

# *Recent changes in British wage inequality: evidence from large firms and occupations*

Article

Accepted Version

Schaefer, D. and Singleton, C. ORCID: <https://orcid.org/0000-0001-8247-8830> (2020) Recent changes in British wage inequality: evidence from large firms and occupations. *Scottish Journal of Political Economy*, 67 (1). pp. 100-125. ISSN 1467-9485 doi: <https://doi.org/10.1111/sjpe.12225> Available at <https://centaur.reading.ac.uk/83562/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1111/sjpe.12225>

Publisher: Wiley

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online



# Recent changes in British wage inequality: Evidence from large firms and occupations\*

Daniel Schaefer<sup>†</sup>

Carl Singleton<sup>‡</sup>

*Published at: Scottish Journal of Political Economy. (2019; In Press)*

## Abstract

Using a linked employer-employee dataset covering large firms, we present new evidence on British wage inequality trends over the past two decades. Differences between firms in the average wages they paid did not drive these trends. Between 1996 and 2005, greater wage variance within firms accounted for eighty-six percent of the total increase in wage variance among employees. In the following decade, wage inequality between firms continued to increase, whereas overall wage dispersion decreased. Approximately all the contribution to inequality dynamics from estimated firm-specific factors, throughout the employee wage distribution, disappears after accounting for the changing occupational content of wages.

*Keywords:* wage inequality, within-firm inequality, occupational wage premiums

*JEL codes:* D22, E24, J31

---

\*We are grateful for comments and advice from Richard Dickens, Mike Elsby, Jan Grobovšek, Maia Güell, Ines Lee, Giovanni Razzu, Andy Snell and the anonymous referees, as well as participants in 2017 at the European Association of Labour Economists 29th Conference, Royal Economic Society Symposium of Junior Researchers, Society for Economic Dynamics Annual Meeting and International Workshop on Establishment Panel Analyses (IAB Nuremberg). This paper is mostly based on data from the New Earnings Survey Panel Dataset (Crown copyright 2016), having been funded, collected and deposited by the Office for National Statistics (ONS) under secure access conditions with the UK Data Service. Neither the ONS nor the Data Service bear any responsibility for the analysis and discussion of the results in this paper.

<sup>†</sup>University of Edinburgh, [daniel.schaefer@ed.ac.uk](mailto:daniel.schaefer@ed.ac.uk)

<sup>‡</sup>University of Reading, [c.a.singleton@reading.ac.uk](mailto:c.a.singleton@reading.ac.uk); with thanks to the Economic and Social Research Council (UK) for funding support under Grant No. ES/J500136/1, in addition to the Scottish Economic Society.

# 1 Introduction

The long-term trend of rising wage inequality in Great Britain has been extensively documented ([Hills et al., 2010](#); [Machin, 2011](#); [Belfield et al., 2017](#)). The majority of this increase in Britain occurred in the 1980s, and has tended to occur within observable sub-groups of workers and firms, such as education levels and industry sectors, as in the United States and several other countries. Although well-studied, some ambiguity remains over what drives changes in the wage distributions of labour markets such as Britain's. This paper contributes to answering this question in two ways. First, we use a new source of linked employer-employee data to ask how much have differences between firms, relative to within, accounted for recent wage inequality trends in Britain. The answer to this question is particularly relevant for policy makers, given the recent public focus on income inequality: for example, if firms were the main drivers of rising inequality, then corporation taxes could be more suitable to reverse inequality trends than income taxes. Second, unlike similar datasets in other countries, our data include detailed descriptions of employee occupations, and thus we ask how occupational polarisation affects between- and within-firm inequality trends.

Among full-time employees in Britain, we find that over eighty percent of the increase in the variance of log weekly wages between 1996 and 2005 occurred within firms. In the subsequent decade overall inequality decreased, whereas the dispersion of average firm wages increased. We document similar results for hourly and annual wages. Previously, [Faggio et al. \(2010\)](#) have found that rising wage inequality in Britain in the fifteen years prior to 1999 was almost entirely accounted for by an estimate of between-firm variance. A contribution of our paper is to extend these previous results, using the same survey data of wages and hours, but by instead matching a representative sample of employees to the majority of large firms. This provides us with a robust sample of employer-employee linked jobs, as opposed to using some separate source to estimate firm average wages; [Faggio et al. \(2010\)](#) lacked data on the wages within specific firms. [Mueller et al. \(2017\)](#) have also studied British wage inequality from the firm's perspective. Using data on average pay at hierarchy levels in a small number of firms, they found substantial within-firm inequality in the years 2004-13 and that this tended to increase as firms grew. They also suggested that overall wage inequality trends could be related to an increasing concentration of employment in larger firms.

Several studies of other countries have documented that trends in employee wage inequality and the dispersion of firm productivity or firm average wages tend to coincide (see among others for the US: [Davis and Haltiwanger, 1991](#); [Dunne et al., 2004](#); [Barth et al., 2016](#); [Song et al., 2019](#). For Sweden: [Nordström Skans et al., 2009](#); [Akerman et al., 2013](#). For West Germany: [Card et al., 2013](#). For Brazil: [Alvarez et al., 2018](#); [Helpman et al., 2017](#). For Eastern Europe: [Gromadzki et al., 2019](#). See also the literature review by [Card et al., 2018](#)). At first look our results for Britain would appear to conflict with this wider literature. The data that we use are a one percent random sample of employees, and so we mostly limit our attention to large firms, representing approximately forty percent of all UK employment. The overall patterns in wage inequality for this forty percent are not dissimilar to the whole UK economy. But [Song et al. \(2019\)](#) have shown that the contribution of between-firm wage dispersion to overall changes is smaller among larger US firms. Although our results are robust to varying how we select the sample of employees, and consequently to increasing or decreasing the average firm size, we would advise some caution in comparing our findings to those found elsewhere using more widely representative data.

Nonetheless these British data offer some clear advantages. They are considered to be accurate records from firms' payrolls (Nickell and Quintini, 2003), giving measures of annual and weekly earnings, and their constituent components, including hours worked. Using information which firms provide on their employees' occupations, we ask how much of the estimated contribution from firm-specific differences to wage inequality changes is accounted for by the observed occupational content of wages. The answer is approximately all of it, throughout the wage distribution. Some combination of changes to between-occupation inequality and the sorting of occupations across firms accounts for changes in the dispersion of firm-specific wages over the last two decades in Britain.

We consider this inclusion of occupations in the analysis a contribution over previous studies because it is well-known that the polarisation of employment across occupations – the increase in shares of employment in relatively high- and low-skilled occupations – accounts for a significant part of long-run wage inequality changes in the UK, US and elsewhere (Goos and Manning, 2007; Autor et al., 2008; Williams, 2013). Song et al. (2019) have suggested that while skill-biased technological change could account for overall wage inequality increases and the polarisation of employment in the US, across firms there may have been greater specialisation and concentration of occupations, thus explaining some part of the rising dispersion in average firm wages. This theory has been largely untested due to a lack of comprehensive data covering long periods of employee wages, detailed occupations and firm identities. A notable exception is Weber-Handwerker and Spletzer (2016), who made some progress on this for the US between 2000 and 2011, finding that the changing composition of employment over occupations, as opposed to industries, accounted for almost half of between establishment wage growth. They also found that the changing extent of occupational concentration within firms, measured by Herfindahl indices, explains only a small part of this between establishment wage growth. In contrast to Weber-Handwerker and Spletzer, we also allow for changes to estimated occupational wage premiums, as well as their composition over firms and employment. As such, our results suggest that the polarisation of the UK labour market can account for approximately all of the changes in firm-specific inequality throughout the wage distribution over the past two decades.

This new evidence from Great Britain suggests that the estimated importance of between-firm inequality in driving overall employee wage variance, found elsewhere, could also represent an important role for the occupational transformation of firms and labour markets, which has yet to be accounted for.

The remainder of the paper proceeds as follows: Section 2 describes the data, Section 3 presents the results from decompositions of wage variance over the last two decades in Britain, Section 4 describes the dynamics of inequality throughout the wage distribution, and Section 5 summarises. Further information is presented in the Online Appendix concerning the data, sample construction, mathematical details and additional results.

## 2 Data

The data we use are from the Office for National Statistics (ONS) New Earnings Survey Panel Dataset (NESPD), 1975-2015. It is a continuing sample of approximately one percent of all Pay As You Earn (PAYE) taxpayers in Britain, covering up to 180,000 employee jobs per year.<sup>1</sup> The small number of jobs

---

<sup>1</sup>PAYE approximately covers all UK employees. It describes the system by which an employer deducts income taxes due and pension contributions before an employee subsequently receives their net wages.

not registered for PAYE in Britain, which tend to be of very low pay, are not sampled. Employees who are not paid in the reference period are also excluded. These are both potential sources of composition bias in measuring inequality changes, which could vary over the economic cycle. But it is an advantage that the dataset is a long-running panel, since we can expect many repeated observations of employer-employee matches.<sup>2</sup> The data are collected via a questionnaire issued to employers, who are required by law to complete it with reference to payrolls. They return the gross weekly earnings and hours worked of employees, and their detailed components, as well as an employee's occupation and other information related to remuneration, such as pensions and collective agreements. The reference period for the survey is always a week in April. Gross annual earnings for the year to April have been recorded since 1999.

It is an advantage of these data that we can consider the robustness of results across different frequencies of pay. For example, the compositional differences in two jobs samples from the NESPD which contain either non-missing observations of weekly or annual wages could be large, given that for the latter individuals must have been with the same employer for at least twelve months. Related, the timing of bonus payments tends to be seasonal in Britain. Approximately half of all such payments economy-wide, and over seventy percent in the financial and insurance activities sector, occur in the 'bonus season' of December-March.<sup>3</sup> This seasonal pattern is consistent across years. Therefore, the measures of weekly and hourly wages in April will only capture a small part of this pay component. The fraction of total employee remuneration which is from bonus payments has been constant at six to seven percent over the past two decades, but in the financial and insurance activities sector this has fluctuated between twenty and thirty-five percent around the Great Recession. If bonus payments have significantly affected trends in the employee wage distribution, then we could expect our results to differ between using annual vs. weekly or hourly measures.

Information on employer size and industry classification was added to the NESPD from 1996 onwards by the [ONS](#), using Her Majesty's Revenue and Customs' Inter-Departmental Business Register (IDBR), an administrative census of all UK registered companies. Only tiny businesses consisting of the self-employed are not found on the IDBR. The employer reporting unit observed in the NESPD is generally the enterprise or a local unit thereof. The 'firm' is an enterprise for the vast majority of the data used in the following analysis.<sup>4</sup> For a sub-period, 2002-15, the enterprise of all jobs is identified, and we use this as a robustness check of whether our less precise definition of a firm could qualitatively affect any of the main results. We can do this using the annual cross-section datasets of the Annual Survey of Hours and Earnings, from which the NESPD in later periods is derived. These sources of earnings data are considered to be unusually accurate, at least when compared with household based surveys ([Nickell and Quintini, 2003](#)). The NESPD has undergone minor methodological changes over its lifetime, but the principal aim of collecting detailed and precise information on hours, pay and occupations has remained consistent. In [Online Appendix A](#) we summarise the relevant changes, as well as providing greater detail than what follows on the construction of our analysis sub-samples and variables.

---

<sup>2</sup>We do not exploit this feature fully since in the current publicly available form of the dataset employers can only be robustly identified over time for 2002-2015.

<sup>3</sup>See ONS statistical bulletin: "Average weekly earnings, bonus payments in Great Britain: financial year ending 2016".

<sup>4</sup>We are comfortable that the enterprise is a typical definition of the firm, as defined for UK government administrative purposes. IDBR definition: "An Enterprise can be defined as the smallest combination of legal units (generally based on VAT and/or PAYE records) that is an organisational unit producing goods or services, which benefits from a certain degree of autonomy in decision-making, especially for the allocation of its current resources. An enterprise carries out one or more activities at one or more locations. An enterprise may be a sole legal unit."

## 2.1 Creating a large firms sample of the NESPD

For all sub-samples of the NESPD we include only those aged 16-64 and exclude jobs where pay in the reference period has been affected by absence or leave. For weekly wages and hours worked we use the reported values excluding any overtime. Hourly rates of pay are derived from gross weekly pay and basic (usual) hours worked. When considering annual earnings, we only ever use observations where the employee is reported as having been in the same job for at least a year. All monetary values are deflated to 1997 prices using the ONS Retail Price Index from April, to match the reference period of the NESPD.<sup>5</sup> To analyse and estimate a within-firm component of wage dispersion, we have to match sufficient numbers of employees to each observed firm. Hence we construct a large firms sample of the NESPD. We consider only jobs in each year at enterprises with 250 employees or more according to the IDBR.<sup>6</sup> In the baseline sample we keep only full-time jobs, defined as working over thirty hours in a week before overtime, and in each year then keep firms for which there are ten or more job observations with non-missing values of pay and hours worked. This firm-based selection imposes a de facto minimum firm size of more than a thousand employees. We construct several other sub-samples, which are discussed in the results, where we vary the minimum number of job observations required per firm and add part-time workers.

Throughout the following analysis and results one can generally replace any reference to ‘firms’ with ‘large firms,’ or even ‘very large firms.’ This is clear when we compare the enterprise size distributions in 2013 of the UK population and the firms in our baseline NESPD sub-sample (Online Appendix Table A1). Over seventy percent of UK enterprises with at least 250 employees have less than a thousand. But in our baseline sample such firms are only five percent of the total number. On the other hand, firms with more than two thousand employees are relatively over represented: the sample includes a similar number of firms with over five thousand employees as there are in fact such UK enterprises.<sup>7</sup> Though in our baseline analysis we do not represent the whole firm size distribution of Britain, we can nonetheless claim to sample employees from practically all very large enterprises. As such we are able to study a significant fraction of jobs and wages: in 2013 the firms in our sample represented approximately forty percent of employee jobs.<sup>8</sup>

The focus on large firms implies that our results are likely to provide an upper bound on the estimated contribution of the within-firm component to inequality more generally, both in the cross-section and over time. This assessment is based on [Mueller et al. \(2017\)](#), who documented larger within-firm pay differences in larger UK companies, and that these differences were increasing over time as firms grew. They also suggested that convexities in firm production functions could explain these findings. Intuitively, if talent and firm size are multiplicative in the creation of value, then the efficient assignment of managers and professional workers should allocate more talent to larger firms. Even if those workers receive only a share

---

<sup>5</sup> Accessed from the ONS website 25/05/2016. We choose this particular price deflator, over for instance the Consumer Price Index, because of its much longer time series. Also, we study changes in log wage inequality and so this choice does not affect any of our results.

<sup>6</sup> The cut-off between the definition of Small and Medium Enterprises (SMEs) and Large firms in the UK is typically at 250 employees.

<sup>7</sup> Part of the non-sampling discrepancy is due to the NESPD being British as opposed to UK. In 2013 ONS data suggests there were thirty enterprises in Northern Ireland with over a thousand employees. Using enterprise identifiers from the ASHE to define firms in 2013, and otherwise the same criteria to construct the baseline large firms sample, gives us 598 enterprises with over five thousand employees.

<sup>8</sup> According to the ONS Labour Market Statistics Workforce Jobs series, there were approximately 27.5 million employee jobs in Great Britain in 2013.



of their marginal product, then we would still expect to observe more within-firm pay inequality in larger firms.<sup>9</sup>

Given we study the dynamics of wage inequality, we briefly document how the baseline sample's firm size distribution has evolved over time, between 1997 and 2007 for example (Online Appendix Table A2). The share of firms with more than two thousand employees increased by over thirteen percentage points in this period, with the largest increase among those with five thousand or more. The share of employee observations in very large firms similarly increased. The true distribution of these firms was relatively unchanged over the period, according to their administrative IDBR enterprise level of employment. We believe this difference reflects a shift since 2004 in the employer reporting unit of the earnings survey towards more commonly being the enterprise, as opposed to the local unit.<sup>10</sup> We also describe the sample's changing industrial make-up over the same ten years (Online Appendix Figure C1).<sup>11</sup> The share of firms associated with the manufacturing sector decreases notably, while real estate and business services firms are increasingly represented. We observe similar trends in the labour force shares of sectors, though in this case there is also a decline in the share of employees in public administration and defence.

An advantage of our data over those used in similar studies is the presence of employer descriptions of jobs and their assignment to a detailed occupational classification. Throughout the paper, occupations refer to the International Standard Classification of Occupations 1988 (ISCO88), unless stated otherwise. Due to inconsistencies in source data classifications we only consider occupations for the sub-period 1996-2010. Comparing the incidence of major occupation groups in the sample over time, some occupations are less prevalent in 2007 than in 1997, with a large decrease for professionals (Online Appendix Table A3). At the same time, the share of elementary occupations has increased by almost the same amount.

## 2.2 Describing wages in large firms and the NESPD

Since economy-wide trends in wages have been extensively documented elsewhere using the NESPD (e.g. Machin, 2011), here we focus on whether recent patterns among jobs in large firms have been notably different. Figure 1 compares selected percentiles of real log wages for full-time employees between our baseline sub-sample and the whole NESPD. Online Appendix Figure C2 similarly compares mean values. Median real wages were relatively stagnant during the early to mid 2000s. But they have seen a substantial decline since 2008, especially compared with the periods following previous UK recessions. The dispersion of log weekly wages increased for the whole NESPD persistently from 1975 to 1995 (Figure 1A). The variation of wages in our baseline sample is somewhat lower than in the whole NESPD. This is due to a tighter distribution of wages above the median among those working in larger firms. Generally though the pattern of wages across the large firms distribution is similar to the whole NESPD: for example, both show a steep increase in real hourly and weekly wages for top earners in the early 2000s, as well as a decline in dispersion at the onset of the Great Recession, driven by relatively higher earnings at the bottom. Figure 2A further demonstrates these changes by plotting real log weekly wages relative to 1996 for selected employee percentiles of the large firms sample.

---

<sup>9</sup>For a formal exposition of this theory and an application to US data see Gabaix and Landier (2008).

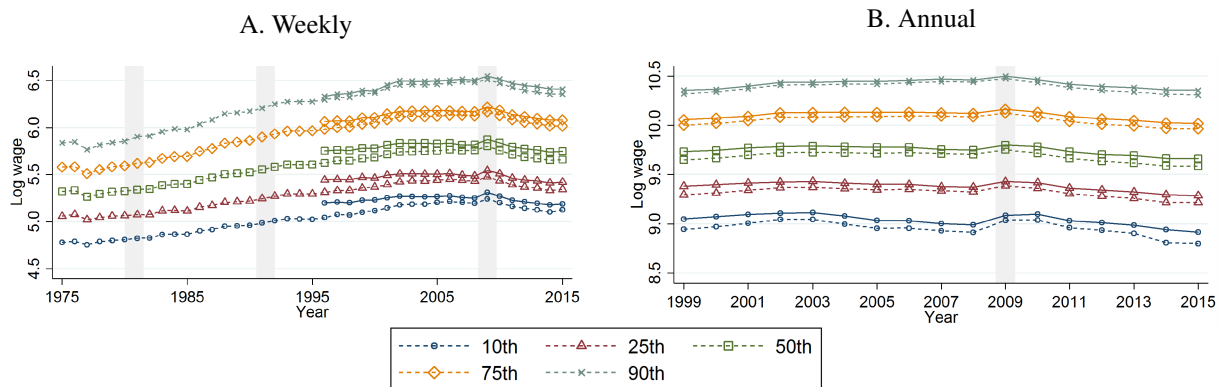
<sup>10</sup>This coincides with the replacement of NESPD with ASHE. Despite studying the documentation we cannot find any noteworthy reason for such a sizeable shift. As we show in what follows, we are confident that this does not qualitatively drive any of the main results.

