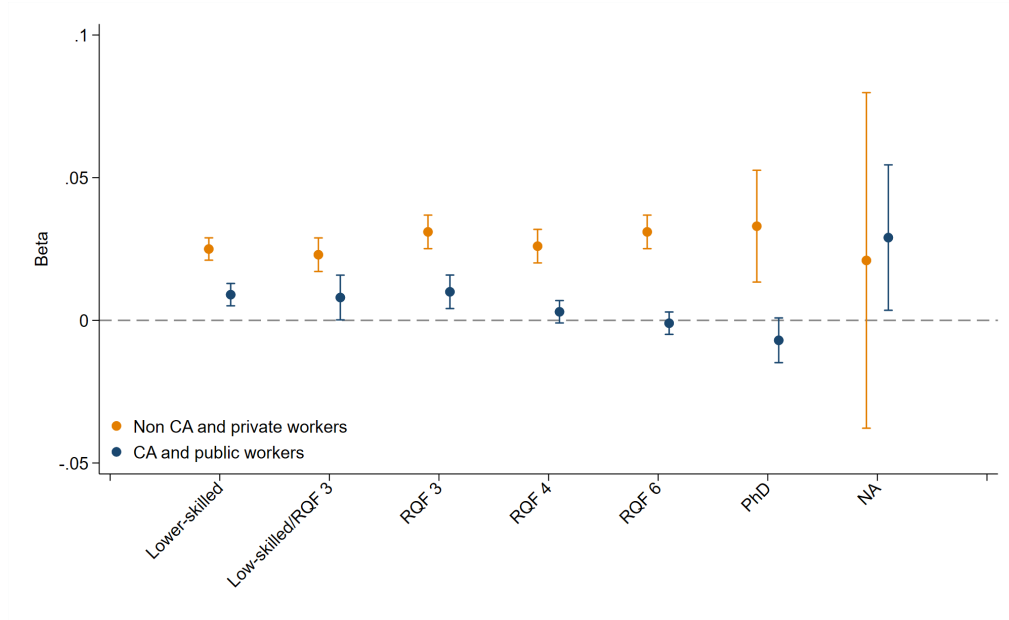
Notes: This figure shows the coefficients and corresponding 95% confidence intervals found when estimating equation 15, interacting log revenue per worker with occupation group.

Table 8: Coefficients on log revenue per worker interacted with skill groups

	All workers	Private, no CA workers	Public or CA workers
Occupation Group	(1)	(2)	(3)
Lower-skilled	0.019 (0.002)	0.025 (0.002)	0.009 (0.002)
RQF 3/Lower-skilled	0.016 (0.003)	0.023 (0.003)	0.008 (0.004)
RQF 3	0.023 (0.002)	0.031 (0.003)	0.010 (0.003)
RQF 4	0.015 (0.002)	0.026 (0.003)	0.003 (0.002)
RQF 6	0.014 (0.004)	0.031 (0.003)	-0.001 (0.002)
PhD	0.014 (0.006)	0.033 (0.010)	-0.007 (0.004)
Unclassified	0.016 (0.010)	0.021 (0.030)	0.029 (0.013)
Obs	1534631	738,048	714,953
R^2	0.897	0.911	0.906

Notes: This table shows the coefficients and corresponding standard errors found when estimating equation 15, interacting log revenue per worker with the skill level of the occupation, as determined by the UK Home Office.

Figure 15: Relation between job fixed effect and firm revenue per worker, by skill group (2011-19)

