The Anatomy of the Wage Distribution. How do Gender and Immigration Matter?

Rasmus Lentz* Suphanit Piyapromdee† Jean-Marc Robin‡

July 17, 2016

Preliminary and incomplete

Abstract

In this paper, we propose an estimation method that allows for unrestricted interactions between worker and firm unobserved characteristics in both wages and the mobility patterns. Related to Bonhomme, Lamadon and Manresa (2014) (BLM), our method identifies double sided unobserved heterogeneity through an application of the EM-algorithm where the firm classification is repeatedly updated so as to improve on the likelihood function. In Monte Carlo simulations we demonstrate that the cyclic updating of the firm classification provides a significant performance improvement. Firm classification is a result of both wage and mobility patterns in the data. We estimate the model on Danish matched employer-employee data for the period 1985-2011. The estimation includes gender, education, age and time controls. We find an increased sorting pattern over time, although overall sorting is modest. The wage gap between genders is decreasing over time. The contribution to the wage gap from mobility differences between the genders is also decreasing over time.

Keywords: On-the-job search; Heterogeneity; Wage distributions

JEL codes: E24; E32; J63; J64

*University of Wisconsin-Madison. E-mail: rlentz@wisc.edu

†University College London. Email s.piyapromdee@ucl.ac.uk

‡Sciences-Po, Paris and University College London. Email jmarc.robin@gmail.com

1 Introduction

The studies of inequality and discrimination are both vexed by the challenges of measuring the returns to unobservable characteristics and the correlation between observables and unobservables. The challenge is further complicated by the fact that the relevant unobservables reside on both sides of the employment relationship; with both firms and workers. For example, are women paid less because they are women? Or because their gender is correlated with particular worker side unobservables? Or because they tend to work for firms or in jobs that pay less regardless of the gender of the worker?

In this paper, we propose an estimation method that allows for unrestricted interactions between worker and firm unobserved characteristics in both wages and the mobility patterns. Related to Bonhomme et al. (2015) (BLM), our method identifies double sided unobserved heterogeneity through an application of the EM-algorithm.

The log wage of a match is assumed to be normally distributed with a mean and variance that depends on both observed and unobserved worker and firm heterogeneity. Unobserved heterogeneity is described through groupings where all workers in a given group are identical according to the unobserved characteristics, similarly for firms.

Workers move between firms according to transition probabilities that depend both on worker and firm characteristics. Employed workers meet other employment opportunities at a rate that depends on the type of the current firm. The type of the employment opportunity is drawn from a distribution that is common to all workers and independent of the type of the current firm. Finally, the probability that the worker decides to quit the current job and move to the new firm is assumed to depend both on the worker's type as well as the types of the two firms involved. In our particular specification, the choice probability takes the shape of a binomial logit. Hence, one interpretation of the mobility patterns in the paper is that of a standard on the job search model with random utility.

The estimator maximizes the likelihood of observing workers' wage and employment histories given a data structure that records the identity of the worker's employer at any point in time as well non-employment. We do not consider the earliest part of a worker's history and furthermore assume the worker's state at the beginning of the observation window is drawn from the steady state. Wages are observed for each employment spell and can change within the spell at an annual frequency. Unobserved heterogeneity on the worker side is modeled as a random effect whereas the firm side is modeled as a fixed effect because the likelihood function evaluation is infeasible in case of a firm side random effect. The maximization of the likelihood of the data is implemented through an application of the EM-algorithm with double sided unobserved heterogeneity.

The estimator is initialized by some firm classification. We choose to follow BLM and group firms by the k-means algorithm using wage data. The estimator proceeds to update the firm classification in the iteration steps. For a given firm classification, the algorithm performs an EM step where a worker classification (in terms of a posterior on a worker's type) is obtained in the E step for given wage and mobility parameters, and in the M-step the wage and mobility parameters are updated subject to the worker and firm classifications so as to maximize worker posterior conditional expected log-likelihood of the data. The mobility parameters are close to those of a Bradley-Terry model but nevertheless involve a substantial non-linearity in the likelihood function with respect to the mobility parameters. We formulate an MM algorithm that allow a fast solution for mobility parameters that improve on the likelihood function - a substantial contribution for the feasibility of the estimator. With the likelihood improved by the EM step, the firm classification is then occasionally updated before a new EM step so as to improve on the expected log-likelihood. The firm classification updates are done subject to the entire likelihood of the data and can change substantially from its initialization.

The model will be estimated on Danish matched employer-employee data where the worker's employment match state is observed at a weekly frequency. The wage data are observed specifically for the given match in question and can change at an annual frequency. We observe both hours and earnings in the match. The model allows us to provide a detailed measurement of the returns to observable characteristics such as gender, ethnic origin, age, and education. And given the dynamic structure of the model we can also distinguish between wage and value variance and their decompositions.