<sup>11</sup>Throughout industry sectors refer to the Standard Industrial Classification (SIC) 2003.



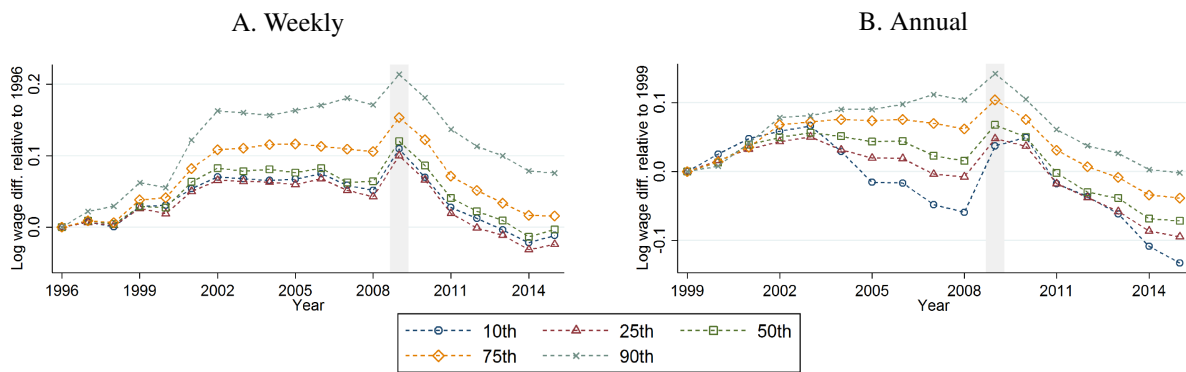
The increase in the variance of log annual wages, which include all performance related payments, was more substantial between 1999 and 2007 than for log weekly wages. As shown in Figures 1B and 2B, this is explained by real wages falling at the lower percentiles and only marginally rising at the median, while the ninetieth percentile increased consistently throughout this period. These patterns and comparisons are similar when we consider all employees and not only those working full-time (Online Appendix Figures C3-C5).

FIGURE 1: Percentiles of real log wages in large firms, full-time employees only, and comparison with the whole NESPD sample, 1975-2015



Note.- age 16-64 and full-time employees only. ‘Weekly’ exclude overtime. ‘Annual’ are for employees with the firm at least one year. See the text for further details of sample construction. Shaded areas represent official UK recessions. Solid lines are the series for a large firm sub-sample of the NESPD.

FIGURE 2: Percentiles of real log wages in large firms, full-time employees only: differences relative to 1996/9



Note.- see Figure 1.

### 3 Wage inequality trends: the role of between-firm variance

To account for how much of the variance in logarithmic employee wages is explained by differences in the average wages paid by firms, we use the decomposition of Davis and Haltiwanger (1991). As a measure of inequality, the variance of log wages has the advantage of being simple and tractable, relates to the standard Mincer-style wage regression model, and maintains mean independence. However, it does not satisfy stricter criteria for inequality measures, such as the Lorenz criterion or the Pigou-Dalton principle (see Sen, 1973).

The total variance of the natural logarithm of wages across a set of firms and their employees,  $V_e$ , can be decomposed into estimates of a within-firm component,  $V_{wf}$ , and the variance of average log wages

between firms,  $V_{bf}$ . Denoting the total number of firms in a given year by  $J$ , and the number of employees we observe in firm  $j = 1, \dots, J$  by  $N_j$ , such that the total sample number of employees is  $N = \sum_{j=1}^J N_j$ , then we can write

$$\underbrace{\frac{1}{N} \sum_{j=1}^J \sum_{i=1}^{N_j} (w_{ij} - \bar{w})^2}_{\text{Overall - } V_e} = \underbrace{\frac{1}{N} \sum_{j=1}^J \sum_{i=1}^{N_j} (w_{ij} - \bar{w}_j)^2}_{\text{Within-firm - } V_{wf}} + \underbrace{\sum_{j=1}^J \frac{N_j}{N} (\bar{w}_j - \bar{w})^2}_{\text{Between-firm - } V_{bf}}, \quad (1)$$

where  $w_{ij}$ ,  $\bar{w}$ , and  $\bar{w}_j$  denote respectively the log wage of employee  $i$  in firm  $j$ , the sample mean of log wages, and the sample mean of log wages within firm  $j$ .<sup>12</sup> For convenience, we leave implicit the dependence of  $j$  on  $i$  throughout the paper. The term capturing the between-firm component of wage dispersion weights by employment share the observed distance of a firm's estimated average wage to the overall average wage, such that larger firms have a potentially greater influence on wage dispersion than smaller firms. There are two possible choices of how to weight firms: by their shares of employee observations in the sample, or by their relative size as indicated by the IDBR recorded numbers of employees. Our preference throughout is the former, but we find this choice has no qualitative effect on results (see for example Online Appendix Figure C6).

We exclude the top one percent of all earners from the variance calculations in this section, though we will include them again later in Section 4. This is because the data are not top-coded and a small sample, and variance decompositions can be sensitive to any outliers or measurement error in these top earnings. Throughout the remainder of the paper we mostly focus on weekly wages, as these are recorded in the data independently of an employer's response for the hours worked of their employees. Hours responses to the earnings survey could potentially be quite subjective across firms, compared with payroll wages, and thus contribute significantly to any measurement error. Furthermore, the sample of weekly wages includes jobs which are less than a year old. These jobs would be excluded from an analysis of annual wages and their importance within the true wage distribution could vary over time.

Figure 3 plots the estimated components of (1) for each year between 1996 and 2015 for full-time employees. Table 1 summarises the decomposition results discussed throughout this section. Overall weekly wage dispersion increased by two log points over the entire sample period (column (9), Table 1). However there is an observable difference pre and post the 2008 financial crisis. The latter period experienced falling inequality, more than accounted for by the decreasing variance of wages within firms, while at the same time between-firm inequality continued to increase. Prior to 2008, the increase in within-firm inequality explained the majority of the overall three log points increase (over 80 percent: column (8), Table 1). The overall variance of log weekly wages mirrors closely the pattern of the within-firm component at shorter frequencies, as seen in Figure 3A. Both the pre-2008 increase and the post-2008 decrease in inequality were driven mostly by the within-firm component in Britain (see also Online Appendix Figure C7A).

---

<sup>12</sup>Sampling errors in the measures of firm average wages (or hours) will generally induce a positive bias in between-firm variance estimates and their shares of the overall variance. We do not attempt to correct this, and instead rely on our analysis being focused on trends, since the size of this bias is unlikely to vary significantly over the period studied. The literature in this area, such as Card et al. (2013), also acknowledges the bias from sampling error, and similarly tends to ignore it, by arguing that trends are unlikely to be affected. Here we are especially reliant on any changes to the NESPD/ASHE sample frame or method not affecting the level of bias over time. We are confident that this is qualitatively the case, given our knowledge of the timing of any such changes, as discussed in Online Appendix A.

TABLE 1: Summary of decomposition results for the variance in real log weekly wages, full-time employees

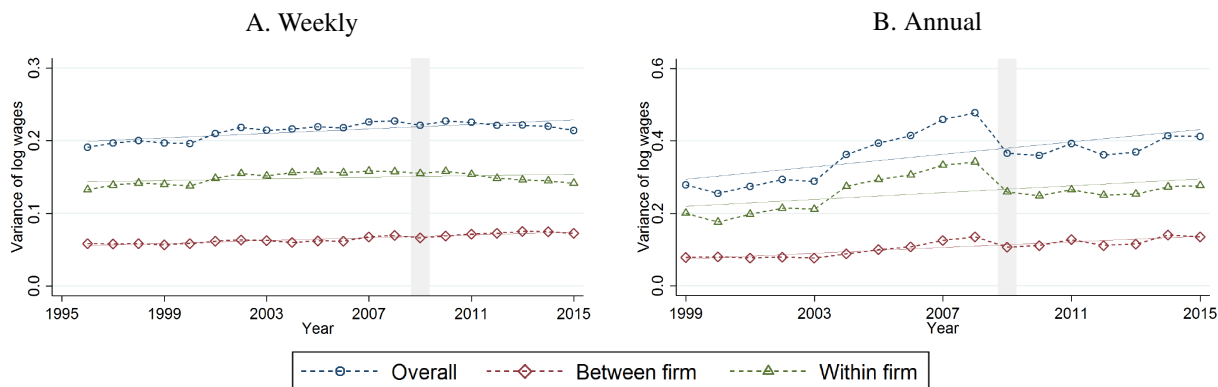
	1996*		2005		2015		Change 1996-2005		Change 1996-2015	
	Level (1)	Share (2)	Level (3)	Share (4)	Level (5)	Share (6)	Level (7)	Share $of^{**}$ (8)	Level (9)	Share $of^{**}$ (10)
Actual weekly wages:	0.191		0.219		0.214		0.028		0.023	
Between: total	0.058	0.305	0.062	0.284	0.072	0.339	0.004	0.14	0.014	0.61
Hourly wages <i>bf</i>	0.074	0.387	0.072	0.331	0.083	0.389	-0.001	-0.04	0.009	0.39
Usual hours <i>bf</i>	0.005	0.029	0.004	0.016	0.004	0.016	-0.002	-0.07	-0.002	-0.08
Covariance <i>bf</i>	-0.021	-0.110	-0.014	-0.063	-0.014	-0.067	0.007	0.25	0.007	0.30
Within: total	0.133	0.695	0.157	0.716	0.141	0.661	0.024	0.86	0.009	0.39
Hourly wages <i>wf</i>	0.138	0.723	0.161	0.734	0.139	0.650	0.023	0.82	0.001	0.04
Usual hours <i>wf</i>	0.003	0.016	0.006	0.027	0.006	0.028	0.003	0.10	0.003	0.13
Covariance <i>wf</i>	-0.008	-0.044	-0.010	-0.045	-0.003	-0.016	-0.001	-0.05	0.005	0.22
Unobs. weekly wages:	0.082		0.090				0.008			
Between	0.024	0.277	0.019	0.211			-0.005	-0.63		
Within	0.058	0.713	0.071	0.789			0.013	1.63		

\* Some cells are empty because we only use occupational data for 1996-2010.

\*\* Values give the fraction or share of the overall change in log employee wage variance accounted for by each component.

Note.- *bf* and *wf* refer to the between- and within-firm components of the total variance. "Unobs." wages control for 3-digit occupations (see Section 3.1). Relevant rows may not sum exactly due to rounding.

FIGURE 3: Within- and between-firm components of the variance in log employee wages, 1996-2015



Note.- age 16-64 and full-time employees only. ‘Weekly’ exclude overtime. ‘Annual’ wages are for employees with the firm at least one year. The data is for all large firms in the NESPD who have at least ten full-time employee wage observations in that year. The top one percent of wage values in each year are excluded from calculations here. Shaded areas represent official UK recessions. Lines without markers are estimated linear trends to show the overall change in the sample period.

We also consider three sub-samples of firms, each consisting of approximately a third of the employee observations: the public sector, SIC 2003 sectors G-H (wholesale, retail, hotels, restaurants etc.), and the remainder of the private sector (Online Appendix Figure C7). The overall results are qualitatively robust to looking at the private sector only. The within-firm share of employee weekly wage variance is higher in the public sector and has been trending upwards over the past decade. This is in contrast with the private sector, which has experienced an increasing share of between-firm inequality over this whole period. Both of these trends may in part relate to changes in the composition of UK employment, within or between firms, brought about by the Great Recession and the subsequent period of fiscal austerity, but further investigation of these differences is beyond the scope of this study. Given we consider only full-time employees up to this point, unsurprisingly our results are qualitatively unchanged for hourly wage inequality (Online Appendix Figure C8). Where we can identify firms exactly at the enterprise level, using the ASHE datasets for 2002-15, results are also not qualitatively different (Online Appendix Figure C10).

Similar conclusions also hold for annual wages (Figure 3B; see also Online Appendix Figure C9A). Just over forty percent of the thirteen log points increase in annual wage inequality was accounted for by between-firm variation in Britain over the past two decades (column (10), Online Appendix Table D1). This compares, for example, with the finding by Song et al. (2019) that greater between-firm inequality accounted for sixty percent of the total thirty log point increase in US annual wage inequality for employees in very large firms between 1981 and 2013. The decrease in British employee wage dispersion during the financial crisis was also more pronounced in annual wages, and accounted for mostly by less inequality within firms.

The changes in weekly wage variance discussed here are small in magnitude compared with those measured for annual earnings, among approximately the same sets of firms. We believe that there are three potential reasons for this difference. First, employees with less than a year of tenure in jobs, not represented in the annual earnings decomposition, could have had increasingly similar wages. However, even if this were the case, the sample weight of these employees is just not large enough for it to be a plausible explanation. Barth et al. (2016) found that ignoring job changers’ wages in fact relatively dampens the measured change in US annual wage inequality. Second, though the share of bonuses in total pay was approximately constant over the period, there could have been increasing variance in these payments. If this were an important factor, then we would expect to see greater increases at the top of

the annual wage distribution when compared with the weekly, which we do not see in Figure 2. A third likely explanation, which we cannot precisely identify in this data, is that hours worked in jobs throughout the year have become significantly more variable, especially in low-paying jobs. Although our sample of annual earnings conditions on workers who are full-time in April, their hours worked could fluctuate over the year; we cannot observe the number of weeks worked, nor the hours worked within a week, outside of the reference period in April. Our view is that this explanation is the most likely.<sup>13</sup>

For weekly measures we can identify the determinants of actual wage inequality, as opposed to earnings, by further decomposing the variance estimates into the components which account for the variance in log hourly wage rates,  $\theta$ , log weekly hours worked,  $h$ , and their covariance:

$$V_{wf} = V_{wf}^{\theta} + V_{wf}^h + 2cov_{wf}(\theta, h), \quad (2)$$

$$V_{bf} = V_{bf}^{\theta} + V_{bf}^h + 2cov_{bf}(\theta, h) \quad (3)$$

(see Online Appendix B.1 for exact definitions and derivations of these terms). The covariance terms are potentially large, since both individual and firm average wages and hours are known to be strongly correlated.

Unlike other related studies, we can show explicitly that the variation in hours worked does not affect the decomposition results for full-time employees. Both between- and within-firm hours variance components together account for less than five percent of weekly wage variance throughout the period (column (2), (4), & (6), Table 1). This offers some support to results in other studies which could not directly observe hours, but restricted their attention to full-time employees, such as in Card et al. (2013). When we contrast this with a decomposition of weekly wage variance which includes those working part-time, changes in the variance of hours worked within firms most closely determine overall inequality changes: in the last two decades firms have been increasingly using a mix of part- and full-time employees (see Online Appendix Figure C11, which plots the decomposition described by (2)-(3) for the weekly earnings of full-time and all employees, and Online Appendix Table D2, which summarises this decomposition for all employees). We also see a sharp increase in wage variance among all employees in 2004-05, which coincides with a methodological shift in the survey, whereby more low-paid and part-time jobs without PAYE numbers were sampled. We discuss this further in the Online Appendix, but it is a good reason why we mostly focus on only full-time workers throughout the analysis here. In terms of levels, the combined hours components account for as much as forty percent of overall wage inequality among all employees. The covariance in hours and wages, both within and between firms, is also a significant part, accounting for as much as twenty percent, reflecting the tendency of part-time jobs to be more commonly low-wage.

### 3.1 Observable vs unobservable wage inequality

In Section 2 we described how the baseline sample has changed over time in terms of firm size, and the industry sectors and occupations represented. Further, we cannot be certain that the results are unaffected by changes in how much some characteristics of jobs are rewarded: for example, the recent rise in the London wage premium, which could potentially manifest as greater between-firm inequality.<sup>14</sup> To account

<sup>13</sup>This pattern could also be related to the substantial rise in so-called “Underemployment” in the UK over the same period, documented for example by Bell and Blanchflower (2013).

<sup>14</sup>For example, ONS published results from the ASHE for the nominal median weekly pay of full-time employees show an increase between 1997 and 2007 of forty-five percent in London, compared with thirty-five percent in the North East of England.

for this, we regress log wages in each year of our sub-samples of the NESPD on employee characteristics, and then describe inequality in the resulting unobservable part: i.e. *for each year* between 1996 and 2015, and each measure of wages, we estimate

$$w_{ij} = \mu + \beta \mathbf{x}_{ij} + \underbrace{\alpha_{j(i)} + \varepsilon_{ij}}_{\text{unobs.} - \psi_{ij}}, \quad \text{with } E[\varepsilon_{ij} | \mu, \mathbf{x}_{ij}, \alpha_{j(i)}] = 0 \quad (4)$$

where in separate wage regressions for each year in our sample period we include a minimum set of controls in the vector  $\mathbf{x}_{ij}$  for sex, age and its square, and the region of employment, and  $\beta$  contains coefficients, which can vary by year. What we call the unobservable part of wages is given by  $\psi_{ij}$ . This includes a firm-specific component  $\alpha_{j(i)}$ , where the subscript simply highlights that all workers in some firm  $j$  share a common firm effect. It additionally includes the remaining heterogeneity in wages, which is left in the error term  $\varepsilon_{ij}$ , and which reflects the quality of the employee and firm match. We call this part of the wage “unobservable” because it is the difference between an employee’s actual wage and the part which can be observed, or predicted, without knowing the exact identity of their employer. We use the estimated values  $\hat{\psi}_{ij}$  for each year to study how any additional controls included in  $\mathbf{x}_{ij}$ , in particular for employee occupations, could allow us to more precisely determine the sources of wage inequality trends.

We do not have information on years of education, or some other explicit proxy for levels of human capital. The only way we can mitigate the resulting concern, that this missing information would be correlated with occupation controls, is by considering the robustness of any results while varying the detail of the occupational classification used. For example, in our data, the two-digit minor occupation group “office clerks” contains the three-digit occupations “secretaries and keyboard operating clerks” and “numerical clerks”, which we might expect to typically require different levels of education or qualifications. One alternative interpretation of occupational data is that they approximate the bundles of tasks carried out by workers. Occupational classifications can also be tightly related to collectively bargained wage floors, such as in other European labour markets like Portugal and Spain. However, it is worth noting that this is not the case for any substantial part of the British workforce in our sample period.

