Economics Letters 118 (2013) 63–67

Contents lists available at SciVerse ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Wage sorting trends

Jesper Bagger [a], Kenneth L. Sørensen [b,c], Rune Vejlin [b,*]

[a] Department of Economics, Royal Holloway College, University of London, Egham, Surrey TW20 0EX, United Kingdom
[b] Department of Economics and Business, Aarhus University, Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark
[c] KORA, Danish Research Institute for Local and Regional Governments, Denmark

ARTICLE INFO

Article history:
Received 20 August 2012
Received in revised form
19 September 2012
Accepted 19 September 2012
Available online 27 September 2012

JEL classification:
J30
J31
J62

Keywords:
Matched employer–employee data
Firm fixed effects
Worker fixed effects
Wage sorting
Wage inequality

ABSTRACT

We document a strong trend towards more positive assortative wage sorting using Danish Matched Employer–Employee data from 1980 to 2006. The pattern is not due to compositional changes in the labor market and primarily occurs among high wage workers.

1. Introduction

The seminal paper of Abowd et al. (1999), refined and extended in Abowd et al. (2002), investigates whether "high wage firms" employ "high wage workers". The empirical analysis builds on a log wage regression with worker and firm fixed effects. Here, a worker (firm) fixed effect is a time invariant log wage component specific to a given worker (firm). A high wage worker is a worker with a relatively high worker fixed effect (analogously for firms). The authors compute the empirical correlation between worker and firm fixed effects, pooling annual cross sections, and find that it is negative in France (correlation −0.28 using data from 1976 to 1987) and in the US (correlation −0.03 using data from 1984 to 1993).[1] Similar studies have since been conducted on a number of different datasets.[2] We refer to the correlation between worker and firm fixed effects as wage sorting.[3]

The purpose of this paper is to document and examine trends in wage sorting. We use a Danish full population Matched Employer–Employee (MEE) panel for 1980–2006. Pooling across annual cross sections, the correlation between worker and firm fixed effects is 0.05. We show that this estimate masks a systematic nonstationarity. By computing cross section specific correlations we find that the correlation between worker and firm effects increases from a low −0.07 in 1981 to a high 0.14 in 2001. The trend towards positive assortative wage sorting occurs almost exclusively in the top quartile of the distribution of workers effects, i.e. among high wage workers, where the increase is from

* Corresponding author. Tel.: +45 87 16 52 64.
E-mail addresses: jesper.bagger@rhul.ac.uk (J. Bagger), ksoerensen@econ.au.dk (K.L. Sørensen), rvejlin@econ.au.dk (R. Vejlin).

1 These results are reported in Abowd et al. (2002).

2 See e.g. Gruetter and Lalive (2004), (1990–1997, correlation −0.22, Austria), Andrews et al. (2008), (1993–1997, correlation −0.21 to −0.15, Germany), Sørensen and Vejlin (forthcoming), (1980–2006, correlation −0.06 to 0.11, Denmark).

3 This notion of wage sorting is not linked to economic theory, and is distinct from that of productivity sorting, i.e. sorting on worker and firm productivity. A number of recent studies of productivity sorting (see e.g. Eeckhout and Kircher (2011), Bagger and Lentz (2012) and Bartolucci and Devicienti (2012)) find that it is difficult to identify productivity sorting from wage data alone.

0165-1765/$ – see front matter
doi:10.1016/j.econlet.2012.09.021

Table 1
Summary statistics.

Year	Obs.	Avg. ln w	S.d. ln w	Share women	Avg. age	Avg. years of education	Avg. experience
1980	767,088	5.069	0.304	0.24	36.43	10.45	21.50
1985	787,526	5.103	0.293	0.24	36.47	10.81	20.14
1990	777,097	5.246	0.296	0.26	37.09	11.19	19.59
1995	778,641	5.257	0.303	0.28	38.82	11.49	19.91
2000	816,112	5.291	0.326	0.31	41.44	11.67	21.11
2005	799,643	5.299	0.335	0.32	43.06	11.78	21.86

a low −0.20 to a high 0.37. The change in the wage sorting is economically important: it comprises 41% of the increase in the standard deviation of log wages between 1980 and 2006.

We ascertain that the nonstationary wage sorting pattern is due to nonstationarity in the covariance between firm and worker effects, and that it is not driven by compositional changes in the labor force in terms of education, age, and gender. Finally, we show that the wage sorting trend is associated with worker reallocation via voluntary quits, and with entry and exit of workers over the period we consider.

2. Data

Our empirical analysis is based on a Danish register-based annual MEE panel covering 1980–2006. The unit of observation is a given individual in a given year. Measures of actual labor market experience are available from 1964. For workers entering the labor market prior to 1964 (born before 1948) we add the potential pre-1964 experience net of education.[4]

The raw data consists of 60,847,593 observations. We inflate wages to 2006 levels. We discard (i) public sector jobs and individuals under education (19,191,599 observations), (ii) observations with missing data (6,103,607 observations), and (iii) observations preceding observed labor market entry or if the individual enters later than age 35 (13,804,815 observations). We trim the within-experience-education group wage distribution (top and bottom 1% deleted, 503,454 observations) and select the maximal set of connected workers and firms (99,953 observations deleted).[5] The analysis data contains 21,144,165 observations.