2 The Model

Workers are indexed by $i \in \{1,...,I\}$ and firms are indexed by $j \in \{0,1,...,J\}$, where $j=0$ reflects non-employment. For each worker i, we observe $(w_{it}, j_{it}, x_{it})_{t=1}^{T}$, where $j_{it} \equiv j(i,t) \in \{0,1,...,J\}$ is the ID of worker i's employer at time t.[1] x_{it} are observed worker controls, and w_{it} is the worker's log-wage rate at time index t. Note that although the number of repeated observations many vary across individuals, we assume a balanced panel to simplify the notations. The survey covers a specific calendar time period, but observations for worker i may start at any date and end at any later date. So x_{it} is a vector that contains at least calendar time, say τ_{it}, measured for example as different time intervals like five-year periods. In the application, x_{it} will also contain information on gender, education and age. We assume that x_{it} is discrete, allowing to estimate different parameters for different control values. Wages are measured at annual frequency within a given match but mobility between employment states and firms is measured at a weekly frequency. Thus, wage observations are missing except for the first week of the match and the first week of the year.

[1]The first employment state j_{i1} is necessarily with a firm: $j_{i1} > 0$.

We assume that empℓoyers (firms) can be clustered into L different groups indexed by $\ell \in \{1,\dots,L\}$ and that wor*k*ers can be clustered into K different groups indexed by $k \in \{1,...,K\}$. The index ℓ_j is the type of firm j and k_i is the type of worker i. We shall be treating the unobserved firm types $\mathscr{L} = (\ell_1,...,\ell_J)$ as a a fixed effect (i.e. a parameter to be estimated) and the worker type as a random effect mixing conditional distributions of workers' trajectories. Unemployment is observable and is denoted by $\ell = 0$.

Let $f_{k\ell}(w|x)$ denote the wage density, conditional on the worker's type k, the employer's type ℓ and observable control x. In the expressions below, adopt the convention that $f_{k0}(\cdot|x) = 1$ since there is no wage attached to unemployment spells. We assume that w_{it} and $w_{it'}$ are independent conditional on $\ell_{j(i,t)}, \ell_{j(i,t')}$. We further denote the probability for a worker of type k of making a transition from a firm of type ℓ to a firm of type ℓ' (possibly $\ell' = \ell$) at the end of period t as $M_{k\ell\ell'}(x)$, and the probability of staying with the same employer is $\overline{M}_{k\ell}(x) = 1 - \sum_{\ell'=0}^{L} M_{k\ell\ell'}(x)$. With this, $M_{k\ell 0}(x)$ is the layoff rate in a match between a type k worker and type ℓ firm. The state of unemployment is special in that by definition, $M_{k00}(x) = 0$. The worker type k conditional job finding rate in the model is $1-\overline{M}_{k0}(x)$. Finally, let $\pi_{k\ell}(x)$ denote the probability for initial matches (for $t = 1$) to be of type (k,ℓ) given x.

3 Estimation

In this section we develop a Classification Esperance Maximization (CEM) algorithm for estimating the mixture model.

3.1 Likelihood with observed firm types

Let us first consider the case where the employers' types ℓ_j are observed. Let $\ell_{it} = \ell_{j(i,t)}$ denote, by some abuse of notation, the type of the firm employing worker i in period t. Let also

$$D_{it} \equiv D(i,t) = \begin{cases} 1 & \text{if } j_{i,t+1} \neq j_{it} \\ 0 & \text{if } j_{i,t+1} = j_{it} \end{cases}$$

indicate an employer change between t and $t+1$.

For a value $\beta = (f, M, \pi)$ of the parameters and a classification $\mathscr{L}$ of firms, the likelihood for one worker i is

$$\sum_{k=1}^{K} L_i(k; \beta, \mathscr{L}),$$

where $L_i(k;\beta,\mathscr{L})$ is the individual likelihood conditional on worker type k, i.e.

$$L_i(k;\beta,\mathscr{L}) = \pi_{k,\ell_{i1}}(x_{i1}) \prod_{t=1}^{T} f_{k,\ell_{it}}(w_{it}|x_{it}) \prod_{t=1}^{T-1} \overline{M}_{k,\ell_{it}}(x_{it})^{1-D_{it}} M_{k,\ell_{it},\ell_{i,t+1}}(x_{it})^{D_{it}}, \tag{1}$$

where by convention $f_{k\ell}(\cdot|x) = 1$ if the wage observation is missing, and assuming that for the last observation period we do not know whether a mobility occurs or not by the end of it.

3.2 An EM algorithm with known firm types

The EM algorithm (Dempster et al., 1977) consists of iterating the calculation of the posterior probabilities of worker types (E-step) and the expected log-likelihood maximization using the type probabilities calculated in the E-step (M-step). See Arcidiacono and Jones (2003); Bonhomme and Robin (2009); Arcidiacono and Miller (2011) for recent applications in economics. The firm classification is in the data unobserved. It is infeasible to evaluate the likelihood function for the formulation of the model where a firm's unobserved type is a latent variable symmetric to the unobservable worker type formulation in equation (1). The difficulty lies with accounting for the co-dependency between a firm's workers resulting from their matches to a common firm type in a setup where workers move between firms. Consequently, we estimate the model for a given firm classification $\mathscr{L}$. We shall explain in the next subsection how we set and update $\mathscr{L}$.