Several other studies have estimated firm wage premiums using variants of the two-way worker- and firm-fixed effects model of [Abowd et al. \(1999\)](#) (AKM). However, as discussed at length by [Card et al. \(2018\)](#) and [Borovičková and Shimer \(2017\)](#), estimates obtained from this model are prone to several sources of bias. For example, there is a well-known and typically substantial negative bias in small sample settings on the estimated covariance of worker- and firm-effects, with a coinciding positive bias on the variances of each separate set of effects ([Andrews et al., 2012](#)). We anticipate that these biases would be large using small one percent samples of firm employees, making the interpretation of any results a significant challenge. What we call firm-specific effects here are not comparable to the firm wage premiums identified by job-switching in [Card et al. \(2013\)](#) and [Song et al. \(2019\)](#) among others. Instead estimates  $\hat{\alpha}_j$  should be interpreted as measuring the composition- or covariate-adjusted differences of firm average wages from the overall employee sample mean, in each year. Their variance is surely biased upwards by not addressing unobservable worker heterogeneity (ability) and its distribution across firms. It is also likely that this bias will be partially reduced when occupational controls are added to (4), since their estimated contribution to wages would be correlated with worker ability. Our approach has the advantage of allowing the sample and firm-specific effects to vary each year, which are limitations of the AKM approach. It is also a tractable way to assess the combined roles of the changing observable composition of employment and wage premiums.

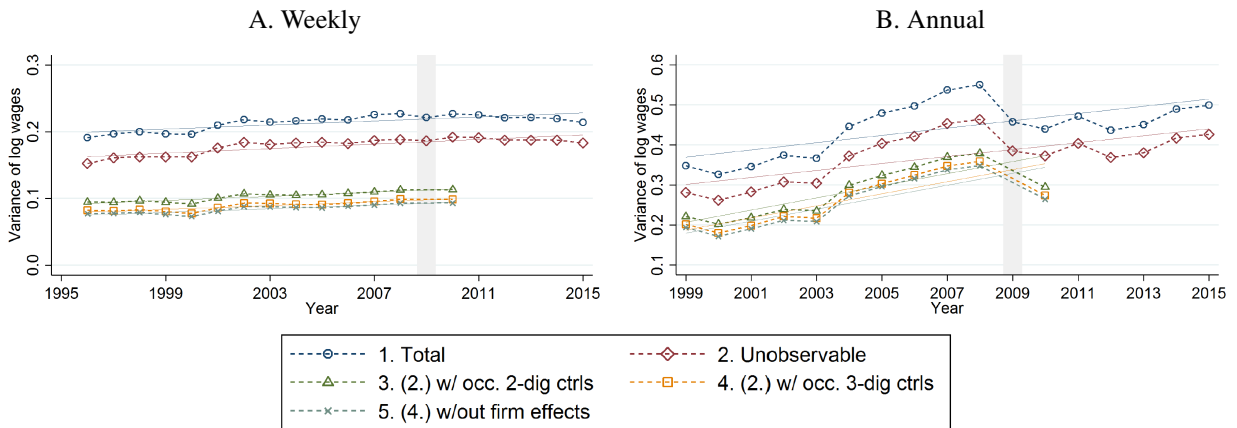


Figure 4 compares the total variance of log weekly and annual wages with the variance of the estimated values for their unobservable parts, using alternative specifications of the yearly wage regressions given by (4). That is, we can use our estimates to write for each year

$$V_e = \text{var}(\widehat{\beta}\mathbf{x}_{ij}) + 2\text{cov}(\widehat{\beta}\mathbf{x}_{ij}, \widehat{\alpha}_j) + \text{var}(\widehat{\psi}_{ij}) . \quad (5)$$

We compare over time  $V_e = \text{var}(w_{ij})$  and  $\text{var}(\widehat{\psi}_{ij})$ , where the difference between these values equals the variance in the estimated observable part of wage heterogeneity plus twice its covariance with the firm-specific effects. The contribution to the level of wage inequality from the covariance term is small, being less than ten percent of the overall amount in all years for both weekly and annual wages when (4) includes controls for occupations. Comparing the variances of actual observed wages and their unobservable part, the patterns over time appear to be mostly unaffected by the inclusion of controls in (4) for regions, age groups and gender (comparing series 2. with 1. in Figure 4). This implies that any dynamic changes in the overall composition of our baseline sample of the NESPD and/or wage premiums for these observable employee characteristics are insignificant. In other words, the changes in British wage inequality over this period are to some extent unobservable changes, occurring within sex, age groups and regions. However, adding controls for the occupational content of wages not only significantly explains a large part of the level, but also decreases the amount of the increase in wage inequality in the decade prior to the Great Recession which is accounted for by greater variance in unobservable heterogeneity (comparing series 3. and 4. with 1. and 2. in Figure 4, or column (7) in Tables 1 & D1). This is increasingly the case when we control for a more detailed group of occupations. Changes in between-occupation inequality and the sorting of workers across occupations are important contributors to total changes in earnings inequality. We also show estimates of the residual variance, excluding firm effects from the estimated wage regressions. Though these are less robust, since they overestimate the role of occupations in the overall level, and potentially in the changes over time, the results show a similar pattern to those which include estimated firm-specific effects.

FIGURE 4: Variance of estimated unobservable (residual) log employee wages, 1996-2015



Note.- age 16-64 and full-time employees only. The top one percent of wage values and residuals in each year are excluded from calculations here. ‘Total’ gives the total variance  $V_e$ . All unobservable log wages are estimated using regressions with controls for sex, age, age squared and major regions. (2.)-(4.) include estimates of firm-specific effects, and respectively (3.) and (4.) add controls for ISCO 2- and 3-digit groups. (5.) is the variance of residuals from an estimation of the wage regression without firm-specific effects. Lines without markers are estimated linear trends.



We can also account for the role of between-firm differences in the unobservable part of wage inequality changes, by replacing the values and statistics for  $w$  with estimates  $\hat{\psi}$  in (1). Online Appendix Figure C12 shows that, after conditioning employee wages on 3-digit occupation groups, the share in the overall variance level of the within-firm component,  $\text{var}(\hat{\epsilon}_{ij})/\text{var}(\hat{\psi}_{ij})$ , is increased on average from two-thirds to three-quarters for weekly wages, and from three-quarters to four-fifths for annual earnings (comparing series 4. with 1. and 2., where 1. gives  $V_{wf}/V_e$ ). Unsurprisingly, a part of the difference in average wages across firms is accounted for by the types of workers they employ. More importantly for our focus on trends, the share of unobservable wage variance which is within-firm is increasing over time, relative to the equivalent share of the observed total wage variance, after addressing the role of changing occupation-specific wage premiums and their composition in our sample. This suggests that changes to the observed between-firm component,  $V_{bf}$ , represent changes to between-occupation inequality and the distribution of occupations across firms.

## 4 Inequality changes throughout the wage distribution

In analysing the dynamics and components of an aggregate measure of wage inequality, we could be neglecting a more complex evolution of the cross-sectional wage distribution. To determine the role of firms in changes across and within the distribution of wages, we employ a graphical method popularised by Juhn et al. (1993), and subsequently adapted by Song et al. (2019) among others. We can write employee log wages as

$$\underbrace{w_{ij}}_{\text{Employees}} = \underbrace{\bar{w}_j}_{\text{Firms}} + \underbrace{[w_{ij} - \bar{w}_j]}_{\text{Employee/firm}}. \quad (6)$$

We then compute estimates of the averages of each term in (6) within each percentile bin of the employee wage distribution in every year. By considering the resulting differences across percentiles and between years, we can then account for the role of firm average wages, as opposed to the relative difference between employees' wages and their firms' averages, in driving wage inequality changes. We also report a heuristic measure of what moves the wage distribution, which captures the share of the variance across percentiles in average wage changes accounted for by covariance with changes in the 'Employee/firm' (within-firm) component:

$$\gamma = \frac{\text{cov}(\Delta\{\overline{w_{ij} - \bar{w}_j}\}_p, \Delta\bar{w}_p)}{\text{var}(\Delta\bar{w}_p)}, \quad (7)$$

where  $\Delta\bar{w}_p$  gives the change between two periods in the sample average log employee wage in percentile bin  $p$ .

### 4.1 Actual employee wages

Figure 5 represents this graphical decomposition for the change in real log weekly wages before the Great Recession, between 1997 and 2007, using the baseline sub-sample of full-time employees. The relatively smooth 'Employees' series plots the change in the average log wage of workers in each percentile between the two years. To avoid confusion, these are unlikely to be the same individuals: this is a comparison of annual cross-sections. Each percentile is decomposed using around four to five hundred job observations in each year. A positive slope across percentiles indicates that in some portion of the wage distribution inequality has increased. For example, wages at the median increased by 5 log points (5%) over this period, but by 10 points (11%) at the seventy-fifth percentile, and 20 points (22%) at the ninety-fifth. Representing

the evolution of wage inequality in this way shows that small changes in the time series of overall log wage variance can belie starker inequality dynamics. For instance, here we see that inequality fell among the very lowest earners, potentially due to the introduction of the National Minimum Wage in 1999. By construction, the average level of the ‘Firms’ (between-firm) components across percentiles is the same as that for ‘Employees’, and the ‘Employee/firm’ (within-firm) component is centred about zero. For the graphical analysis it is the slopes of these series across the percentiles which concern us. The between-firm component contributes somewhat to the rise in wage inequality at the top of the wage distribution, but the within-firm component also contributes, increasing across the percentiles from the twentieth onwards. The within-firm change accounts for over two-thirds of the overall movement at percentiles across the distribution:  $\gamma = 0.68$ .

FIGURE 5: Change in the average real log weekly wage by percentile of employees, and the contribution from firms



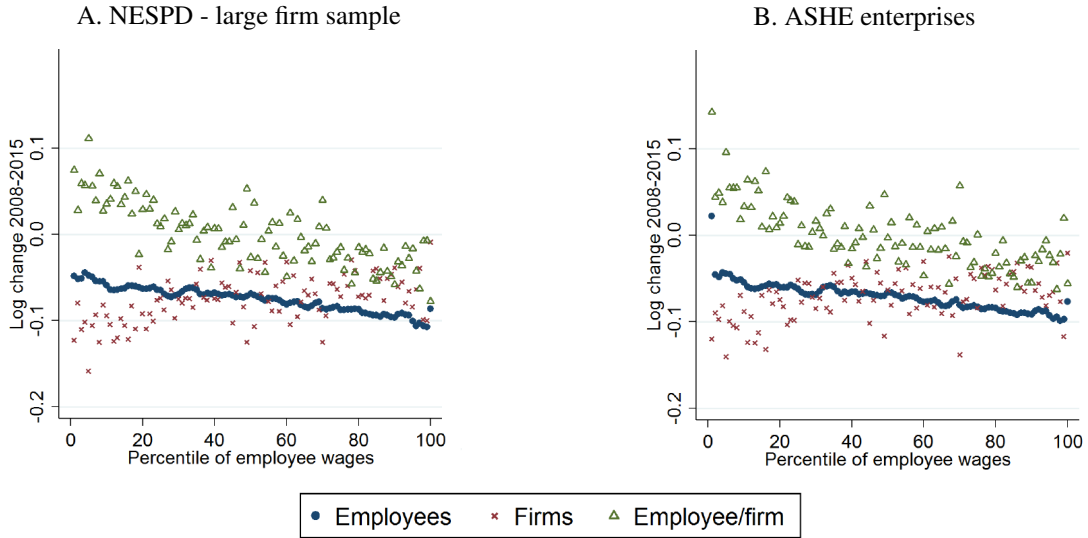
Note.- age 16-64, full-time employees only, excluding overtime. The ‘Employees’ values are computed by taking the average log real wages of employees within each percentile, increasingly ordered by the level of wages in each year, and taking the difference between years. The ‘Firms’ values are computed by taking the average across workers, in each percentile, of the average log wages of the firms they work for, in each year, and then taking the difference across years. The ‘Employees/Firms’ values are the residual difference between these other two lots: equivalently, the average across workers, in each percentile, of the log difference in employee wages from their firms’ average value, in each year, and taking the difference across years.  $\gamma = 0.68$ .

For this graphical decomposition we retain the top one percent of earners in the sample. The very top of the income and wage distribution has drawn significant attention recently, especially in the US (see [Piketty, 2014](#); [Song et al., 2019](#)). Although based on a small sample of these top earners in Britain, we can see from [Figure 5](#) that average weekly wages in the top one percent for 1997-2007 did not experience greater relative increases than those in the top decile. As discussed before, the April reference period of these data is unlikely to capture the role of bonus payments in weekly wage inequality changes. When we turn to annual wages later, we see in fact that the wages of the top one percent increased approximately twenty log points more than for employees at the ninety-fifth percentile over the same period ([Online Appendix Figure C20A](#)). Over eighty percent of the log weekly wage increase for the top one percent occurred within firms, notably higher than for all other percentiles besides the bottom one percent.

Before progressing further, we also represent the change in inequality since the Great Recession, for 2008-2015, in the same way ([Figure 6](#)). We can also demonstrate here that our results are unaffected if we instead use the administrative definition of an enterprise from the ASHE datasets to define firms. Real wages

decreased across the whole distribution since the financial crisis, but inequality also fell. However, there is no suggestion in the data that this can be accounted for by changes in the differences in average wages between firms. In fact, the within-firm component more than explains the changes across the distribution:  $\gamma > 1$ .

FIGURE 6: Change 2008-2015 in the average real log weekly wage by percentile of employees and the contribution from firms: NESPD large firms sample vs ASHE large enterprises



Note.- see Figure 5 and Online Appendix Figure C10: for NESPD  $\gamma = 1.87$ , for ASHE  $\gamma = 1.72$ .

We consider how consistent these results are across the whole time period. Still focusing on full-time weekly wages, Figure 7A plots the changes for selected quantiles of the employee weekly wage distribution relative to 1996 (i.e. we now decompose changes in average employee wages in five percentile bins between 1996 and all subsequent years). The average wages paid by firms can account for some of the relatively greater increase in the top five percent of employee wages in the early 2000s (Figure 7B). But changes to wage dispersion within firms explain the vast majority of the inequality dynamics throughout the employee wage distribution over the last two decades (Figure 7C).

## 4.2 Unobservable wage heterogeneity

Figure 8 shows an equivalent decomposition for 1997-2007 as Figure 5, but only for the estimated unobservable part of wages  $\hat{\psi}_{ij}$ , including firm-specific effects, and controlling for occupation groups:

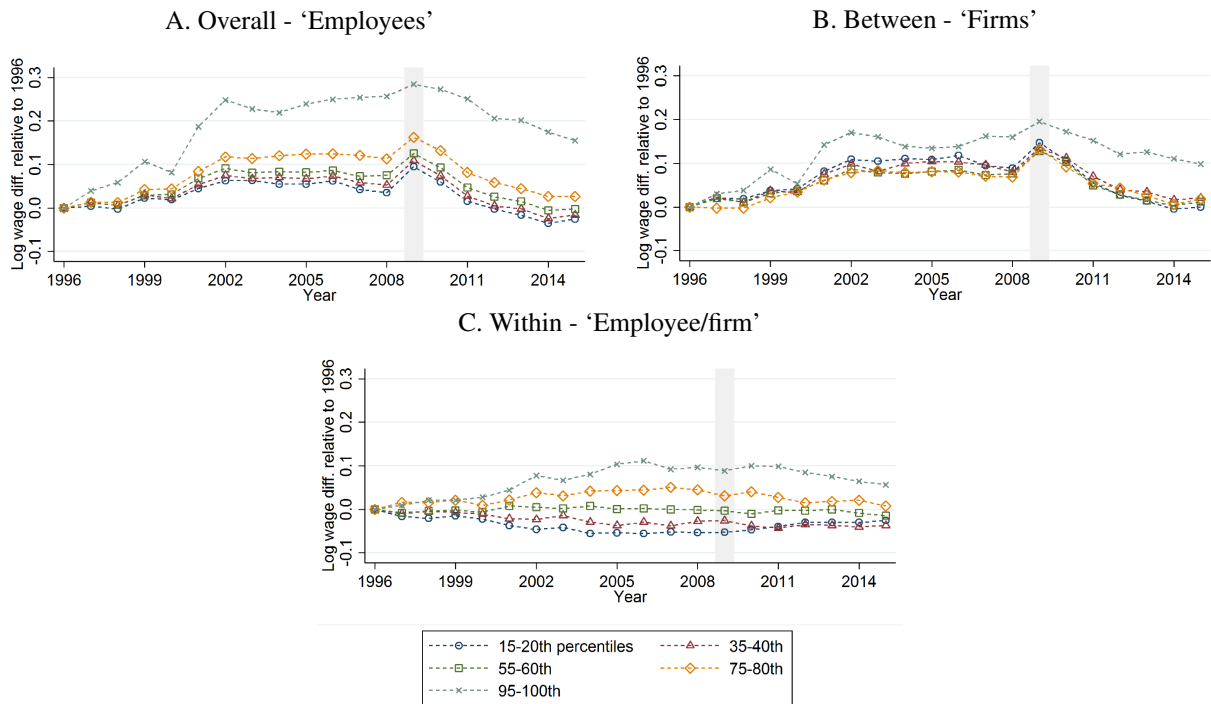
$$\underbrace{\hat{\psi}_{ij}}_{\text{Employees}} = \underbrace{\bar{\hat{\psi}}_j}_{\text{Firms}} + \underbrace{[\hat{\psi}_{ij} - \bar{\hat{\psi}}_j]}_{\text{Employee/firm}}, \quad (8)$$

where  $\bar{\hat{\psi}}_j = \hat{\alpha}_j$  and  $[\hat{\psi}_{ij} - \bar{\hat{\psi}}_j] = \hat{\epsilon}_{ij}$ .

The pattern of unobservable employee wage changes across percentiles noticeably differs, comparing with Figure 5. There is rising inequality across the distribution, with the slope becoming steeper from the eightieth percentile upwards. The estimates of firm-specific wages did not account for these dynamics: neither for rises below the eightieth percentile, nor the greater increase in the highest wages:  $\gamma = 1.00$ .<sup>15</sup> In Online Appendix Figure C13 we consider alternative estimates of the unobservable part of wages. In

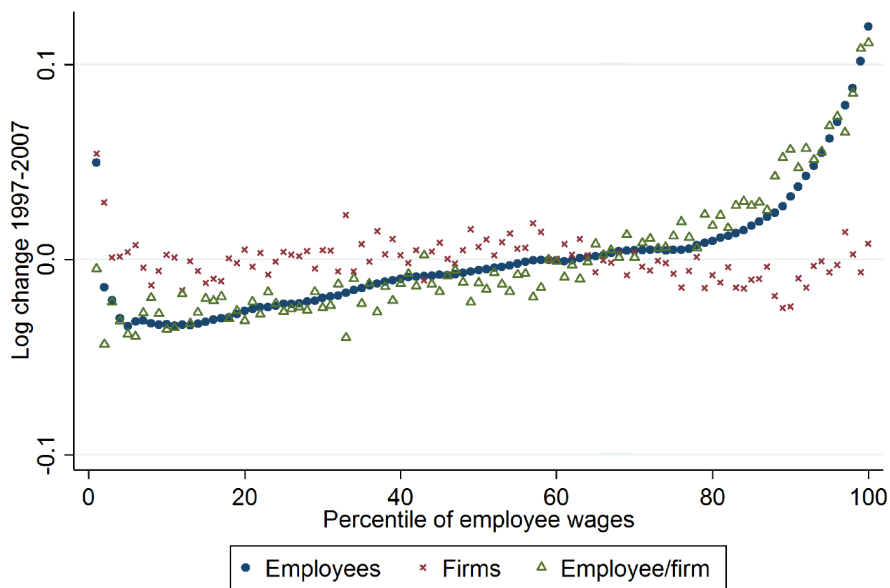
<sup>15</sup> $\gamma$ , with the inclusion of firm-specific effects, is measured by  $cov(\Delta\bar{\hat{\epsilon}}_p, \Delta\bar{\hat{\psi}}_p)/var(\Delta\bar{\hat{\psi}}_p)$ .