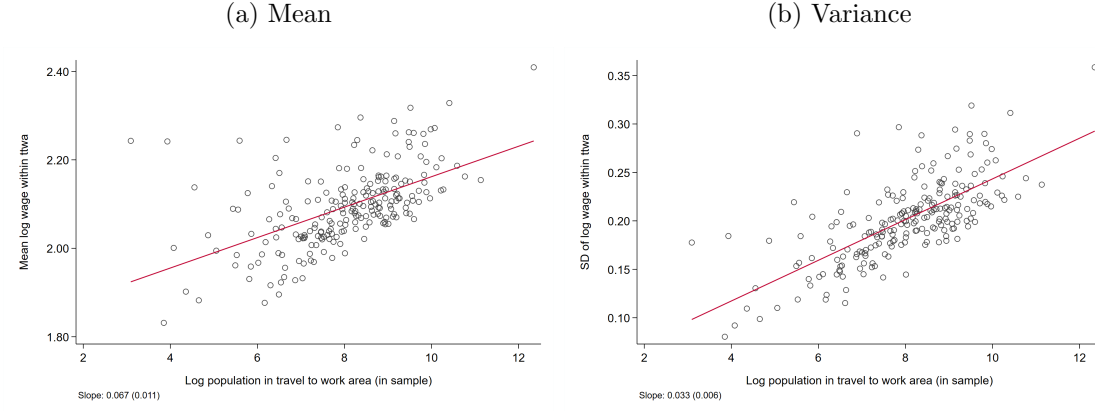


Notes: This figure shows the coefficients and corresponding 95% confidence intervals found when estimating equation 15, interacting log revenue per worker with the skill level of the occupation, as determined by the UK Home Office.

A.9 Appendix figures and tables relating to wage inequality across geography

Panel 16a reproduces the well-known fact that mean wages are higher in larger labour markets, while panel 16b shows that the variance of log-wages is also increasing log-linearly with log-population within a TTWA.

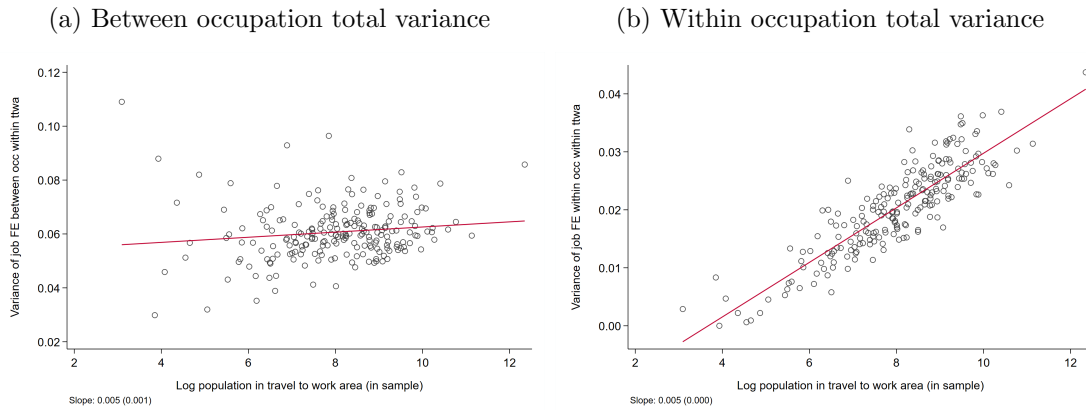
Figure 16: Within TTWA log population vs moments of log pay (2011-19)



Notes: Panel 16a plots the mean of log-hourly wage (y-axis) against log-population (x-axis) in the travel-to-work area. Panel 16b plots the variance of log hourly wage (y-axis) against log-population (x-axis) in the travel-to-work area.

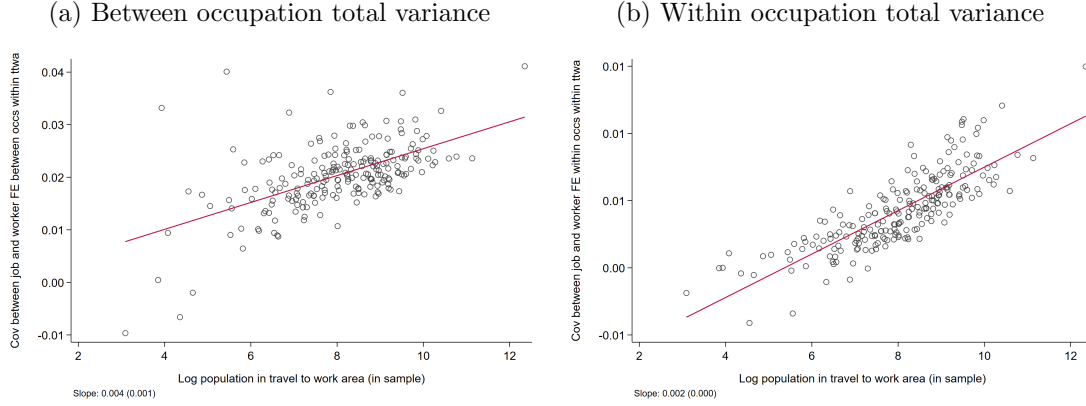
In figure 17, we plot scatterplots of the absolute values of the part of log-wage variance explainable by between-occupation job fixed effect variance (panel 17a), and the part explainable by within-occupation job fixed effect variance (panel 17b). Analogously, figure 18 plots the absolute values of the covariance of job fixed effects and individual fixed effects attributable to between-occupation sorting (panel 18a) and within-occupation sorting (panel 18b).