For a given value of $\beta = (f, M, \pi)$, the posterior probability of worker i to be of type k given all wages and controls (all the available information) is

$$p_i(k;\beta,\mathscr{L}) = \frac{L_i(k;\beta,\mathscr{L})}{\sum_{k=1}^{K} L_i(k;\beta,\mathscr{L})}. \tag{2}$$

Then, define

$$Q_i(f;\beta^{(m)},\mathscr{L}) = \sum_{k=1}^{K} p_i(k;\beta^{(m)},\mathscr{L}) \left[\sum_{t=1}^{T} \ln f_{k,\ell_{it}}(w_{it}|x_{it})\right]. \tag{3}$$

It is the expected log-likelihood of worker i's wages for a given value $\beta^{(m)}$ of the parameter. The worker posteriors are determined by the model parameters and firm classification $\left(\beta^{(m)},\mathscr{L}\right)$, where the superscript is used to denote the given EM-algorithm iteration. Also, let

$$H_i(M;\beta^{(m)},\mathscr{L}) = \sum_{k=1}^{K} p_i(k;\beta^{(m)},\mathscr{L}) \left[\sum_{t=1}^{T-1} \left\{(1-D_{it})\ln \overline{M}_{k,\ell_{it}}(x_{it}) + D_{it} \ln M_{k,\ell_{it},\ell_{i,t+1}}(x_{it})\right\}\right]. \tag{4}$$

It is the expected log-likelihood of worker i's employment history conditional on the first state ℓ_{i1}.

The EM algorithm iterates the following steps.

E-step For $\beta^{(m)} = (f^{(m)}, M^{(m)}, \pi^{(m)})$ and $\mathscr{L}$ calculate posterior probabilities $p_i(k; \beta^{(m)}, \mathscr{L})$.

M-step Update $\beta^{(m)}$ by maximizing $\sum_i p_i(k;\beta^{(m)},\mathscr{L}) \ln L_i(k;\beta,\mathscr{L})$ subject to $\sum_{k,\ell} \pi_{k\ell}(x) = 1$ for all x, that is

$$f^{(m+1)} = \arg\max_f \sum_{i=1}^{I} Q_i(f; \beta^{(m)}, \mathscr{L}), \tag{5}$$

$$M^{(m+1)} = \arg\max_M \sum_{i=1}^{I} H_i(M, \mathscr{L}; \beta^{(m)}, \mathscr{L}), \tag{6}$$

$$\pi_{k\ell}^{(m+1)}(x) = \frac{\sum_{i=1}^{I} p_i(k;\beta^{(m)},\mathscr{L})\,\mathbf{1}\{x_{i1} = x, \ell_{i1} = \ell\}}{\#\{i : x_{i1} = x\}}. \tag{7}$$

3.3 Firm classification

We propose an iterative algorithm where the firm classification is repeatedly updated so as to maximize the total likelihood of the data. We must however confront that parts of the likelihood cannot be sensibly compared across firm classifications. This concerns the likelihood of the observed mobility between firms as well as the likelihood of the worker's initial match. To see this, consider a simple example where there is only a single firm and worker type and there is no unemployment. Hence, in truth, the steady state probability is fully placed on a single combination $\pi(k,\ell) = 1, k = \ell = 1$. But the estimation divides firms arbitrarily into $L = 2$, half of the firms in one group and half in the other. The model parameters are the same across the two firm types since there is no difference between them. However, in the evaluation of the likelihood in equation (1) it follows that $\pi(k,\ell) = 1/2$. The likelihood of the worker's mobility pattern is similarly reduced. Thus, the likelihood mechanically drops as a result of the arbitrary split of a firm group into two identical groups. Had the likelihood understood that a firm is equally likely to be one or the other firm type, the likelihood would have remained unchanged. But the fixed effect approach to the firm classification does not allow this consideration.

Consider the expected complete log-likelihood given firm types:

$$L(\beta, \mathscr{L}) = \mathbb{E} \sum_{i=1}^{I} \sum_{k=1}^{K} y_{ik} \ln L_i(k;\beta,\mathscr{L}) = \sum_{i=1}^{I} \sum_{k=1}^{K} p_i(k;\beta,\mathscr{L}) \ln L_i(k;\beta,\mathscr{L}),$$

where y_{ik} is equal to 1 if individual i is of type k and is otherwise equal to 0. By analogy with Lloyd's algorithm for k-means, given an initial value $\beta, \mathscr{L}$, where β can be obtained given $\mathscr{L}$

using the previous EM algorithm, we suggest to update $\mathscr{L}$ firm by firm as

$$\begin{aligned}\widehat{\ell}_j(\beta,\mathscr{L}) = \arg\max_{\ell_j} & \sum_{i=1}^{I}\sum_{k=1}^{K} p_i(k;\beta,\mathscr{L}) \left[\sum_{t=1}^{T} \ln f_{k,\ell_j}(w_{it}|x_{it}) \times \mathbf{1}\{j(i,t)=j\}\right] \\ & + \sum_{i=1}^{I}\sum_{k=1}^{K} p_i(k;\beta,\mathscr{L}) \left[\ln \pi_{k,\ell_j}(x_{i1}) \times \mathbf{1}\{j(i,1)=j\}\right] \\ & + \sum_{i=1}^{I}\sum_{k=1}^{K} p_i(k;\beta,\mathscr{L}) \left[\sum_{t=1}^{T-1} \left\{ (1-D_{it}) \ln \overline{M}_{k,\ell_j}(x_{it}) + \right.\right. \\ D_{it} & \left.\left. \left[\mathbf{1}\{\ell_{j(i,t+1)}=0\} \ln M_{k,\ell_j,0}(x_{it}) + \mathbf{1}\{\ell_{j(i,t+1)} \neq 0\} \ln \left[\sum_{\ell'=1}^{L} M_{k,\ell_j,\ell'}(x_{it})\right]\right]\right\} \times \mathbf{1}\{j(i,t)=j\}\right].\end{aligned} \tag{8}$$