FIGURE 7: Average real log weekly wage of employees in selected quantiles, relative to 1996, and contributions from firms



Note.- see Figure 5.

FIGURE 8: Change 1997-2007 in the average real unobservable log weekly wage by percentile of employees, and the contribution from firms



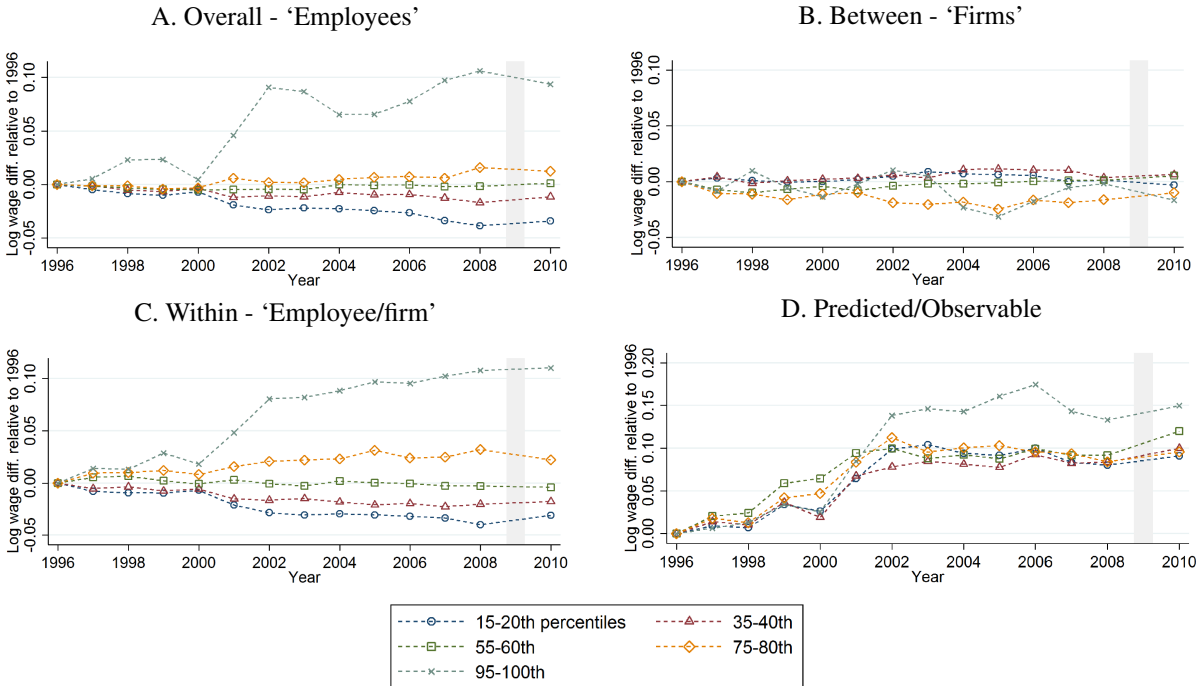
Note.- Unobservable log wages are estimated using a regression with controls for sex, age, age squared, major regions, ISCO 3-digit groups and firm-specific effects.  $\gamma = 1.00$ . See notes for Figure 5 or the text for a description of how series are calculated and interpretation.

panel A, we can see that other controls, including the firm-specific effects, are not driving this result. In panel B, controlling for less detailed occupation groups still relatively reduces the estimated role for firm-level differences. In panels C and D we show that not including firm-specific effects would lead us to overestimate the role of occupational inequality. These results allow us to conclude that any change in the differences in average wages between large firms is mostly due to some combination of changes to

between-occupation inequality and the concentration of relatively high- or low-wage occupations within firms. Further, given that this result does not qualitatively depend on firm-specific controls, it is possible that results in other studies, which assign some of the importance of changes in the between-firm component to industrial change, are to an extent misrepresenting a more significant role of occupations.

Figure 9 replicates Figure 7 but instead for unobservable wages. There is no substantial contribution from between-firm inequality (Figure 9B) to the dynamics of the unobservable part of the wage distribution since 1996 (Figure 9A). Figure 9D also shows the contribution from the observed part of the real wage distribution,  $\Delta\{\hat{\mu} + \hat{\beta}'\mathbf{x}_{ij}\}_p$ . Here we can see explicitly how wage premiums and the prevalence of high paying occupations in our sample account for the rise in inequality at the top of the employee wage distribution. Online Appendix Figure C14 further demonstrates the robustness of this result across all percentiles, considering changes over other ten-year periods, each beginning in a year between 1996-2000.

FIGURE 9: Average real (un)observable log weekly wage of employees in selected quantiles, relative to 1996, and contributions from firms



Note.- see Figure 8.

### 4.3 Additional results and robustness checks

So far in this section we have only discussed the dynamics of weekly wages for full-time employees working for firms with at least ten job observations in the NESPD in any given year. We can also check whether results change, based on the pre-recession period 1997-2007, when we alter these aspects of the sample. Online Appendix Figures C15-C17 decompose the log change in the weekly wages of full-time employees who are employed by large enterprises with at least one, five or twenty employee job observations. For actual wages, as we increase the sample size and include some smaller firms, it becomes clear graphically that the firms component cannot explain inequality dynamics. Considering the unobservable part of wages, with controls for occupations, the results are also qualitatively unchanged as we vary the average firm size and how much of UK employment our sample of employees represents.

In Online Appendix Figure C18 we return to our baseline sample, but now study only private sector employees. Again the results are unaffected. Further, Online Appendix Figure C19 shows that there is no qualitative difference in results if we decompose hourly wage dynamics as opposed to weekly. For annual earnings, Online Appendix Figure C20 demonstrates that for actual wage inequality the majority of the dynamics across percentiles are explained by the changing picture within firms. This is also the case when we turn to the increased inequality in unobservable annual wages. We also see that the greater increase in the variance of annual wages over this period, compared with weekly wages, is mostly explained by a rise in inequality among employees with the lowest annual earnings. Since this description only includes workers who stayed in the same job for at least a year, this most likely reflects a large increase in the week-to-week volatility of working hours among the lowest paid employees and jobs in the British labour market. Finally, we also consider the picture for weekly wages including part-time employees, and after conditioning on employee characteristics, there is no suggestion in Online Appendix Figure C21 that firm average wages have driven inequality dynamics in this case also.

Returning to weekly wages and our baseline sample, Online Appendix Figure C22 looks at changes across all percentiles for five-year sub-periods. Notably for robustness, these are periods where the classification of occupations used in the NESPD is constant, and thus cross-walking was not necessary. The majority of recent increases in employee wage inequality occurred in the five years to 2001. There is no contribution to this from the firms component for unobservable log wages. This is also the case for actual wage inequality, apart from some contribution to greater changes above the ninety-fifth percentile. For the periods 2002-07 and 2005-10, the rise in wage inequality is small, and is driven by greater wage changes for only the highest earners. But in both later sub-periods, once we account for the observable content of wages, the role of firm average wage differences is reduced.

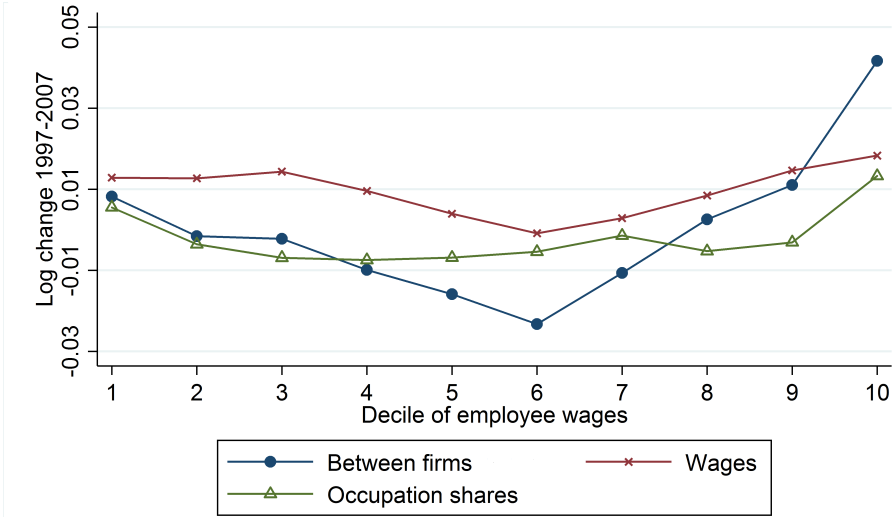
To expand on these findings further, we focus on the ‘Firms’ (between-firm) component of the change between 1997 and 2007 in unobservable weekly wages for full-time employees, represented by Online Appendix Figure C13A: i.e. the estimated firm-specific effects, with controls for some employee observables but not occupations in the log wage regressions. Averaging these across employee wage percentiles, we carry out a shift-share decomposition. This allows us to account for the role of the changing occupational structure of the firms represented in each decile, in addition to changes in the wages paid by the firms with relatively high or low paid employees for specific occupations (see Online Appendix B.2 for details). Figure 10 shows the results of this decomposition. The ‘Wages’ component for each decile is computed as follows: (1) identify the firms for which the employees in that decile are working, separately for 1997 and 2007, (2) compute average occupational wages across those firms, separately for 1997 and 2007, (3) keep the occupational shares constant at the observed level in those firms in 1997, (4) compute the average over time of the occupational wages obtained in step (2) and multiply with the shares from step (3). The ‘Occupation shares’ component is obtained in a similar way, the only difference being that we average over time the occupational shares and hold occupational wages constant at their 1997 levels. Within each decile, the difference between the sum of the ‘Wages’ and the ‘Occupation shares’ components, on the one hand, and the ‘Between firms’ term, on the other hand, is given by changes in covariance terms (not displayed).

Assuming that the occupational composition of the firms represented among the employees in each decile had not changed between 1997 and 2007 (cross markers), Figure 10 suggests that between-firm inequality would have increased less through the effect of changes only in occupational wages than



estimated overall. The changes to the occupational composition of firms (diamond markers), keeping occupational wages fixed, can account for the majority of how firm-specific wages contributed to employee earnings changes in the bottom and top deciles. In other words, the increased between-firm contribution to inequality at the top of the earnings distribution was mostly explained by firms who employed the highest paid workers increasingly focusing on high-paying occupations.

FIGURE 10: Decomposing the firm component of employee real wage inequality patterns, 1997-2007: the role of changing firm occupation shares vs wages



Note.- this figure takes the average over deciles of the firm component of wage changes in Online Appendix Figure C13A (circle markers), and carries out a shift-share decomposition. The two components thereof reported here are as follows: first holding the average across employees of firm occupation shares constant, and considering only average occupational wage changes (cross markers), and second holding the average across employees of firm average occupational wages, in a decile, constant, but varying only firm occupation shares (diamond markers).

### 5 Summary and further discussion

We have used well-known methods to answer whether recent trends in British wage inequality, viewed through a sample of employees at mostly very large firms, can be accounted for by between-firm inequality. We have found substantial evidence that in the last two decades this has not been the case. This is also clear when we consider estimates of unobservable wage heterogeneity, controlling for changes to occupational premiums and firm-specific effects. At first look, this would appear to contradict what is becoming a stylised fact, across several countries, that between-firm wage inequality is the most important driver of overall trends. But this is not the first paper to suggest that some part could be accounted for by the changing supply and demand of occupations across firms and labour markets (see Card et al., 2013; Song et al., 2019). We further believe our results can be reconciled with some of these previous studies. The analysis here is dominated by the very largest firms in Britain. Already Song et al. (2019) have shown that in the US firm size matters. Larger firms come from the starting point of having more diverse workforces and complex pay structures, and so there is far more scope for changes over time in the dispersion of wages within as they evolve. Second, we believe our results chime with another hypothesis from Song et al. (2019): the reason why within-firm inequality cannot account for overall dynamics, in most studies, could be due to the increasing occupational concentration, or specialisation, of firms, coinciding with falling costs of outsourcing work tasks, and their greater tendency to focus on so-called ‘core-competencies.’ The very large and long-lived firms, which dominate our sample, are where we might expect such changes



in specialisation to mostly occur. Adding to this the continued trend of increasingly polarised demand for occupations in the British labour market (Goos and Manning, 2007; Williams, 2013), it is then not surprising that once we focus on the inequality dynamics of the estimated unobservable part of wages, with controls for changing occupational premiums and the composition of the workforce, the role of firm-specific differences becomes markedly weaker, or even non-existent.

The results here suggest that future analyses of this kind should attempt to seek out data which can address the possible role of the changing occupational structure of firms. Otherwise it could be challenging to identify whether inequality changes are accounted for by some greater segregation of workers across firms, or whether this to some extent reflects only the combined effects of changes to the occupational polarisation of employment and firm-level specialisation. In other words, the estimated role of assortative matching over innate firm and worker productivities could be overstated, if firms simultaneously alter their demand for occupations and skills.

The results here only reflect what has happened for wages in mostly very large firms. However, it is well known that there is a substantial wage premium paid by large firms on average, and most of this is not accounted for by the selection of different worker types over firm sizes (see for a summary Oi and Idson, 1999). There is evidence from the US that this wage premium has shrunk markedly in recent decades, particularly driven by a decrease in the part accounted for by large firms paying more than small firms for otherwise identical workers (Bloom et al., 2018). If there have been similar changes in this wage premium in Britain during the period we study, then it would be consistent with us having overestimated the role of between-firm wage inequality increases in our sample, which is dominated by large firms, when compared with whatever the reality is in the overall economy.

A limitation of the analysis here is that we are restricted to studying repeated cross-sectional data of jobs and wages, since representative samples of employers cannot be identified reliably across time for an extended period in the NESPD. We believe this is the limit of what can be achieved using currently available British data sources in terms of measuring the general role for firms in aggregate wage inequality dynamics, without small sample biases totally confounding any analysis. There are however some avenues for further research. Focusing on the largest British firms, it is possible to link these samples of their employees' current and historical wage data, as well as job characteristics, with more detailed firm characteristics, such as profits, value added, labour productivity, export and ownership status, and levels and types of investment activity. This information could allow a rich analysis of what factors relate to changes in the occupational specialisation of firms and the levels of wage inequality within them. However, to go even further, we hope that existing UK administrative earnings data, for all employees and their employers, will be made available for research in the near future. This would allow as comparably detailed and enlightening analysis as Song et al. (2019) have recently carried out for the US. Only then can the continuing large evidence gap regarding the determinants of British wage inequality be more completely addressed, with the NESPD's more detailed records of job characteristics, such as hours and occupations, serving as a useful supplementary data source.

## References

- Abowd, J., F. Kramarz, and D. Margolis.** 1999. "High wage workers and high wage firms." *Econometrica*, 67(2): 251–334.
- Akerman, A., E. Helpman, O. Itskhoki, M.-A. Muendler, and S. Redding.** 2013. "Sources of Wage Inequality." *American Economic Review: Papers & Proceedings*, 103(3): 214–219.
- Alvarez, J., F. Benguria, N. Engbom, and C. Moser.** 2018. "Firms and the Decline of Earnings Inequality in Brazil." *American Economic Journal: Macroeconomics*, 10(1): 1–43.
- Andrews, M., L. Gill, T. Schank, and R. Upward.** 2012. "High wage workers match with high wage firms: Clear evidence of the effects of limited mobility bias." *Economics Letters*, 117(3): 824–827.
- Autor, D. H., L. F. Katz, and M. S. Kearney.** 2008. "Trends in U.S. Wage Inequality: Revising the Revisionists." *The Review of Economics and Statistics*, 90(2): 300–323.
- Barth, E., A. Bryson, J. C. Davis, and R. Freeman.** 2016. "It's Where You Work: Increases in the Dispersion of Earnings across Establishments and Individuals in the United States." *Journal of Labor Economics*, 34(S2): S67–S97.
- Belfield, C., R. Blundell, J. Cribb, A. Hood, and R. Joyce.** 2017. "Two Decades of Income Inequality in Britain: The Role of Wages, Household Earnings and Redistribution." *Economica*, 84(334): 157–179.
- Bell, D. N., and D. G. Blanchflower.** 2013. "Underemployment in the UK Revisited." *National Institute Economic Review*, 224(1): F8–F22.
- Bloom, N., F. Guvenen, B. S. Smith, J. Song, and T. von Wachter.** 2018. "The Disappearing Large-Firm Wage Premium." *AEA Papers and Proceedings*, 108 317–22.
- Borovičková, K., and R. Shimer.** 2017. "High Wage Workers Work for High Wage Firms." NBER Working Papers 24074, National Bureau of Economic Research, Inc.
- Card, D., A. R. Cardoso, J. Heining, and P. Kline.** 2018. "Firms and Labor Market Inequality: Evidence and Some Theory." *Journal of Labor Economics*, 36(S1): S13 – S70.
- Card, D., J. Heining, and P. Kline.** 2013. "Workplace Heterogeneity and the Rise of West German Wage Inequality." *The Quarterly Journal of Economics*, 128(3): 967–1015.
- Davis, S. J., and J. Haltiwanger.** 1991. "Wage Dispersion between and within U.S. Manufacturing Plants, 1963-86." *Brookings Papers on Economic Activity. Microeconomics*, 1991 115–200.
- Dunne, T., L. Foster, J. Haltiwanger, and K. R. Troske.** 2004. "Wage and Productivity Dispersion in United States Manufacturing: The Role of Computer Investment." *Journal of Labor Economics*, 22(2): 397–429.
- Faggio, G., K. G. Salvanes, and J. V. Reenen.** 2010. "The evolution of inequality in productivity and wages: Panel data evidence." *Industrial and Corporate Change*, 19(6): 1919–1951.
- Gabaix, X., and A. Landier.** 2008. "Why has CEO pay increased so much?" *The Quarterly Journal of Economics*, 123(1): 49–100.
- Goos, M., and A. Manning.** 2007. "Lousy and Lovely Jobs: The Rising Polarization of Work in Britain." *The Review of Economics and Statistics*, 89(1): 118–133.
- Gromadzki, J., I. Magda, and S. Moriconi.** 2019. "Firms and wage inequality in Central and Eastern Europe ." IBS Working Papers 03/2019, Instytut Badan Strukturalnych.
- Helpman, E., O. Itskhoki, M.-A. Muendler, and S. J. Redding.** 2017. "Trade and Inequality: From Theory to Estimation." *The Review of Economic Studies*, 84(1): 357–405.