Table 1 documents that average (real) log wages and their dispersion are increasing over our data period. Moreover, average education increases by around 1.5 years over the data period, the labor force ages due to the general demographic development, average experience is stable, and female (private sector) labor force participation is increasing.[6]

3. Econometric framework

Let i index individuals, j index employers, and let t index annual cross sections. The function $\mathrm{J}(i, t)$ maps individual observations into employer IDs. Consider a log-linear two-way error component wage equation:

$$\ln w_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + \theta_i + \psi_{\mathrm{J}(i,t)} + \varepsilon_{it}, \tag{1}$$

where $\ln w_{it}$ is the log-wage, $\mathbf{x}_{it}'$ contains time-varying regressors: experience, experience squared and a set of year dummies, θ_i is a time-invariant worker effect, $\psi_{\mathrm{J}(i,t)}$ is a time-invariant firm effect, and ε_{it} is the residual log-wage. Throughout we maintain the assumption that $\mathrm{E}[\varepsilon_{it}|\mathbf{x}_{it}', \mathrm{J}(\cdot,\cdot), i, t] = 0$.[7] Conditioning on workers and firms being connected ensures that the matrix of regressors in (1) has full column rank. (1) is a widely used representation of log wages.

Wage sorting is often measured by $\widehat{\rho}$, the correlation between the estimated worker and firm effects computed by pooling all available cross sections. We report cross section specific wage sorting estimates, denoted $\widehat{\rho}_t$. Formally, let $\widetilde{\theta}_{it} = (\widehat{\theta}_i - \widehat{\mu}_{\theta,t})/\widehat{\sigma}_{\theta,t}$ and $\widetilde{\psi}_{\mathrm{J}(i,t)t} = (\widehat{\psi}_{\mathrm{J}(i,t)} - \widehat{\mu}_{\psi,t})/\widehat{\sigma}_{\psi,t}$ be worker and firm effects standardized with respect to cross section t averages and standard errors. Let N be the total number of observations and let $\mathbb{I}_t$ be the index set of workers present in cross section t. Then,

$$\widehat{\rho}_t = \frac{1}{|\mathbb{I}_t|}\sum_{i=1}^{N}\mathbf{1}(i \in \mathbb{I}_t)\widetilde{\theta}_{it}\widetilde{\psi}_{\mathrm{J}(i,t)t}, \tag{2}$$

where $\mathbf{1}(\cdot)$ is an indicator function. Notice that $\widehat{\rho}_t = \sum_{k=1}^{K}\widehat{\pi}_{kt}\widehat{\rho}_{kt}$, where $\widehat{\pi}_{kt} = |\mathbb{I}_{kt}|/|\mathbb{I}_t|$ is the share of cross section t workers belonging to group k ($\mathbb{I}_{kt}$ is the index set of workers in group k in cross section t), and $\widehat{\rho}_{kt} = \sum_{i=1}^{N}\mathbf{1}(i \in \mathbb{I}_{kt})\widetilde{\theta}_{it}\widetilde{\psi}_{\mathrm{J}(i,t)t}/|\mathbb{I}_{kt}|$ measures the statistical dependence between $\widetilde{\theta}_{it}$ and $\widetilde{\psi}_{\mathrm{J}(i,t)t}$ in group k in cross section t. This decomposition allows us to assert the extent to which changes to $\widehat{\rho}_t$ stem from compositional changes, i.e. changes to $\widehat{\pi}_{kt}$, and from within-group changes in wage sorting, i.e. changes to $\widehat{\rho}_{kt}$.

4. Results

Pooling annual cross sections, the correlation between the estimated worker and firm fixed effects is $\widehat{\rho} = 0.05$. Fig. 1 plots the $\widehat{\rho}_t$-profile (solid line) which exhibits a strong upward trend over most of the period we consider. This phenomenon has not been documented in previous studies. Overall, the correlation increases from a low −0.07 in 1981 to a high 0.14 in 2001 at which point the correlation declines slightly.

The dashed line in Fig. 1 plots the time profile of $\widehat{\rho}_t^*$, which is computed similarly to $\widehat{\rho}_t$ (cf. (2)), except that worker and firm effects are standardized using the time-invariant (grand) means and standard errors from the pooled cross sections. Comparing the solid and dashed lines in Fig. 1, we note they are almost coinciding; the rising $\widehat{\rho}_t$-profile is driven exclusively by changes in the covariance between worker and firm effects.

The empirical covariance between estimated worker and firm effects underestimates the true covariance (cf. Andrews et al., 2008): if a firm effect is under-estimated, workers at that firm will have over-estimated worker effects, and vice versa. This could drive the rising ρ_t-profile if the bias is more pronounced in earlier years. To ascertain that this is not the case we retain the allocation of workers to firms as found in the data, but simulate counter

[4] In this specification older workers are assigned too much experience. We have experimented with different forms of pre-1964 experience, including specifications that assign too little experience to older workers. Our results are very robust to these changes.

[5] See Abowd et al. (2002) for an explanation of the necessity of conditioning on workers and firms being connected.

[6] Potential experience is trending upwards while our actual experience measure is stationary. We ascribe this to older cohorts being assigned too much experience, and an increased prevalence of sabbaticals from education during 1980–2006.

[7] See Abowd et al. (1999) and Postel-Vinay and Robin (2006) for discussions of the economic content of this assumption.