The first row is the likelihood contribution of the firm's observed wages conditional on its type. The second and third rows represent the likelihood of the observed job durations in the firm and whether a separation is to unemployment or another firm, conditional on the type of the firm. Because non-employment is an observed state, we can compare the likelihood of this part of the mobility observations across classifications. Furthermore, the probability of staying with a given firm type is also directly comparable across firm classifications.

This leaves the question of initialization of the firm classification, $\mathscr{L}^0$. For this we opt for simplicity: We rank firms by average wage per worker in the firm and divide the firms equally into groups based on the sorting.

4 Empirical specification

4.1 Wage distribution

Wages are assumed lognormal given match type. Specifically,

$$f_{k\ell}(w) = \frac{1}{\sigma_{k\ell}} \varphi\left(\frac{w-\mu_{k\ell}}{\sigma_{k\ell}}\right), \tag{9}$$

with $\varphi(x) = (2\pi)^{-1/2} e^{-x^2/2}$. This specification of the log-wage mean allows for a match-specific mean $\mu_{k\ell}$ and variance $\sigma_{k\ell}^2$.

The M-step update 5 takes the following form:

$$\mu_{k\ell}^{(m+1)}(x) = \frac{\sum_{i=1}^{I} p_i(k;\beta^{(m)}) \sum_{t=1}^{T} \mathbf{1}\{\ell_{it} = \ell, x_{it} = x\} w_{it}}{\sum_{i=1}^{I} p_i(k;\beta^{(m)}) \sum_{t=1}^{T} \mathbf{1}\{\ell_{it} = \ell, x_{it} = x\}}$$
$$\sigma_{k\ell}^{(m+1)}(x) = \frac{\sum_{i=1}^{I} p_i(k;\beta^{(m)}) \sum_{t=1}^{T} \mathbf{1}\{\ell_{it} = \ell, x_{it} = x\}[w_{it} - \mu_{k\ell}^{(m+1)}(x)]^2}{\sum_{i=1}^{I} p_i(k;\beta^{(m)}) \sum_{t=1}^{T} \mathbf{1}\{\ell_{it} = \ell, x_{it} = x\}}.$$

4.2 Transition probabilities

We omit conditioning on x_{it} to simplify the notations. The probability for a worker of type k of a transition from a firm of type $\ell \geq 0$ to a firm of type $\ell' \geq 0$ at time t is

$$M_{k\ell\ell'} = \lambda_\ell \nu_{\ell'} P_{k\ell\ell'}. \tag{10}$$

Parameter $\lambda_\ell \in [0,1]$ is the probability of a meeting with an outside employer when the current state is either unemployment or a job of type ℓ. Parameter $\nu_{\ell'} \geq 0$, with $\sum_{\ell'=0}^{L} \nu_{\ell'} = 1$, is the probability that the outside draw is unemployment or a job of type ℓ'.

The parameter $P_{k\ell\ell'}$ is the probability that the transition from ℓ to ℓ' becomes effective. We assume a Bradley-Terry specification for $P_{k\ell\ell'}$ (see e.g. Agresti, 2003; Hunter, 2004). That is,

$$P_{k\ell\ell'} = \frac{\gamma_{k\ell'}}{\gamma_{k\ell} + \gamma_{k\ell'}} \mathbf{1}\{\ell \neq 0 \vee \ell' \neq 0\}, \quad P_{k00} = 0. \tag{11}$$

Parameter $\gamma_{k\ell}$, with $\sum_{\ell=0}^{L} \gamma_{k\ell} = 1$, measures the quality of the match (k,ℓ). If the worker draws a same-type job we assume the workers moves with probability 1/2.[2] For a currently unemployed worker such same-type transitions do not make sense. A draw $\ell' = 0$ when $\ell = 0$ never generates an observed mobility. So $P_{k00} = 0$.

We assume that the offer distribution out of non-employment is the same as that when employed. Since offers are not always accepted, we do not trivially obtain the offer distribution from the distribution of accepted jobs out of non-employment. Rather the assumption is the basis for the identification of γ_{k0} and λ_0. Note yet that a worker can change job within the same sector ℓ; it makes no sense for unemployment.

[2] We experimented the specification

$$M_{k,\ell,\ell'} = \lambda_\ell \nu_{\ell'} \frac{\gamma_{k,\ell'}}{\theta \gamma_{k\ell} + \gamma_{k,\ell'}},$$

where $\theta > 0$ measures the incumbent's advantage and parametrizes mobility within the same group of firms. However it appeared difficult to disentangle θ from λ.