- Hills, J., M. Brewer, S. P. Jenkins, R. Lister, R. Lupton, S. Machin, C. Mills, T. Modood, T. Rees, and S. Riddell.** 2010. “An anatomy of economic inequality in the UK: Report of the National Equality Panel.” LSE Research Online Documents on Economics 28344, London School of Economics and Political Science, LSE Library.
- Juhn, C., K. Murphy, and B. Pierce.** 1993. “Wage Inequality and the Rise in Returns to Skill.” *Journal of Political Economy*, 101(3): 410–42.
- Machin, S.** 2011. “Changes in UK Wage Inequality Over the Last Forty Years.” In *The Labour Market in Winter*. Eds. by P. Gregg, and J. Wadsworth Oxford University Press.
- Mueller, H. M., P. P. Ouimet, and E. Simintzi.** 2017. “Wage Inequality and Firm Growth.” *American Economic Review*, 107(5): 379–83.
- Nickell, S., and G. Quintini.** 2003. “Nominal Wage Rigidity and the Rate of Inflation.” *Economic Journal*, 113(490): 762–781.
- Nordström Skans, O., E. Per-Anders, and B. Holmlund.** 2009. “Wage Dispersion Between and Within Plants: Sweden 1985-2000.” In *The Structure of Wages: An International Comparison*. Eds. by E. P. Lazaer, and K. L. Shaw University of Chicago Press, 217–260.
- Oi, W., and T. Idson.** 1999. “Firm size and wages.” In *Handbook of Labor Economics*. Eds. by O. Ashenfelter, and D. Card, 3, Part B Elsevier, , 1st edition, Chap. 33 2165–2214.
- ONS.** 2016a. “Annual Survey of Hours and Earnings, 1997-2015: Secure Access. [data collection].” 8th Edition SN: 6689, Office for National Statistics, UK Data Service.
- ONS.** 2016b. “New Earnings Survey Panel Dataset, 1975-2015: Secure Access. [data collection].” 6th Edition SN: 6706, Office for National Statistics, UK Data Service.
- Piketty, T.** 2014. *Capital in the Twenty-first Century*. Harvard University Press.
- Sen, A.** 1973. *On Economic Inequality*. Oxford University Press.
- Song, J., D. J. Price, F. Guvenen, N. Bloom, and T. von Wachter.** 2019. “Firming Up Inequality.” *The Quarterly Journal of Economics*, 134(1): 1–50.
- Weber-Handwerker, E., and J. R. Spletzer.** 2016. “The Role of Establishments and the Concentration of Occupations in Wage Inequality.” In *Inequality: Causes and Consequences (Research in Labor Economics)*. Eds. by L. Cappellari, S. W. Polachek, and K. Tatsiramos, 43 Emerald Group Publishing Limited, 167–193.
- Williams, M.** 2013. “Occupations and British Wage Inequality, 1970s–2000s.” *European Sociological Review*, 29(4): 841–857.

# Recent changes in British wage inequality: Evidence from large firms and occupations

## Online Appendix

Daniel Schaefer<sup>†</sup> Carl Singleton<sup>‡</sup>

2019

### Appendix A. Further description of the data and sample construction

In what follows we give some additional details regarding the datasets used, and how we have constructed the sub-samples thereof. All of the relevant documentation and variable descriptions attached to these datasets are publicly available from the UK Data Service. The ONS has also published various documents concerning the data quality and consistency of the NESPD and ASHE. We will publish our replication files for the analysis and sample construction.

We focus on methodological details through the period 1996-2015. From 1975 to 2003, under its guise as the NESPD, very little changed in the methodology and construction of the longitudinal panel dataset. Throughout this period, it should be a true random sample of all employees in employment, irrespective of employment type, occupation, size of employer etc. Given the legal obligation of employers to respond, and their use of payrolls, it has a very high response rate and is believed to be accurate. There is no cumulative attrition from the panel, as any individual not included in the NESPD in any year, for whatever reason, remains in the sampling frame the following year. Conditional on a hundred percent response, the NESPD would be a true one percent random sample of employees. However there are two major sources of undersampling, both occurring if individuals do not have a current tax record. This could occur for some individuals who have recently moved job, or for those who earn very little (mostly part-time), and so do not have to pay tax or National Insurance. From 2004 the ASHE replaced the NESPD. This aimed to sample some of those employees under-represented in the NESPD. It added supplementary responses for those without a PAYE reference, and also attempted to represent employees whose jobs changed between the determination of the sampling frame in January and the reference period in April. Since the ONS states that the bias these amendments were introduced to address were actually small, we do not believe they could affect our results substantially. The ASHE also introduced some imputations, using similar matched ‘donor’ observations where responses were, for example, missing an entry of basic hours but had recorded pay. These imputations were added for weighting purposes. We ignore these weights throughout our analysis because they are based only on employee data (age, sex, occupation and region of work place) to match population estimates obtained from the Labour Force Survey. Therefore using them would no longer allow us to claim that the ASHE and our results are based on random samples of employees within firms. From 2005, a new questionnaire was also created which was intended to reduce the latitude for respondents’ own interpretations of what was being asked of them. From 2007 there were further notable changes. Beforehand, occupations were classified as follows: either, if the respondent stated an employee’s job had not changed in the past year, the previous year’s occupational classification was applied. Otherwise, it was manually coded. Afterwards an automatic coding, text recognition, tool was used. “The effect of using ACTR was to code more jobs into higher paying occupations. The jobs that tended to be recoded into these higher paying occupations generally had lower levels of pay than the jobs already coded to those occupations. Conversely, they tended to have higher levels of pay than the other jobs in the occupations that they were recoded out of. The impact of this was to lower the average pay of both the occupation group that they had moved from and that they had moved to.” As such, this would certainly increase within

---

<sup>†</sup>University of Edinburgh, [daniel.schaefer@ed.ac.uk](mailto:daniel.schaefer@ed.ac.uk)

<sup>‡</sup>University of Reading, [c.a.singleton@reading.ac.uk](mailto:c.a.singleton@reading.ac.uk); with thanks to the Economic and Social Research Council (UK) for funding support under Grant No. ES/J500136/1, in addition to the Scottish Economic Society.

occupation wage inequality for the highest earners, and reduce it for the lowest earners. Nonetheless, we do not believe this is significant in affecting our results. In the main text, we focus the graphical analysis on changes across the period 1997-2007, but also find our results are unchanged for the periods 1996-2006 and 1996-2001. From 2007, the sample size of the ASHE was reduced by twenty percent, with reductions targeted on those industries that exhibit the least variation in their earnings patterns. However, we do not believe this could have affected our results substantially.

To construct the sub-samples from the panel dataset for 1975-2015, for the analysis of hourly or weekly pay, we first drop a few cases of duplicates over all variables. Then, using the panel identifier, year, the information from the IDBR concerning enterprise status and number of employees, industry classification and gross weekly pay including overtime, we also drop some cases which are then determined to be the same job. We do not drop observations where an individual has multiple jobs. We keep only observations for individuals aged 16-64, and which have not been marked as having a loss of pay in the reference period through absence, employment starting in the period, or short-time working, and which are marked as being on an adult rate of pay (i.e. dropping trainees and apprenticeships). This is practically the same filter applied for ONS published results using the NESPD or ASHE. We also drop all observations with zero or missing values for basic hours, and hourly or weekly pay excluding any overtime. Basic hours are intended to be a record for the employee in a normal week, excluding overtime and meal breaks. Gross weekly pay is the main recorded value in the survey, and from this overtime records are then simply subtracted. Hourly rates are then derived from dividing by basic hours worked. We drop observations with over a hundred basic hours worked, as these could reflect measurement error and inclusion of overtime. Full-time is defined as working over thirty basic hours in a week. But there are a tiny number of discrepancies in some years, we believe relating to teaching contracts, where the definition applied by the ONS differs. We however recode these such that for all observations the thirty hours threshold applies. To further address some potential for measurement error, especially in the recorded basic hours, we drop observations whose hourly rate of pay excluding overtime is less than eighty percent of the National Minimum Wage (NMW) which applies each April, with allowance for the different age-dependent rates of the NMW over time. We set the threshold lower to avoid dropping observations where employers have rounded figures about the NMW, where the degree of rounding could vary with the actual value of the NMW, a behaviour which has been hypothesised by the ONS. To then construct the large firms sample, we drop all employers whose exact enterprise reference number of employees from the IDBR, which is only available from 1996 onwards, is less than 250. We also drop observations where the IDBR status, number of employees or industry classification is missing. We then identify each employer in the dataset using the combination of their 5-digit industry code, IDBR status and exact number of IDBR enterprise employees, within each year. For large firms we are confident this can uniquely identify the reporting organisation of the NESPD. The large firms samples we subsequently analyse then condition on there being a minimum number of remaining job observations per firm in a year. For annual pay, we construct the large firms samples in the same way, except we additionally filter out observations where the employee is reported to not have been in the same job for twelve months, and drop observations with zero or missing values of annual gross pay in place of hours or weekly pay. When handling the ASHE annual cross-section datasets we use the exact same approach, except here there is a unique enterprise level identifier which we can use to identify the firms within each year.

For 1996-2001, occupations are classified using the 3-digit ONS 1990 Standard Occupational Classification (SOC). For 2002-2010, this is replaced with the 4-digit SOC 2000, and for 2011-2015, with the SOC 2010. We experimented using the ONS' publicly available cross-walk from 2010 and 2000 to 1990 classification, but discovered that this causes a large structural break in the distribution of occupations. In particular, it causes a substantial additional degree of polarisation of work from 2002 onwards, which would potentially generate erroneous and large increases in within occupation inequality around this date. To address this we rely on a conversion of SOC 1990 and 2000 to the 1988 International Standard Classification of Occupations (ISCO). We obtain these conversions from the Cambridge Social Interaction and Stratification Scale (CAMSIS) project. For the industry classification, we convert ONS Standard Industrial Classification (SIC) 2007 to 2003, using files made available by the UK Data Service.

This conversion uses the 2008 Annual Respondents Dataset where both classifications were applied, and where any 2007 code mapping to multiple 2003 codes is decided using whichever of the two bore a greater share of economic output. For 1996-2002, the work region of the employee is missing, and so we derive this ourselves consistent with the ONS geo-maps, using the more detailed work area variable.

TABLE A1: Comparison of the baseline sample's firm size distribution, and represented employees, with the UK population of enterprises, 2013

Enterprise size <sup>†</sup>	Number of obs.		Total employees in enterprises (000s) <sup>†</sup>	
	Sample firms <sup>‡</sup>	UK enterprises	Sample firms	UK enterprises
250 - 999	92	6,400	69	2,927
1,000 - 1,999	308	1,050	469	1,455
2,000 - 4,999	644	830	2,098	2,612
5,000+	596	635	8,204	8,805
Total	1,640	8,915	10,804	15,799

<sup>†</sup> Values for sample firms use the IDBR record of the number of employees in the enterprise, which includes the firm. This is not the number of observations of employee jobs in the sample.

<sup>‡</sup> All firms in the baseline sample with a minimum of ten full-time employee observations in the NESPD in 2013, and subject to the other sampling criteria described in the text.

Note.- UK enterprises population figures from UK Business: Activity, Size and Location (IDBR, March 2015).

TABLE A2: Baseline sample's number of firm and employee observations by employer size, 1997 & 2007

Enterprise size <sup>†</sup>	Firms			Employees			IDBR ent. employees (000s) <sup>†</sup>		
	1997	2007	Change in share	1997	2007	Change in share	1997	2007	Change in share
250 - 999	125	43	-0.05	1,729	497	-0.03	89	32	0.00
1,000 - 1,999	352	214	-0.08	5,322	2,814	-0.06	539	328	-0.02
2,000 - 4,999	512	548	0.05	10,068	9,789	-0.03	1,612	1,817	0.03
5,000+	485	569	0.09	26,915	36,242	0.12	9,431	8,350	-0.01
Total	1,474	1,374		44,034	49,342		11,671	10,525	

<sup>†</sup> Values use the IDBR record of the number of employees in the enterprise, which includes the firm.

TABLE A3: Baseline sample's incidence of ISCO88 major occupation groups

Major group <sup>†</sup>	1997	2007	Change
1	0.12	0.15	-0.02
2	0.22	0.16	0.06
3	0.09	0.14	-0.05
4	0.24	0.20	0.03
5	0.11	0.14	-0.03
6 & 7	0.08	0.04	0.04
8	0.08	0.05	0.03
9	0.05	0.11	-0.06

<sup>†</sup> Key: 1. Legislators, senior officials and managers, 2. Professionals, 3. Technicians and associate professionals, 4. Clerks, 5. Service workers and shop and market sales workers, 6. Skilled agricultural and fishery workers, 7. Craft and related trades workers, 8. Plant and machine operators and assemblers, 9. Elementary occupations.

## Appendix B. Mathematical details

### B.1 Variance decomposition - hours and wages

From the main text, we can re-write Equation (1), the total variance of log weekly wages as follows, where  $\omega$  and  $\eta$  denote the hourly non-log wage rate and hours worked respectively,  $\theta$  denotes the log hourly wage rate, and  $h$  denotes log hours. Recalling from the main text that  $i$  denotes a worker and  $j$  a firm, terms with a bar above refer to average values within the given subscript. For example,  $\bar{h}_j$  refers to the sample mean of log weekly hours worked for the employees within firm  $j$ . Terms without a subscript but a bar above refer to averages in the whole sample, across all employees and firms.

$$\underbrace{\frac{1}{N} \sum_j^J \sum_i^{N_j} [\ln(\omega_{ij}\eta_{ij}) - \overline{\ln(\omega_{ij}\eta_{ij})}]^2}_{\text{Overall - } V_e} = \underbrace{\frac{1}{N} \sum_j^J \sum_i^{N_j} [\ln(\omega_{ij}\eta_{ij}) - \overline{\ln(\omega_j\eta_j)}]^2}_{\text{Within-firm - } V_{wf}} + \underbrace{\sum_j^J \frac{N_j}{N} [\overline{\ln(\omega_j\eta_j)} - \overline{\ln(\omega\eta)}]^2}_{\text{Between-firm - } V_{bf}}, \quad (9)$$

with

$$V_{wf} = \underbrace{\frac{1}{N} \sum_j^J \sum_i^{N_j} [\theta_{ij} - \bar{\theta}_j]^2}_{V_{wf}^\theta} + \underbrace{\frac{1}{N} \sum_j^J \sum_i^{N_j} [h_{ij} - \bar{h}_j]^2}_{V_{wf}^h} + \underbrace{\frac{2}{N^2} \sum_j^J \sum_i^{N_j} [(\theta_{ij} - \bar{\theta}_j)(h_{ij} - \bar{h}_j)]}_{2cov_{wf}(\theta, h)} \quad (10)$$

and

$$V_{bf} = \underbrace{\sum_j^J \frac{N_j}{N} [\bar{\theta}_j - \bar{\theta}]^2}_{V_{bf}^\theta} + \underbrace{\sum_j^J \frac{N_j}{N} [\bar{h}_j - \bar{h}]^2}_{V_{bf}^h} + \underbrace{2 \sum_j^J \frac{N_j}{N^2} [(\bar{\theta}_j - \bar{\theta})(\bar{h}_j - \bar{h})]}_{2cov_{bf}(\theta, h)}. \quad (11)$$

### B.2 Shift-share analysis of the change in the between-firm component of employee wages

Let each decile be denoted by  $d$ , where  $N^d$  is all employees observed in a given period in that decile of the distribution of wages which is unobservable in the wage regression Equation (4). Let  $k$  denote an employment type, with  $K$  types in total. The share of all employees, irrespective of decile, in type  $k$  in the firm of an employee  $i$  is given by  $\lambda_{ki}$ , where the dependence of  $k$  on  $i$  is implicit. The mean log weekly



wage of type  $k$  in the firm of employee  $i$  is given by  $\hat{\psi}_{ki}$ . We let this value be zero where a firm does not employ anybody of type  $k$ . Taking the estimated firm-specific component of the log wage  $\hat{\alpha}_j$ , we can write the mean of these values for employees within a decile of the distribution of estimated unobservable wages  $\hat{\psi}_{ij}$  as

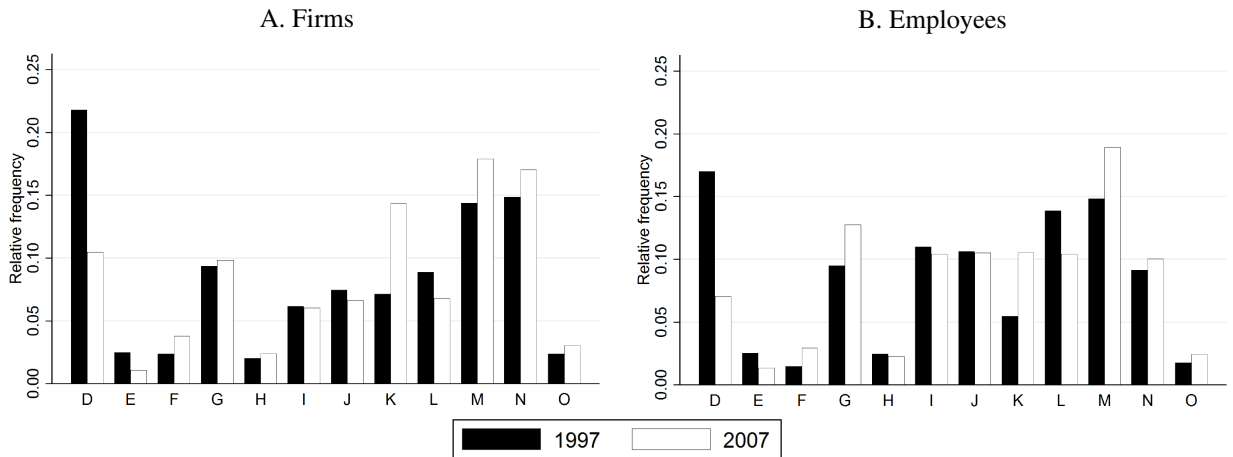
$$\begin{aligned} \frac{1}{N^d} \sum_{i \in d} \{\hat{\alpha}_j\}_i &= \frac{1}{N^d} \sum_k \sum_{i \in d} \lambda_{ki} \hat{\psi}_{ki} \\ &= \sum_k \left[ \underbrace{\left( \frac{1}{N^d} \sum_{i \in d} \lambda_{ki} \right)}_{\bar{\lambda}_k} \underbrace{\left( \frac{1}{N^d} \sum_{i \in d} \hat{\psi}_{ki} \right)}_{\bar{\psi}_k} + \underbrace{\frac{1}{N^d} \sum_{i \in d} (\lambda_{ki} - \bar{\lambda}_k) (\hat{\psi}_{ki} - \bar{\psi}_k)}_{cov(\lambda_k, \hat{\psi}_k)} \right]. \end{aligned} \quad (12)$$

Using (12), denoting historical values by a “ $\bar{\cdot}$ ”, and representing the difference operator by  $\Delta$ , we can write the change over time (between two given years) in the mean of firm-specific log wages for employees in some decile as

$$\sum_k \left[ \underbrace{\bar{\lambda}'_k \Delta \bar{\psi}_k}_{\text{Wages effect}} + \underbrace{\bar{\psi}'_k \Delta \bar{\lambda}_k}_{\text{Shares effect}} + \underbrace{\Delta \bar{\lambda}_k \Delta \bar{\psi}_k}_{\text{Interaction effect}} + \underbrace{\Delta cov(\lambda_k, \hat{\psi}_k)}_{\text{Covar. effect}} \right]. \quad (13)$$