Figure 17: Within TTWA log population vs variance of job FE between and within occupation



Notes: Panel (a) plots a scatterplot of TTWAs with log-population on the x-axis and the between-occupation component of job-class fixed effect variance on the y-axis. Panel (b) plots a scatterplot of TTWAs with log-population on the x-axis and the within-occupation component of job-class fixed effect variance on the y-axis.

Figure 18: Within TTWA log population vs covariance components decomposed into between and within occupation parts



Notes: Panel (a) plots a scatterplot of TTWAs with log-population on the x-axis and the between-occupation component of worker-job-class covariance on the y-axis. Panel (b) plots a scatterplot of TTWAs with log-population on the x-axis and the within-occupation component of worker-job-class covariance on the y-axis.

A.10 Is the relation between size of travel to work area and variance due to occupational composition?

The increase in within-occupation sorting documented in figure 9d could reflect the tendency for jobs in high-wage occupations to locate in larger labour markets, given that we know that the degree of sorting is higher in higher wage occupations. Thus, the relation documented here between TTWA size and positive assortative matching could reflect a composition effect rather than a true correlation with size. To obtain direct evidence on this, we calculate the standard deviation of job fixed effects within each occupation for each travel-to-work area in each period ($\sigma_{kl\tau}$) as well as the correlation between worker fixed effects and job fixed effects within each travel-to-work area, occupation and period ($\rho_{kl\tau}$) (where τ indexes the period). We then perform simple regressions of these moments on log population within the travel-to-work area, controlling for occupation and period fixed effects (denoted by γ_k , and γ_τ respectively for the standard deviation regression, and by δ_k and δ_τ for the correlation regression).

$$\sigma_{kl\tau} = \gamma_0 + \gamma_1 \cdot \ln(pop_l) + \gamma_k + \gamma_\tau + u_{kl\tau} \quad (16)$$

$$\rho_{kl\tau} = \delta_0 + \delta_1 \cdot \ln(pop_l) + \delta_k + \delta_\tau + v_{kl\tau} \quad (17)$$

The results of this regression for γ_1 and δ_1 are reported in table 9 for both the full set of TTWA-occupation-period cells and for the set of cells with at least 30 observations. We find that even after controlling for occupational differences, there is still a residual correlation between travel-to-work area size and both the standard deviation of job fixed effects and the correlation between worker and job fixed effects. This qualitative result is robust to dropping smaller cells for which these moments might not be as reliable. Thus, we conclude that being in a larger local labour market is associated with larger variance of job fixed effects, larger variance of worker fixed effects, and a greater correlation between worker and job fixed effects, even conditional on the occupational composition of the travel-to-work area.

Table 9: Coefficients on log population from regression of occ-ttwa moments

Coefficient on log pop	All > 10	All > 30
Standard deviation of job FE	0.018 (0.001)	0.013 (0.001)
Standard deviation of worker FE	0.015 (0.001)	0.008 (0.001)
Correlation of worker and job fixed effects	0.002 (0.000)	0.001 (0.000)
N	42,328	18,749

Notes: This table shows the results from estimating equations of the form given in 16 of the relationship between the TTWA-occupation-period standard deviation of job and worker FE as well as the correlation between the two and TTWA population. Column one gives results for all cells with more than 10 observations and column two for all cells with more than thirty observations. Standard errors are clustered at the TTWA level.