With this it follows that, for $\ell \geq 0$,

$$\overline{M}_{k\ell} = 1 - \sum_{\ell'=0}^{L} M_{k\ell\ell'} = 1 - \lambda_\ell \sum_{\ell'=0}^{L} \nu_{\ell'} P_{k\ell\ell'} = 1 - \lambda_\ell + \lambda_\ell \sum_{\ell'=0}^{L} \nu_{\ell'}(1 - P_{k\ell\ell'}).$$

4.3 M-step update for transition probabilities

In the M-step of the EM algorithm, we maximize the part of the expected likelihood that refers to transitions, i.e.

$$H(M;\beta^{(m)}) \equiv \sum_{k=1}^{K} \sum_{\ell=0}^{L} \left\{ \overline{n}_{k\ell}(\beta^{(m)}) \ln \overline{M}_{k\ell} + \sum_{\ell'=0}^{L} n_{k\ell\ell'}(\beta^{(m)}) \ln M_{k\ell\ell'} \right\},$$

where

$$\overline{n}_{k\ell}(\beta^{(m)}) = \sum_i p_i(k;\beta^{(m)}) \# \{t : D_{it} = 0, \ell_{it} = \ell, x_{it} = x\},$$
$$n_{k\ell\ell'}(\beta^{(m)}) = \sum_i p_i(k;\beta^{(m)}) \# \left\{t : D_{it} = 1, \ell_{it} = \ell, \ell_{i,t+1} = \ell', x_{it} = x\right\},$$

where $\#\{\}$ denotes the cardinality of a set and where we reintroduce the control $x_{it} = x$ to remind that we are estimating different parameters for all different control values x. This likelihood is similar to the likelihood of a Bradley-Terry model except that when the incumbent firm ℓ wins we do not know against which ℓ'. The likelihood is thus rendered more nonlinear by the presence of the term in $\ln \overline{M}_{k\ell}$. An MM algorithm can still be developed as follows.[3]

Because the logarithm is concave, we can minorize $\ln \overline{M}_{k\ell}$ as follows. With obvious notations,

[3] The MM algorithm works by finding a function that minorizes the objective function and that is more easily maximized. Let $f(\theta)$ be the objective concave function to be maximized. At the m step of the algorithm, the constructed function $g(\theta|\theta_m)$ will be called the minorized version of the objective function at θ_m if

$$g(\theta|\theta_m) \leq f(\theta), \forall \theta, \text{ and } g(\theta_m|\theta_m) = f(\theta_m).$$

Then, maximize $g(\theta|\theta_m)$ instead of $f(\theta)$, and let $\theta_{m+1} = \arg\max_\theta g(\theta|\theta_m)$. The above iterative method guarantees that $f(\theta_m)$ converges to a local optimum or a saddle point as m goes to infinity because

$$f(\theta_{m+1}) \geq g(\theta_{m+1}|\theta_m) \geq g(\theta_m|\theta_m) = f(\theta_m).$$

for $\ell \geq 0$,

$$\ln \overline{M}_{k\ell} = \ln \left(1 - \lambda_\ell + \sum_{\ell'=0}^{L} \lambda_\ell \nu_{\ell'} (1 - P_{k\ell\ell'}) \right)$$
$$\geq \frac{1-\lambda_\ell^{(s)}}{\overline{M}_{k\ell}^{(s)}} \ln \left(\frac{1-\lambda_\ell}{1-\lambda_\ell^{(s)}} \overline{M}_{k\ell}^{(s)} \right) + \sum_{\ell'=0}^{L} \frac{\lambda_\ell^{(s)} \nu_{\ell'}^{(s)} (1-P_{k\ell\ell'}^{(s)})}{\overline{M}_{k\ell}^{(s)}} \ln \left(\frac{\lambda_\ell \nu_{\ell'} (1-P_{k\ell\ell'})}{\lambda_\ell^{(s)} \nu_{\ell'}^{(s)} (1-P_{k\ell\ell'}^{(s)})} \overline{M}_{k\ell}^{(s)} \right).$$

Note that both sides of the inequality are equal if $\beta = \beta^{(s)}$ (no parameter change).

Let

$$\widetilde{n}_{k\ell\ell'}^{(s)} = \overline{n}_{k\ell}^{(m)} \frac{\lambda_\ell^{(s)} \nu_{\ell'}^{(s)} (1-P_{k\ell\ell'}^{(s)})}{\overline{M}_{k\ell}^{(s)}}.$$

This is the predicted number of times that home beats visitor ℓ'. Given initial values $\lambda_\ell^{(s)}, \nu_{\ell'}^{(s)}$ one can update $\gamma^{(s)}$ so as to maximize

$$\sum_{k=1}^{K} \sum_{\ell=0}^{L} \sum_{\ell'=0}^{L} \left\{ \widetilde{n}_{k\ell\ell'}^{(s)} \ln \frac{\gamma_{k\ell}}{\gamma_{k\ell} + \gamma_{k\ell'}} + n_{k\ell\ell'} \ln \frac{\gamma_{k\ell'}}{\gamma_{k\ell} + \gamma_{k\ell'}} \right\},$$

subject to the normalization $\sum_{\ell=0}^{L} \gamma_{k\ell} = 1$.[4] Now, because

$$-\ln(\gamma_{k\ell} + \gamma_{k\ell'}) \geq 1 - \ln(\gamma_{k\ell}^{(s)} + \gamma_{k\ell'}^{(s)}) - \frac{\gamma_{k\ell} + \gamma_{k\ell'}}{\gamma_{k\ell}^{(s)} + \gamma_{k\ell'}^{(s)}}$$