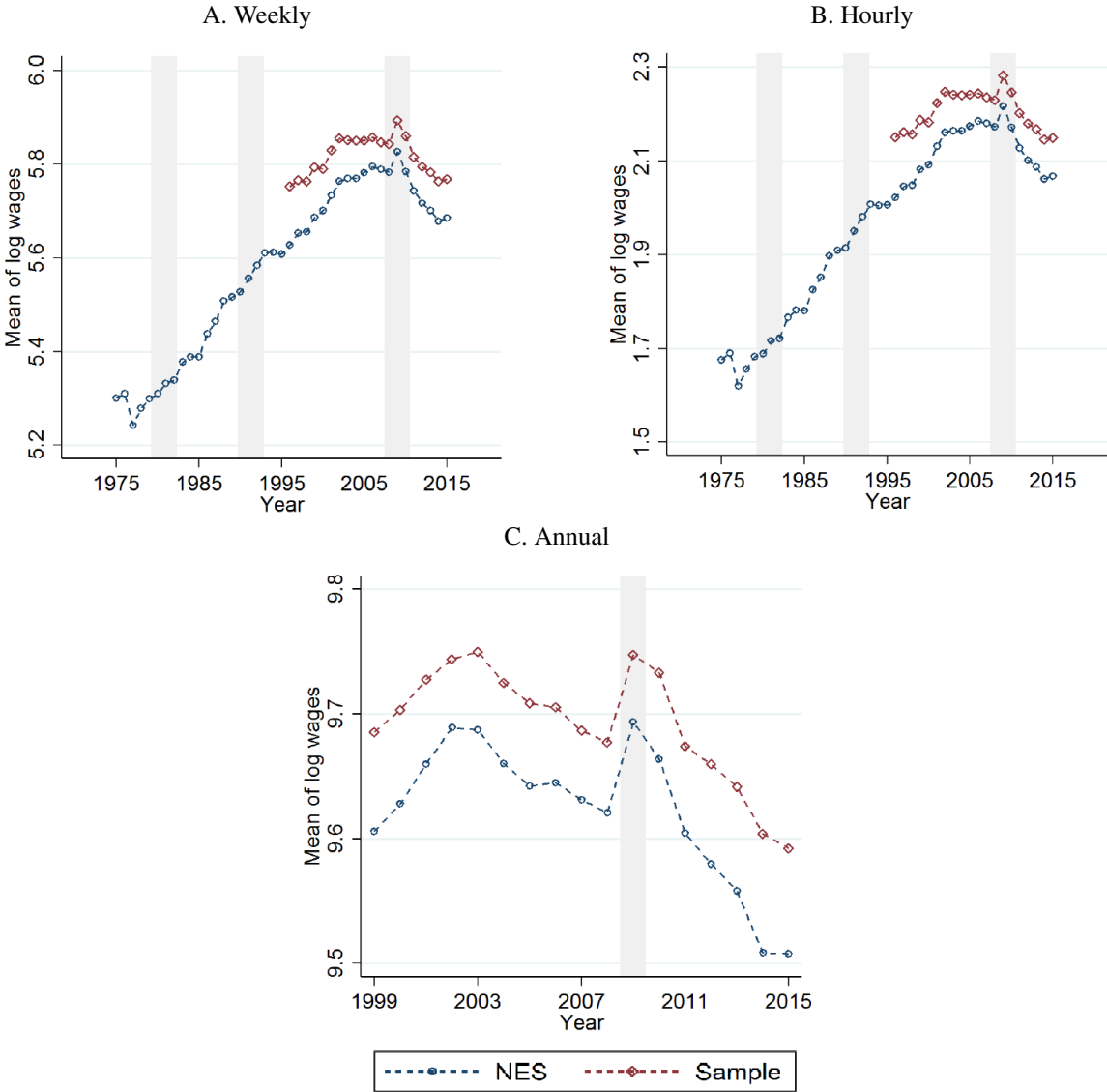
## Appendix C. Additional figures

FIGURE C1: Shares of firms and employees in the baseline sample in SIC 2003 sectors, 1997 & 2007



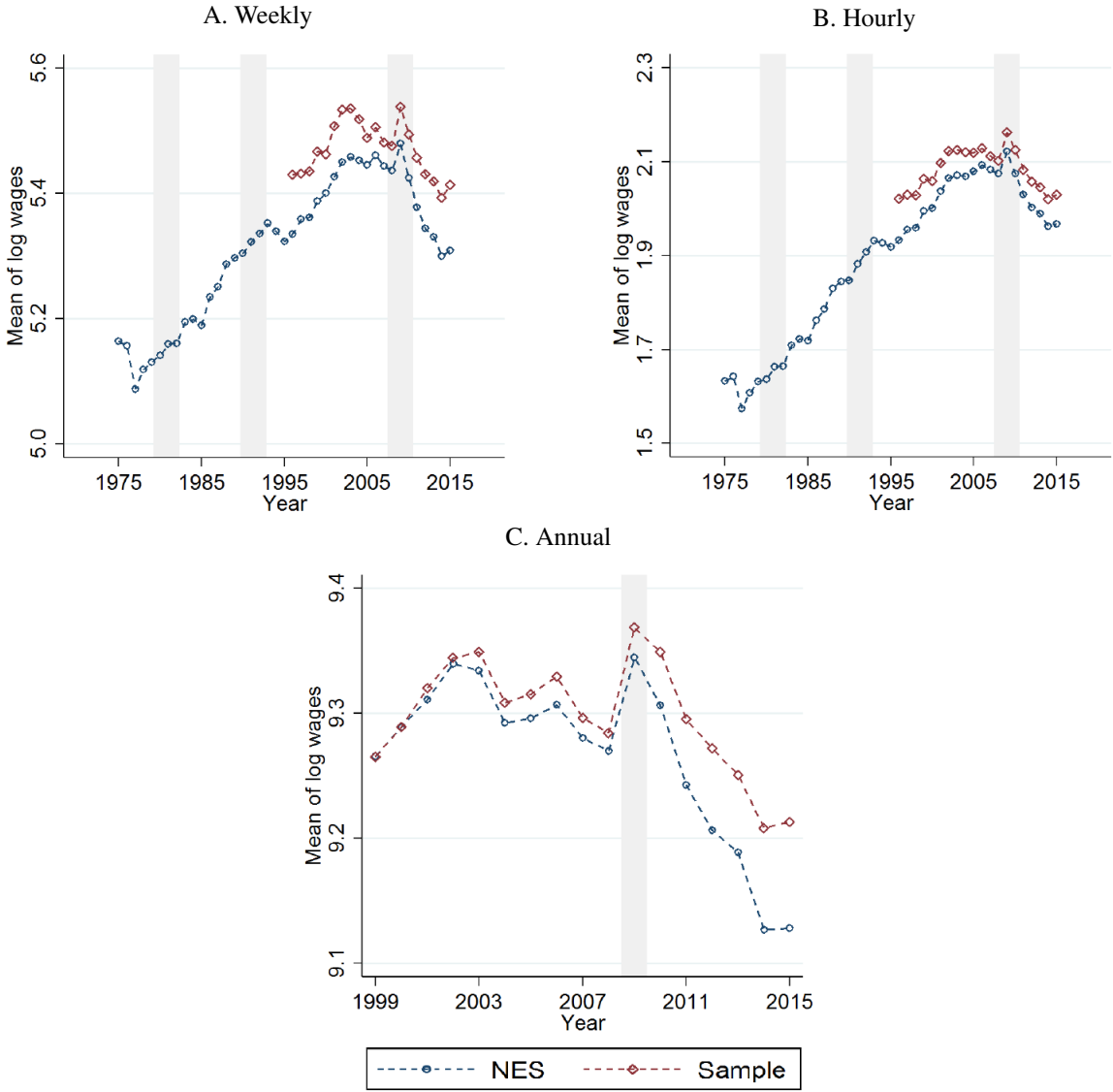
Note.- SIC2003 codes: D. Manufacturing, E. Utilities, F. Construction, G. Wholesale and retail, H. Hotels and restaurant, I. Transport and telecommunication, J. Financial intermediation, K. Real estate, business services, L. Public admin and defence, social security, M. Education, N. Health and social work, O. Other community and social services. SIC2003 categories A-C are omitted due to too small cell sizes.

FIGURE C2: Mean of real log wages in large firms, full-time employees only, and comparison with the whole NESPD sample, 1975-2015



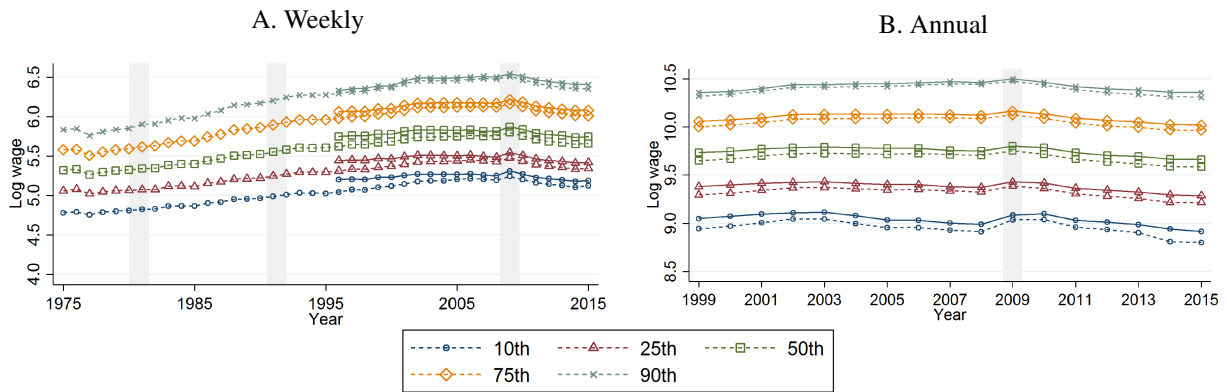
Note.- see Figure 1. The top one percent of wage observations in any year are excluded from all calculations here.

FIGURE C3: Mean of real log wages in large firms, all employees, and comparison with the whole NESPD sample, 1975-2015



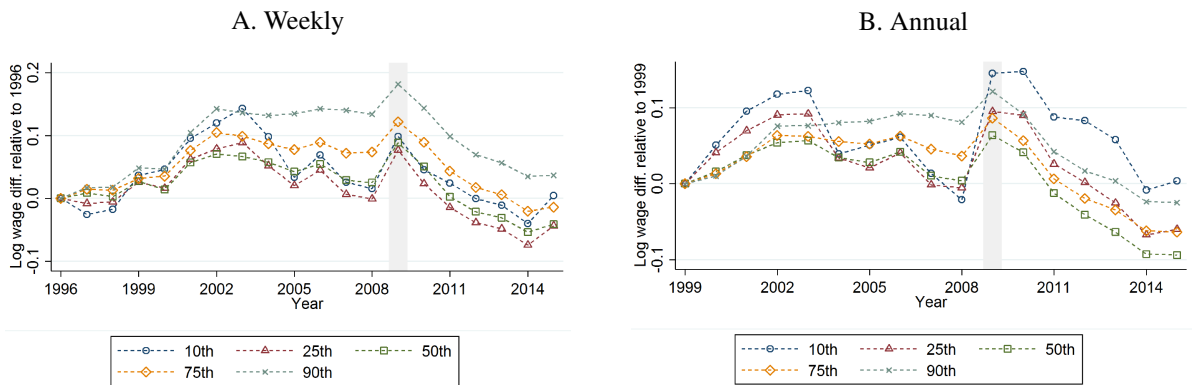
Note.- see Figure 1, except here is with all employees. The top one percent of wage observations in any year are excluded from all calculations here.

FIGURE C4: Percentiles of real log wages in large firms, all employees, and comparison with the whole NESPD sample, 1975-2015



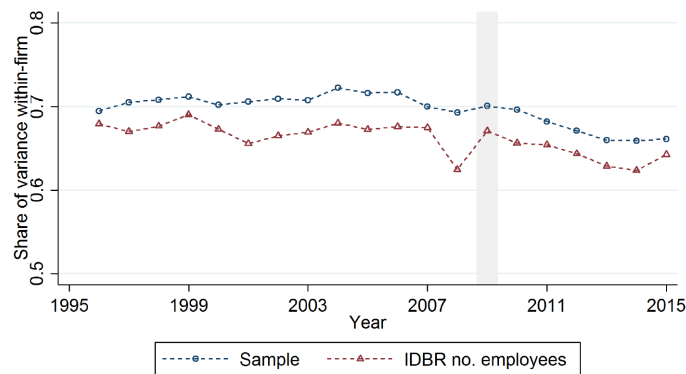
Note.- see Figure 1, except here is with all employees. Solid lines are the series for the large firms sample of the NESPD.

FIGURE C5: Percentiles of real log wages in large firms, all employees: differences relative to 1996/9



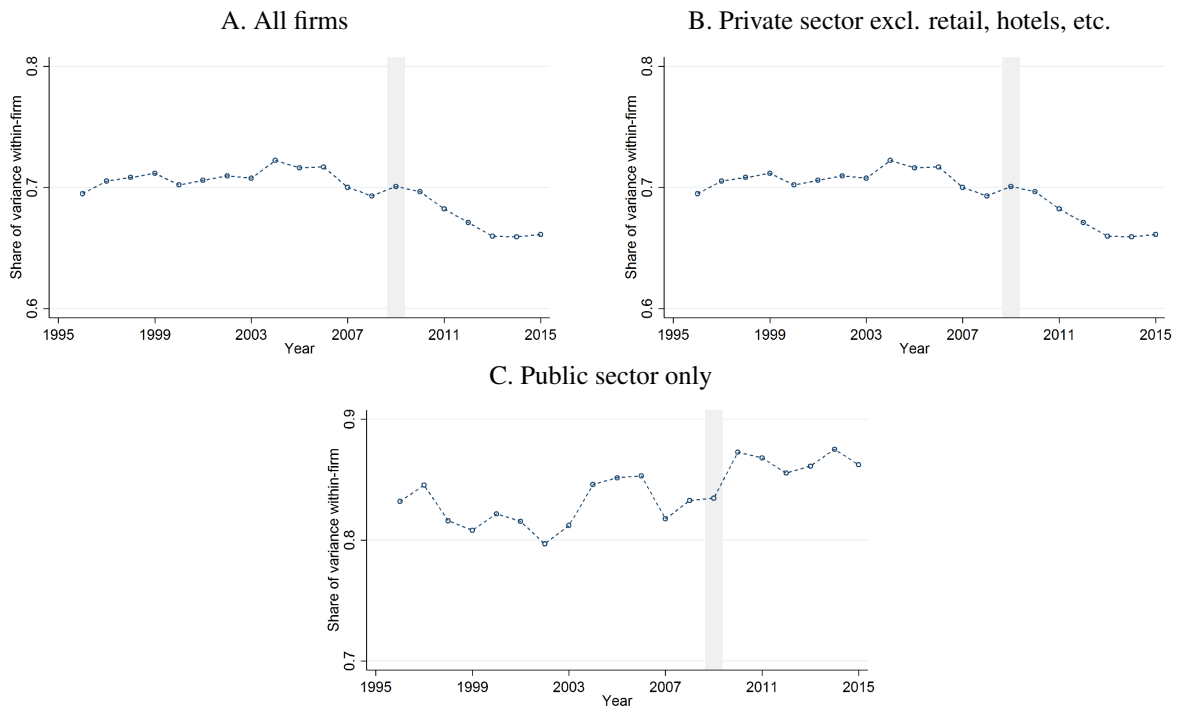
Note.- see Figure 2, except here is with all employees.

FIGURE C6: Share of variance in log weekly employee wages from within-firm component, 1996-2015: comparison of firm weights



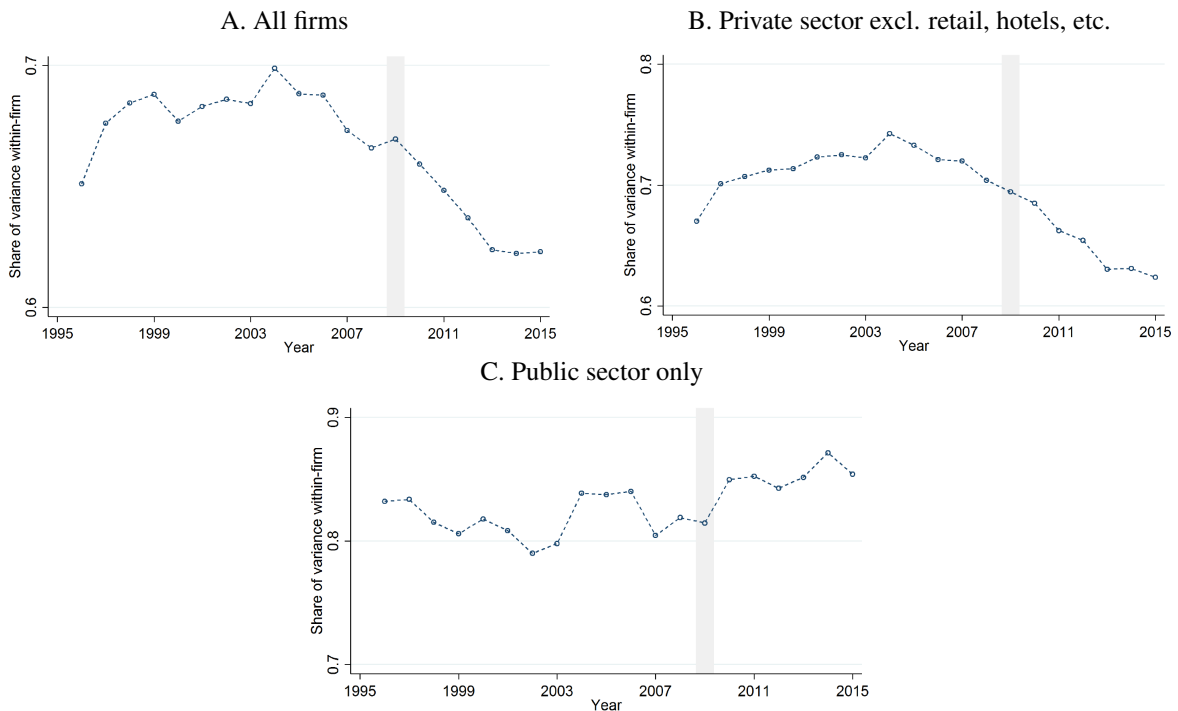
Note.- see Figure 3. ‘Sample’ gives results where firms are weighted using their share of sample observations in that year. ‘IDBR...’ gives results where firms are weighted using their administrative record of enterprise size from the IDBR.