(see Hunter, 2004), we can instead maximize

$$\sum_{k=1}^{K} \sum_{\ell=0}^{L} \left(\sum_{\ell'=0}^{L} (\widetilde{n}_{k\ell\ell'}^{(s)} + n_{k\ell'\ell}) \right) \ln \gamma_{k\ell} - \sum_{k=1}^{K} \sum_{\ell=0}^{L} \sum_{\ell'=0}^{L} \left((\widetilde{n}_{k\ell\ell'}^{(s)} + n_{k\ell\ell'}) \frac{\gamma_{k\ell} + \gamma_{k\ell'}}{\gamma_{k\ell}^{(s)} + \gamma_{k\ell'}^{(s)}} \right).$$

That is (taking special care to indices),

$$\gamma_{k\ell}^{(s+1)} \propto \left(\sum_{\ell'=0}^{L} (\widetilde{n}_{k\ell\ell'}^{(s)} + n_{k\ell'\ell}) \right) \left[\sum_{\ell'=0}^{L} \frac{\widetilde{n}_{k\ell\ell'}^{(s)} + n_{k\ell\ell'} + \widetilde{n}_{k\ell'\ell}^{(s)} + n_{k\ell'\ell}}{\gamma_{k\ell}^{(s)} + \gamma_{k\ell'}^{(s)}} \right]^{-1},$$

where $X_\ell \propto Y_\ell$ means $X_\ell = Y_\ell / \sum_\ell Y_\ell$, that is $\gamma_{k\ell}^{(s+1)}$ should sum to one over $\ell \geq 0$.

[4] Notice that for $\ell = \ell' = 0$, we have an extra contribution of $(\widetilde{n}_{k00}^{(s)} + n_{k00}) \ln \frac{1}{2}$, but it does not matter because it is independent of parameters.

Update $\lambda^{(s)}$ by maximizing

$$\sum_{\ell=0}^{L}\left(\sum_{k=1}^{K}\left(\overline{n}_{k\ell}\frac{1-\lambda_{\ell}^{(s)}}{\overline{M}_{k\ell}^{(s)}}\right)\ln(1-\lambda_{\ell})+\sum_{k=1}^{K}\sum_{\ell'=0}^{L}\left(\widetilde{n}_{k\ell\ell'}^{(s)}+n_{k\ell\ell'}\right)\ln\lambda_{\ell}\right).$$

Let

$$\lambda_{\ell}^{(s+1)}=\left[\sum_{k=1}^{K}\sum_{\ell'=0}^{L}\left(\widetilde{n}_{k\ell\ell'}^{(m)}+n_{k\ell\ell'}^{(m)}\right)\right]\left[\sum_{k=1}^{K}\left(\overline{n}_{k\ell}^{(m)}\frac{1-\lambda_{\ell}^{(m)}}{\overline{M}_{k\ell}^{(m)}}\right)+\sum_{k=1}^{K}\sum_{\ell'=0}^{L}\left(\widetilde{n}_{k\ell\ell'}^{(m)}+n_{k\ell\ell'}^{(m)}\right)\right]^{-1}.$$

Finally update $\nu^{(s)}$ by maximizing

$$\sum_{\ell'=0}^{L}\left[\sum_{k=1}^{K}\sum_{\ell=0}^{L}\left(\widetilde{n}_{k\ell\ell'}^{(s)}+n_{k\ell\ell'}\right)\right]\ln\nu_{\ell'}\quad\text{s.t.}\quad\sum_{\ell'=0}^{L}\nu_{\ell'}=1.$$

That is

$$\nu_{\ell}^{(s+1)}\propto\sum_{k=1}^{K}\sum_{\ell'=0}^{L}\left[\widetilde{n}_{k\ell'\ell}^{(s,m)}+n_{k\ell'\ell}^{(m)}\right].$$

5 Variance Decomposition

5.1 Log Wage Levels

We are interested in decomposing

$$\operatorname{Var}(w_{it}|x_{it}=x)=\frac{1}{I(x)}\sum_{i=1}^{I}\sum_{t=1}^{T}\mathbf{1}\{x_{it}=x\}\,w_{it}^{2}-\left(\frac{1}{I(x)}\sum_{i=1}^{I}\sum_{t=1}^{T}\mathbf{1}\{x_{it}=x\}\,w_{it}\right)^{2}$$

for any x, and where $I(x)=\sum_{i=1}^{I}\sum_{t=1}^{T}\mathbf{1}\{x_{it}=x\}$.