FIGURE C7: Share of variance in log weekly employee wages from within-firm component



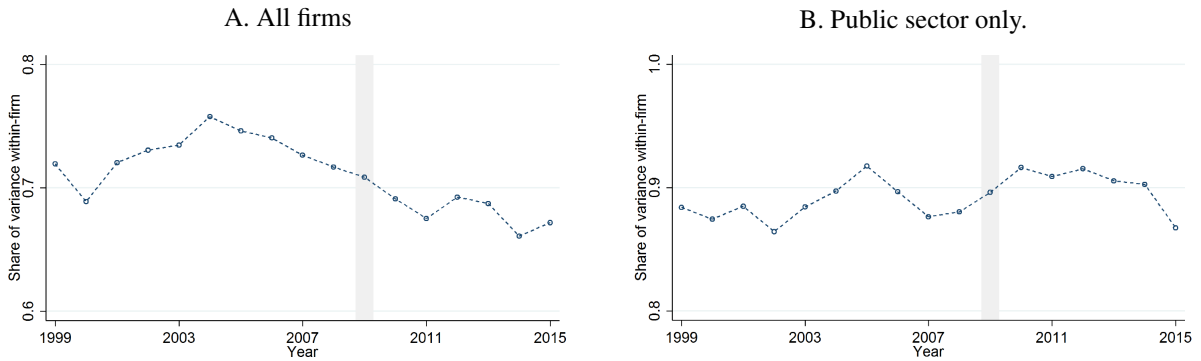
Note.- see Figure 3. Panel B excludes major SIC 2003 sectors G & H. Public sector is represented by public corporation or nationalised industry, central government and local authority employers.

FIGURE C8: Share of variance in log hourly employee wages from within-firm component



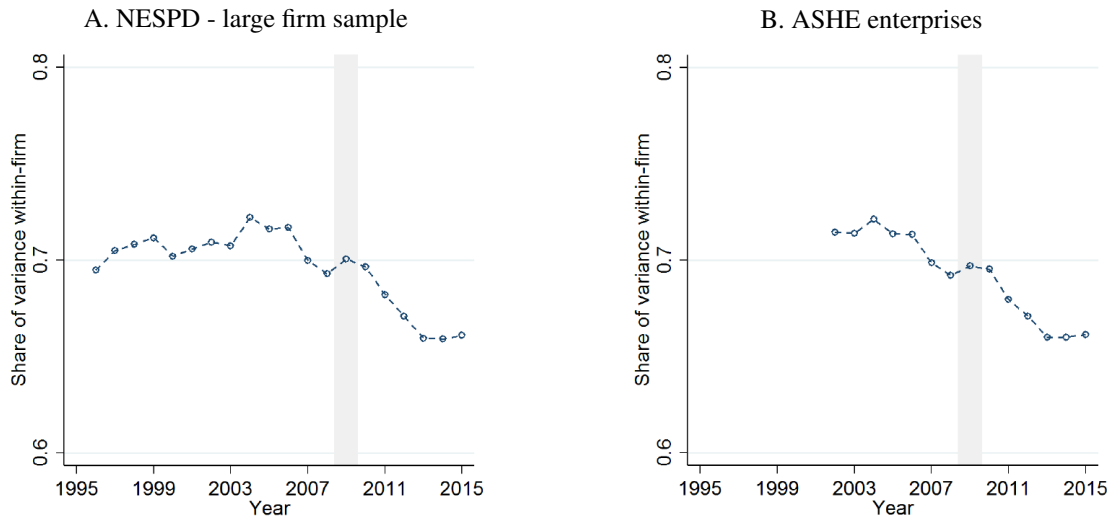
Note.- see Online Appendix Figure C7

FIGURE C9: Share of variance in log annual employee wages from within-firm component



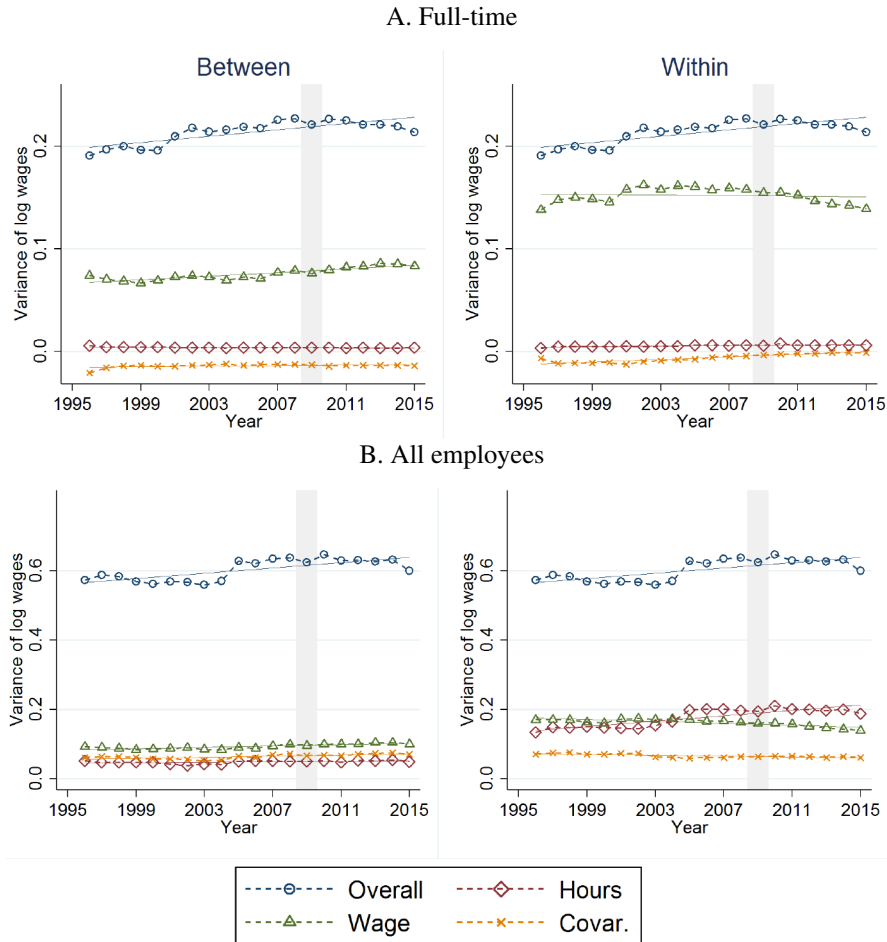
Note.- see Figure 3 and see Online Appendix Figure C7.

FIGURE C10: Share of variance in log weekly employee wages from within-firm component: NESPD large firms sample vs ASHE enterprises



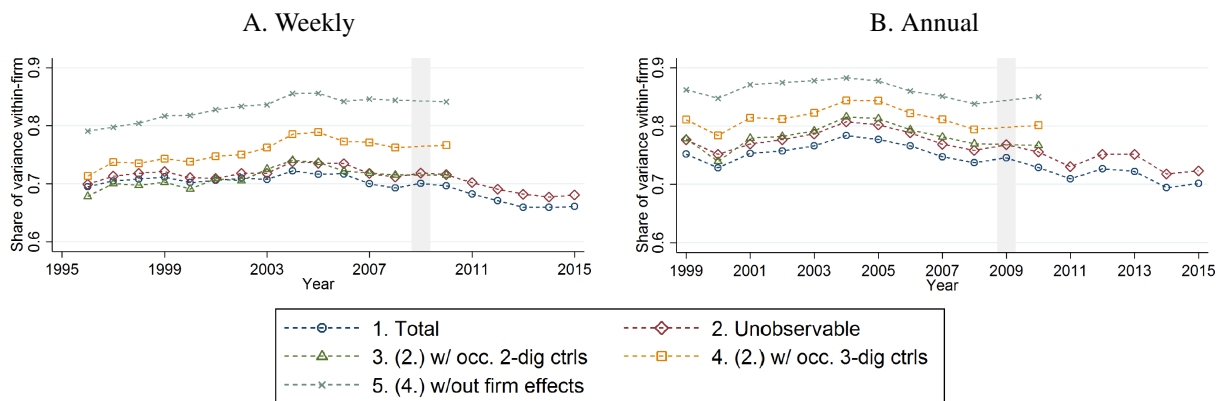
Note.- age 16-64 only, all employees. Weekly wages exclude overtime. In the left panel, the data is for all large firms in the NESPD who have at least ten employee observations in a year. The right panel is the equivalent but using IDBR enterprise identifiers in the ASHE, instead of a looser definition of a 'firm'. Shaded areas represent official UK recessions.

FIGURE C11: Within- and between-firm, hourly rate and usual hours components of the variance in log weekly employee wages, 1996-2015



Note.- age 16-64 only. Wages and hours exclude overtime. The top one percent of wage values in each year are excluded from calculations here. The data is for all large firms in the NESPD who have at least ten (full-time) employee observations in that year. The 'Covar.' series represent twice sample covariance terms. The 'Overall' series, in both left and right panels, is the total sample variance. As such, all other series across both panels sum within year to this total variance. Shaded areas represent official UK recessions. Lines without markers are estimated linear trends.

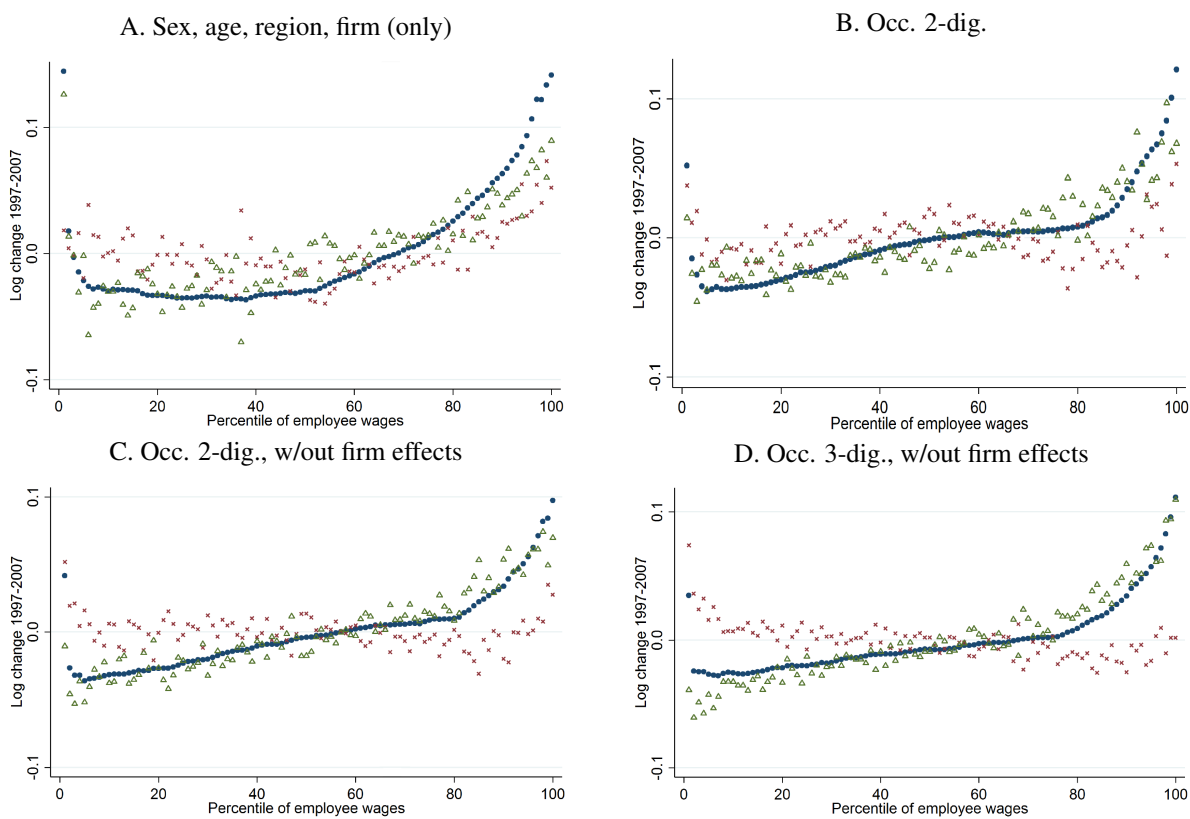
FIGURE C12: Share of within-firm component in the variance of estimated unobservable (residual) log employee wages, 1996-2015



Note.- see Figure 4.

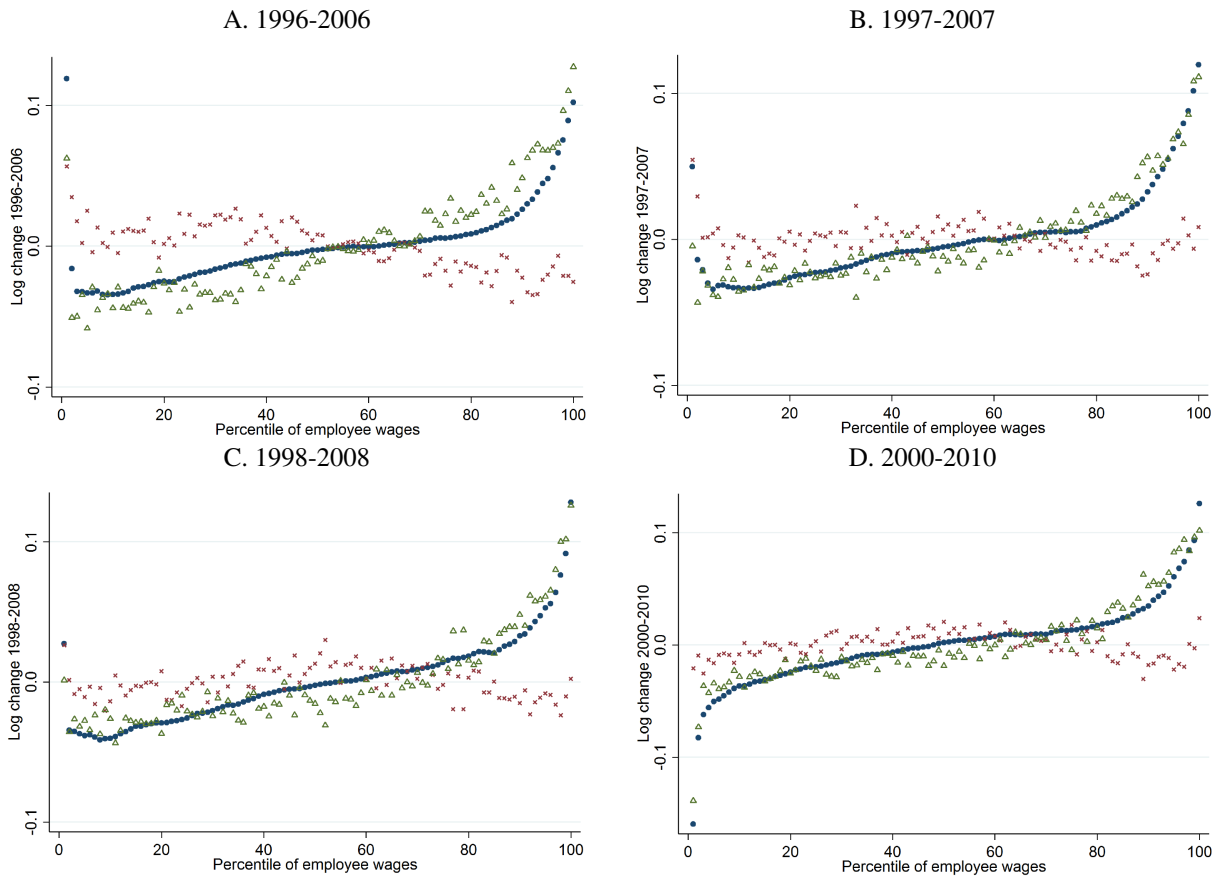


FIGURE C13: Change 1997-2007 in the average real unobservable log weekly wage by percentile of employees and the contribution from firms



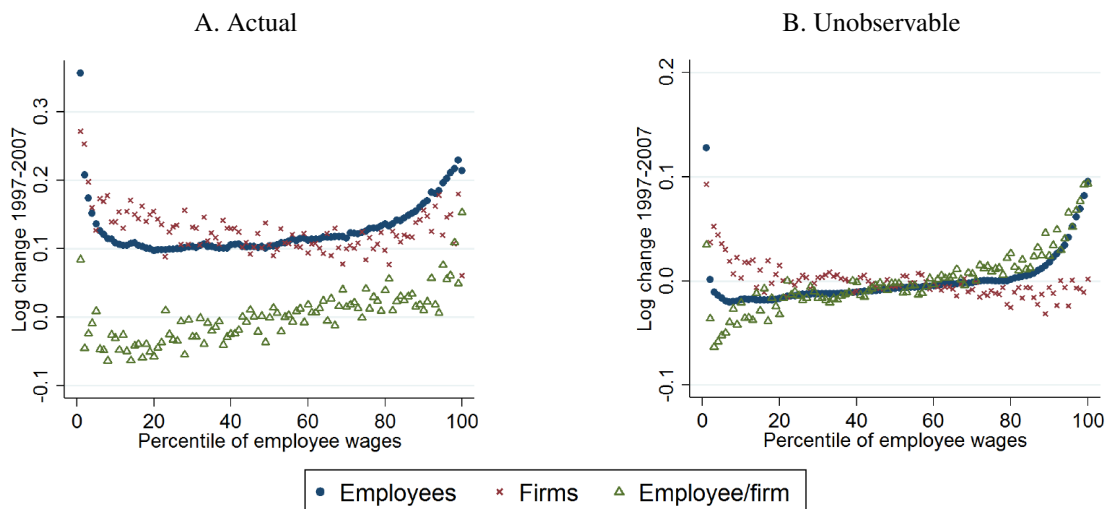
Note.- see Figure 5. Unobservable log wages are estimated using a regression with controls for sex, age, age squared, major regions and firm-specific effects, in addition to those labelled above each panel. A:  $\gamma = 0.67$ . B:  $\gamma = 0.81$ . C:  $\gamma = 0.96$ . D:  $\gamma = 1.14$ .

FIGURE C14: Change in the average real unobservable log weekly wage by percentile of employees and the contribution from firms: other ten-year time periods



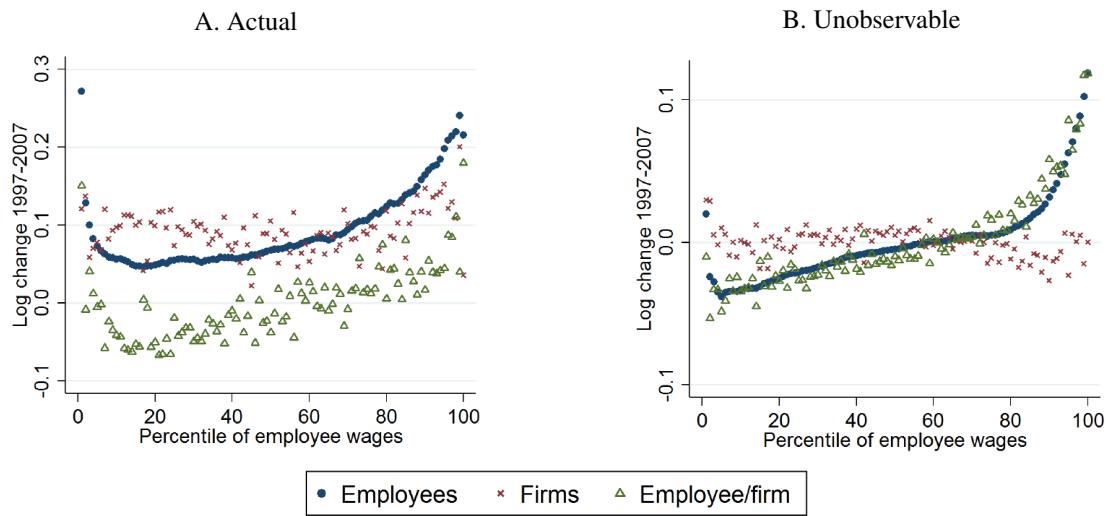
Note.- see Figure 5 and Figure 8. A:  $\gamma = 1.30$ . B:  $\gamma = 1.00$ . C:  $\gamma = 1.03$ . D:  $\gamma = 0.96$ .

FIGURE C15: Change 1997-2007 in the average real log weekly wage by percentile of employees and the contribution from firms: all large firms in the NESPD with 1+ employee observations



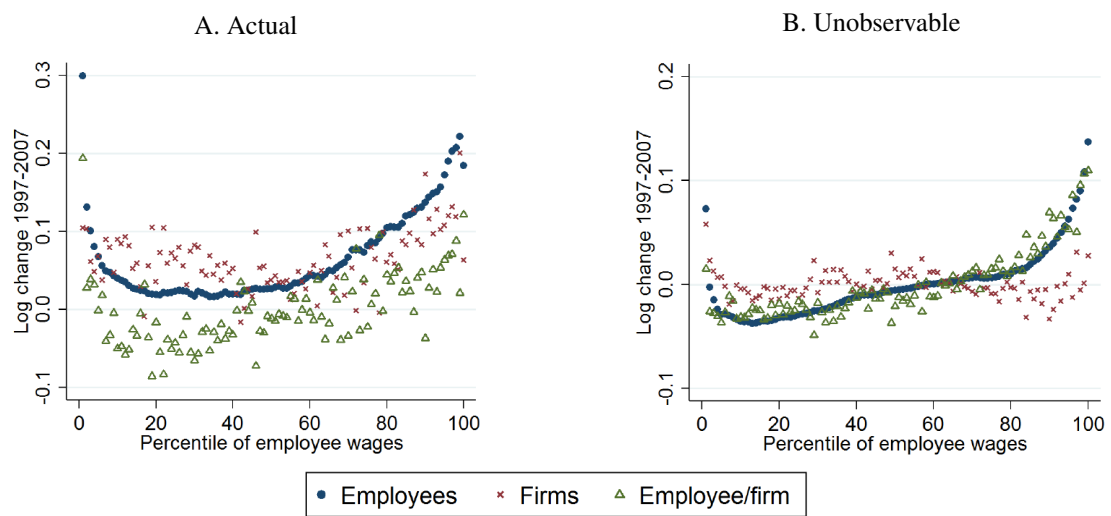
Note.- see Figure 5 and Figure 8. The data used here is for all large firms who have at least one employee observation in the NESPD in a year. A:  $\gamma = 0.61$ . B:  $\gamma = 0.98$ .

FIGURE C16: Change 1997-2007 in the average real log weekly wage by percentile of employees and the contribution from firms: all large firms in the NESPD with 5+ employee observations



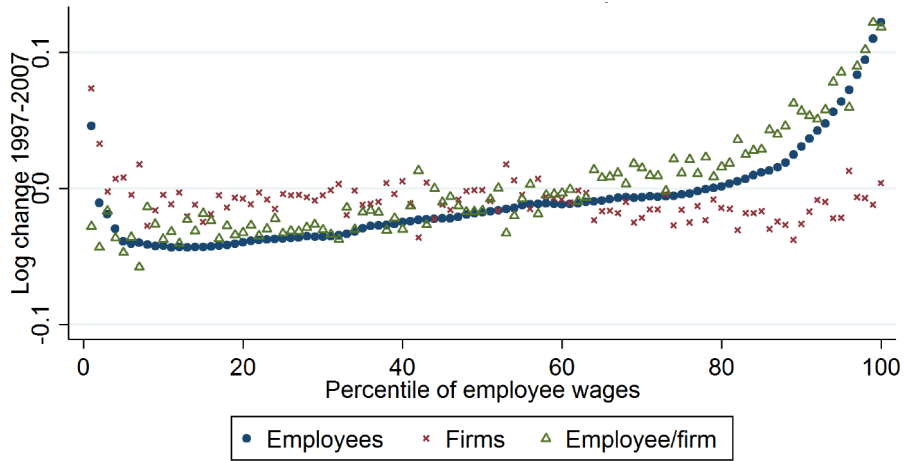
Note.- see Online Appendix Figure C15, except the data used here are for all large firms who have at least five employee observations in the NESPD in a year. A:  $\gamma = 0.75$ . B:  $\gamma = 1.08$ .

FIGURE C17: Change 1997-2007 in the average real log weekly wage by percentile of employees and the contribution from firms: all large firms in the NESPD with 20+ employee observations



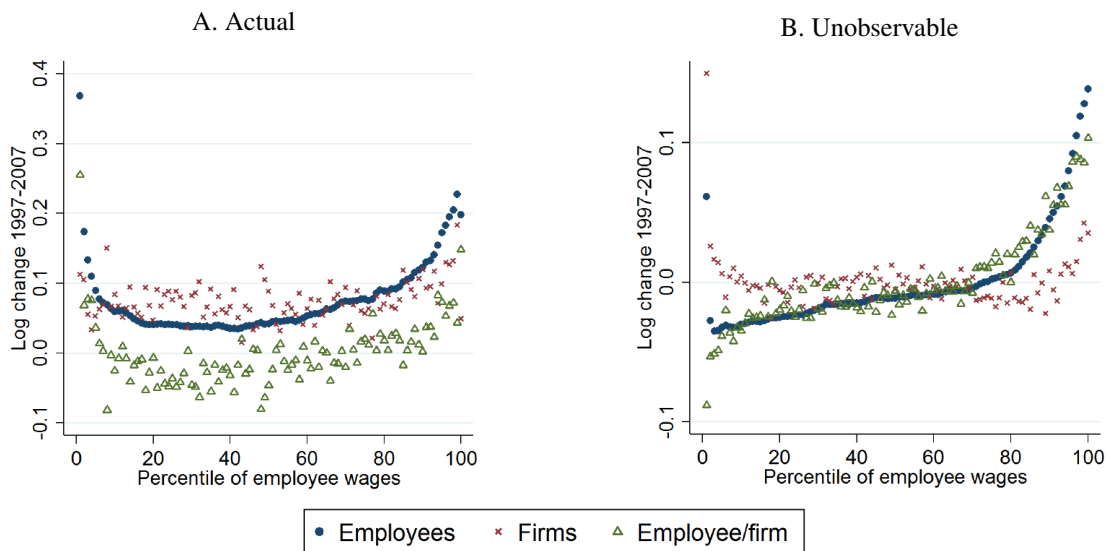
Note.- see Online Appendix Figure C15, except the data used here are for all large firms who have at least twenty employee observations in the NESPD in a year. A:  $\gamma = 0.61$ . B:  $\gamma = 0.93$ .

FIGURE C18: Change 1997-2007 in the average real unobservable log weekly wage by percentile of employees and the contribution from firms: private sector only



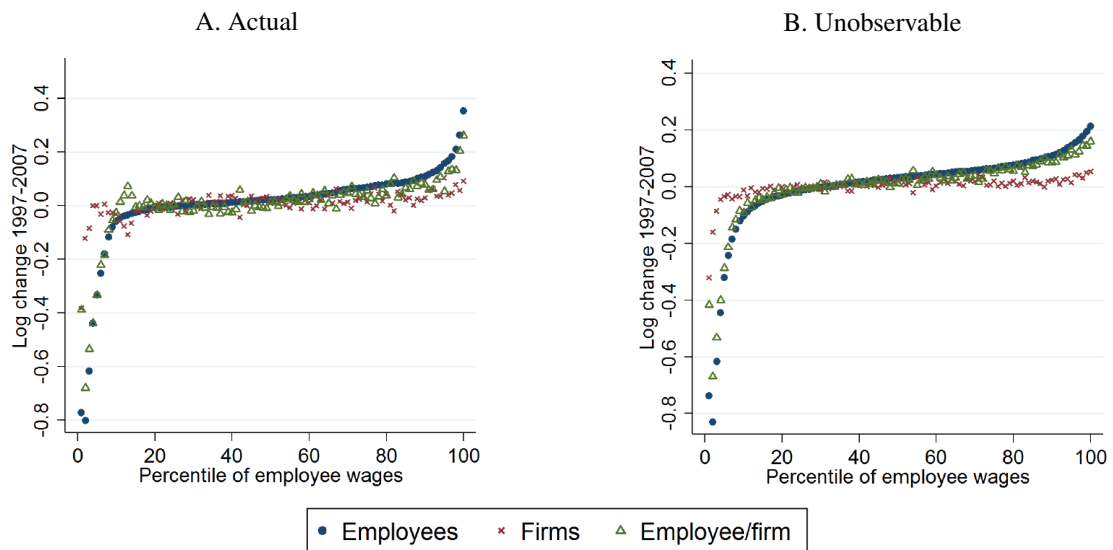
Note.- see Figure 5 and Figure 8, except the data used here is for all large private sector firms in the NESPD who have at least ten full-time employee observations in a year.  $\gamma = 1.00$ .

FIGURE C19: Change 1997-2007 in the average real log hourly wage by percentile of employees and the contribution from firms: comparison with unobservable wages



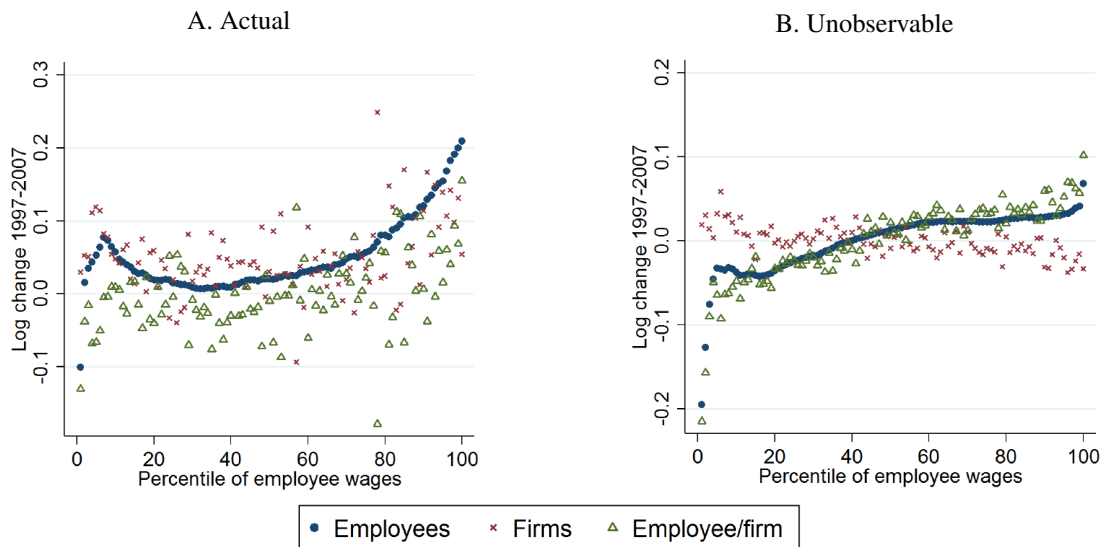
Note.- see Figure 5 and Figure 8. A:  $\gamma = 0.75$ . B:  $\gamma = 0.80$ .

FIGURE C20: Change 1997-2007 in the average real log annual wage by percentile of employees and the contribution from firms: comparison with unobservable wages



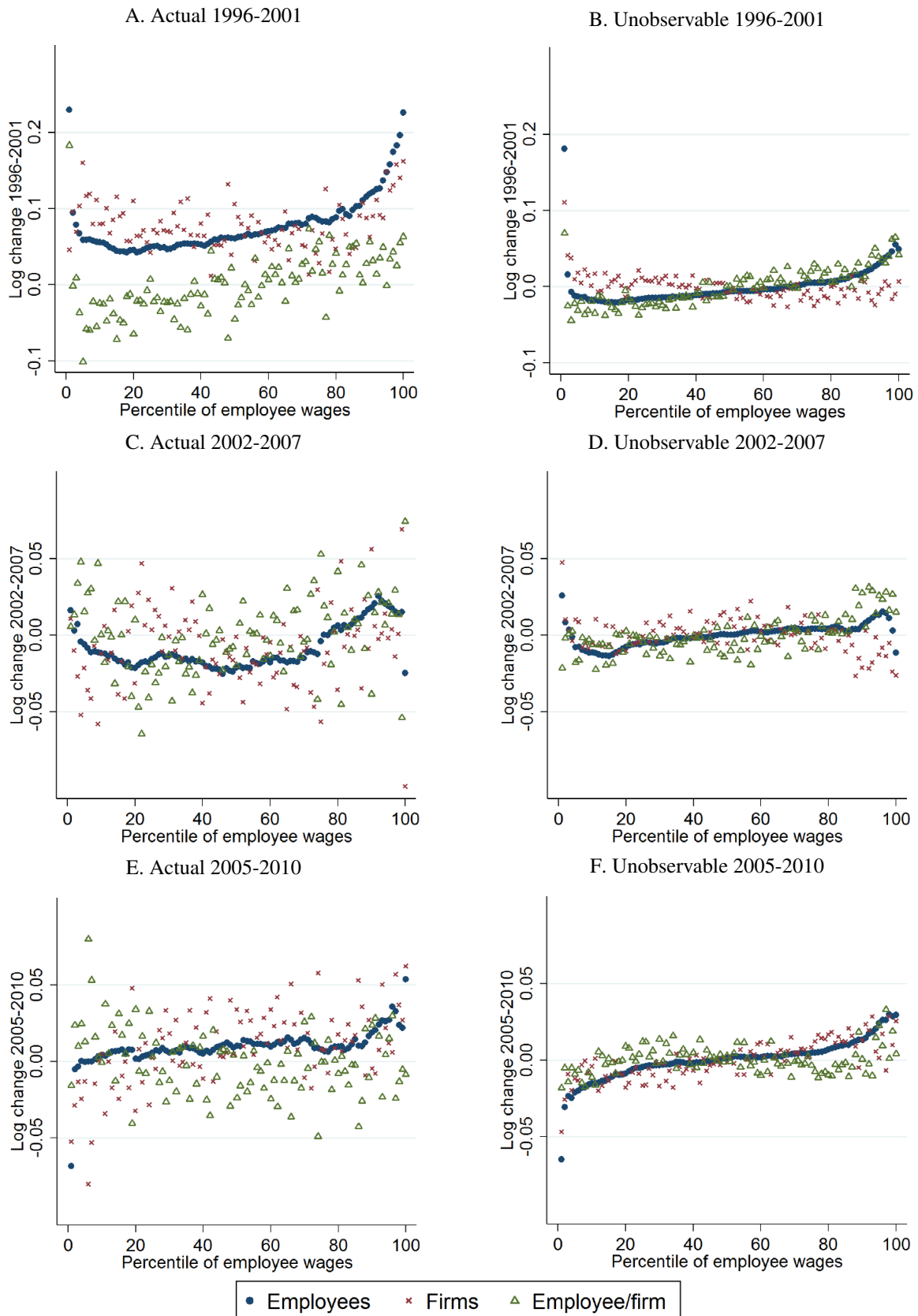
Note.- see Figure 5 and Figure 8, except the data used here is for all large firms who have at least ten employee observations in the NESPD in a year, who have been with the firm at least a year. A:  $\gamma = 0.75$ . B:  $\gamma = 0.77$ .

FIGURE C21: Change 1997-2007 in the average real log weekly wage by percentile of employees and the contribution from firms, full & part-time workers: comparison with unobservable wages



Note.- see Figure 5 and Figure 8, except here the data is for all employees, not full-time only. A:  $\gamma = 0.52$ . B:  $\gamma = 1.25$ .

FIGURE C22: Change in the average real unobservable log weekly wage by percentile of employees and the contribution from firms: other five-year time periods



Note.- see Figure 5 and Figure 8. A:  $\gamma = 0.73$ . B:  $\gamma = 0.77$ . C:  $\gamma = 0.45$ . D:  $\gamma = 0.82$ . E:  $\gamma = -0.14$ . F:  $\gamma = 0.27$ .

## Appendix D. Additional tables

TABLE D1: Summary of decomposition results for the variance in real log annual employee wages

	1999		2005		2015		Change 1999-2005		Change 1999-2015	
	Level (1)	Share (2)	Level (3)	Share (4)	Level (5)	Share (6)	Level (7)	Share of* (8)	Level (9)	Share of* (10)
Actual annual earnings:	0.279		0.393		0.412		0.115		0.133	
Between	0.078	0.281	0.010	0.254	0.135	0.328	0.022	0.19	0.057	0.43
Within	0.201	0.719	0.293	0.746	0.277	0.672	0.093	0.81	0.076	0.57
Unobs. annual earnings:	0.202		0.303				0.101			
Between	0.039	0.189	0.047	0.157			0.008	0.08		
Within	0.163	0.811	0.256	0.843			0.093	0.92		

\* Values give the fraction or share of the overall change in log employee annual wage variance accounted for by each component.  
 Note.- "Unobs.:" wages control for 3-digit occupations (see Section 3.1). Relevant rows may not sum exactly due to rounding.

TABLE D2: Summary of decomposition results for the variance in real log weekly wages, all employees

	1996		2005		2015		Change 1996-2005		Change 1996-2015	
	Level (1)	Share (2)	Level (3)	Share (4)	Level (5)	Share (6)	Level (7)	Share of* (8)	Level (9)	Share of* (10)
Actual weekly wages:	0.574		0.629		0.601		0.055		0.027	
Between: total	0.204	0.356	0.204	0.325	0.218	0.363	0.000	0.00	0.014	0.52
Hourly wages <i>bf</i>	0.092	0.160	0.090	0.143	0.099	0.165	-0.002	-0.04	0.007	0.26
Usual hours <i>bf</i>	0.051	0.090	0.050	0.080	0.049	0.082	-0.001	-0.02	-0.002	-0.07
Covariance <i>bf</i>	0.061	0.106	0.064	0.102	0.070	0.116	0.003	0.06	0.009	0.33
Within: total	0.370	0.644	0.424	0.675	0.383	0.637	0.055	1.00	0.013	0.48
Hourly wages <i>wf</i>	0.169	0.295	0.171	0.272	0.139	0.231	0.002	0.04	-0.031	-1.15
Usual hours <i>wf</i>	0.133	0.232	0.199	0.316	0.188	0.312	0.066	1.20	0.055	2.04
Covariance <i>wf</i>	0.068	0.118	0.055	0.087	0.056	0.094	-0.013	-0.24	-0.011	-0.41

\* Values give the fraction or share of the overall change in log employee wage variance accounted for by each component.  
 Note.- see Table 1.