The model predicts that the cross section variance of log wages is $\operatorname{Var}\mu_{k\ell}(x)+\mathbb{E}\sigma_{k\ell}(x)$, where

$$\operatorname{Var}\mu_{k\ell}(x)=\sum_{k,\ell\neq0}p_{k\ell}(x)\mu_{k\ell}(x)^{2}-\left(\sum_{k,\ell\neq0}p_{k\ell}(x)\,\mu_{k\ell}(x)\right)^{2},$$
$$\mathbb{E}\sigma_{k\ell}(x)=\sum_{k,\ell\neq0}p_{k\ell}(x)\,\sigma_{k\ell}(x),$$

where $p_{k\ell}(x)=\frac{1}{I(x)}\sum_{i=1}^{I}p_{i}(k;\beta)\sum_{t=1}^{T}\mathbf{1}\{\ell_{it}=\ell,x_{it}=x\}$ is the cross-sectional match distribution. Note that the equality between $\operatorname{Var}(\ln w_{it}|x_{it}=x)$ and $\operatorname{Var}\mu_{k\ell}(x)+\mathbb{E}\sigma_{k\ell}(x)$ is exact by construction with

$$\mu_{k\ell}(x) = \frac{\sum_{i=1}^{I} p_i(k;\beta) \sum_{t=1}^{T} \mathbf{1}\{\ell_{it} = \ell, x_{it} = x\} w_{it}}{\sum_{i=1}^{I} p_i(k;\beta) \sum_{t=1}^{T} \mathbf{1}\{\ell_{it} = \ell, x_{it} = x\}},$$

$$\sigma_{k\ell}(x) = \frac{\sum_{i=1}^{I} p_i(k;\beta) \sum_{t=1}^{T} \mathbf{1}\{\ell_{it} = \ell, x_{it} = x\} [w_{it} - \mu_{k\ell}(x)]^2}{\sum_{i=1}^{I} p_i(k;\beta) \sum_{t=1}^{T} \mathbf{1}\{\ell_{it} = \ell, x_{it} = x\}},$$

with β and the firm classification estimated using the CEM algorithm.[5]

Next, let

$$\mu_{k\ell} = a_k + b_\ell + \widetilde{\mu}_{k\ell},$$

where $a_k + b_\ell$ is the linear projection of $\mu_{k\ell}$ on the worker and firm type indicators (that is, regress $\mu_{k\ell}$ on worker and firm type dummies, weighting each (k, ℓ) observation by $p_{k\ell}$).

We can then decompose the variance of mean log wages by match type as

$$\operatorname{Var} \mu_{k\ell} = \operatorname{Var} a_k + \operatorname{Var} b_\ell + 2 \operatorname{Cov}(a_k, b_\ell) + \operatorname{Var} \widetilde{\mu}_{k\ell},$$

with

$$\operatorname{Var} a_k = \sum_{k,\ell \neq 0} p_{k\ell} a_k^2 - \left(\sum_{k,\ell \neq 0} p_{k\ell} a_k \right)^2,$$

$$\operatorname{Var} b_\ell = \sum_{k,\ell \neq 0} p_{k\ell} b_\ell^2 - \left(\sum_{k,\ell \neq 0} p_{k\ell} b_\ell \right)^2,$$

$$\operatorname{Cov}(a_k, b_\ell) = \sum_{k,\ell \neq 0} p_{k\ell} a_k b_\ell - \left(\sum_{k,\ell \neq 0} p_{k\ell} a_k \right) \left(\sum_{k,\ell \neq 0} p_{k\ell} b_\ell \right).$$

5.2 Wage Growth

We can do a similar exercise for wage growth $\Delta \ln w_i = \ln w_{i2} - \ln w_{i1}$ conditional on being employed in two consecutive periods 1 and 2:

$$\operatorname{Var}(\Delta \ln w_i) = \operatorname{Var} \mathbb{E}\left(\Delta \ln w_i | k, \ell, \ell', D\right) + \mathbb{E} \operatorname{Var}\left(\Delta \ln w_i | k, \ell, \ell', D\right),$$

where D is the j2j mobility indicator and ℓ, ℓ' denote the job types in periods 1 and 2.

[5] For example, it is easy to see that

$$\mathbb{E}\mu_{k\ell}(x) = \sum_{k,\ell \neq 0} p_{k\ell}(x)\, \mu_{k\ell}(x) = \frac{1}{I} \sum_{i=1}^{I} w_{it}.$$

We have

$$\mathbb{E}\,\mathrm{Var}\left(\Delta\ln w_i|k,\ell,\ell',D\right)=\mathbb{E}\left(\sigma_{k\ell'}-\sigma_{k\ell}\right)=\sum_{k}\sum_{\ell\neq 0}\sum_{\ell'\neq 0}\sum_{D=0}^{1}P_{k\ell\ell'D}\left(\sigma_{k\ell'}-\sigma_{k\ell}\right),$$

where

$$P_{k\ell\ell'D}=\frac{\pi_k m_{k\ell}\overline{M}_{k\ell}^{1-D}M_{k\ell\ell'}^{D}}{\sum_k\sum_{\ell\neq 0}\sum_{\ell'\neq 0}\sum_{D=0}^{1}\pi_k m_{k\ell}\overline{M}_{k\ell}^{1-D}M_{k\ell\ell'}^{D}}.$$

As for wage levels, we can write the between-variance as

$$\mathrm{Var}\,\mathbb{E}\left(\Delta\ln w_i|k,\ell,\ell',D\right)=\mathrm{Var}\left(\mu_{k\ell'}-\mu_{k\ell}\right)=\mathrm{Var}\left(\Delta a_k+\Delta b_{\ell\ell'D}+\Delta\widetilde{\mu}_{k\ell\ell'D}\right),$$

where $\Delta a_k+\Delta b_{\ell\ell'D}$ denotes the projection of $\mu_{k\ell'}-\mu_{k\ell}$ on worker type dummies and consecutive job type dummies, and $\Delta\widetilde{\mu}_{k\ell\ell'D}$ denotes the residuals (i.e. regress $\mu_{k\ell'}-\mu_{k\ell}$ on k and on job mobility indicators (ℓ,ℓ',D), weighting each observation (k,ℓ,ℓ',D) by $P_{k\ell\ell'D}$). Finally, decompose $\mathrm{Var}\left(\mu_{k\ell'}-\mu_{k\ell}\right)$ further as

$$\mathrm{Var}\left(\mu_{k\ell'}-\mu_{k\ell}\right)=\mathrm{Var}\,\Delta a_k+\mathrm{Var}\,\Delta b_{\ell\ell'D}+2\,\mathrm{Cov}(\Delta a_k,\Delta b_{\ell\ell'D})+\mathrm{Var}\,\Delta\widetilde{\mu}_{k\ell\ell'D}.$$

6 Estimation and Monte Carlo simulations

6.1 Data

We use the matched employer-employee data from Denmark from 1985-2011. Wages are reported at annual frequency and adjusted for the aggregate trend. Mobility data of workers are reported at a weekly level.

6.2 Estimator performance

We illustrate the performance of the estimator by its ability to recover model parameters use to generate simulated data from the model. For this purpose, we use a simplified estimate for all men aged 30-34 over the entire period 1985-2011 that imposes steady state on $\pi_{k\ell}$ and assumes away dependence on covariates x. The estimate is presented in Table 2. The estimation is done subject to $K=4$ and $L=4$. The simulation assumes stationarity whereas the estimator does not. Thus, part of the performance evaluation is whether the estimated $\pi_{k\ell}$ coincides with the simulation's initialization by steady state.

The simulation replicates the sample size which has 1,089,764 workers and 253,150 firms and the worker's initial state is drawn from the stationary distribution associated with the specified mobility parameters. To match the estimations in the following section, each worker in the simulation is simulated to have a 10 year employment history. We first estimate the model on the simulated data for the case of $\hat{K} = K = 4$ and $\hat{L} = L = 4$. To account for simulation noise we repeat the estimation 500 times, where each repetition simulates the data subject to a different random number generator seed and a different initial parameters guess where guesses are drawn randomly. The initial firm classification is always a uniform distribution by firm wages as described in section 3.3.

Table 3 presents the wage parameter estimates denoted by CEM $E[\mu_{k\ell}]$ and CEM $E\left[\sigma_{k\ell}^2\right]$. As can be seen, the estimator performs exceedingly well, matching the true parameters closely. The variation of the estimator over simulation repetitions is very small as well. The estimator requires enough mobility data to perform well, and as can be seen 10 years of employment histories is more than enough. The table also demonstrates the improvement associated with the firm classification updates in the CEM algorithm. The estimates denoted by EM $E[\mu_{k\ell}]$ and EM $E\left[\sigma_{k\ell}^2\right]$ show the wage parameter estimates based on the initial firm classification where the EM algorithm on the remaining model parameters has been run to full convergence. As can be seen, there is a significant discrepancy between the true model parameters and the estimates in this case.

Table 4 presents the mobility parameter estimates using the same labeling as above. Again, it is seen that the full CEM estimator performs very well and matches the true mobility parameters closely. This includes matching the initial match distribution. The EM estimates again demonstrates that the firm classification updating plays an important role in the strong performance of the estimator in that the estimated mobility parameters given the initial firm classification are substantially different from the true model parameters.

6.3 (Preliminary) Estimation Results

We restrict employment spells to include only individuals between the age of 30-50. Furthermore, we stratify the data by education level and 10 year time intervals. Education level is based on the normed number of years of education associated with the worker's highest completed degree. The low education group comprises all degrees normed to less than 12 years of education. The medium education group has a norm of exactly 12 years, and the high eduction group is any education level with a norm greater than 12 years.

Within stratification, the wage parameters are allowed to vary in age, time, and gender. Age is divided into 3 groups; ages 30-36, 37-43, and 44-50. Time is divided into 2 year periods, meaning that a 10 year stratification has 5 different time groups. With the two gender groups, $\mu_{k\ell}(x)$ and $\sigma_{k\ell}^2(x)$ depend on 30 different x groups. The mobility parameters and the initial distribution $\pi_{k\ell}$ are restricted to depend only on gender. It is straightforward to relax this assumption, subject to the identification restrictions that come with the mobility rates in the data. Finally, ν_ℓ is constrained to be same across genders consistent with a random search model where the two genders are searching in the same markets.

We set the number of worker types to $K = 4$ and the number of firm types to $L = 6$. Thus, in total a model estimate consists of 1,564 separate parameters. We create 18 different 10 year intervals (1985-1994, 1986-1995,...) each of which is stratified by education into three groups. In the following we present selected results for the 64 estimations.

6.3.1 Wages and mobility

We sort worker and firm types according to the average wage. Hence, by construction $\mu_{k\ell}$ will on average be increasing in the k and ℓ indices. In Figures 1-3 we show the μ parameters by time window and education stratifications as well as 2 particular sets of controls: 30-36 men and women in the first 2 year period of each time window. It is a robust feature that wages are increasing in firm type for all worker types, in particular, with few exceptions, the highest wage firm type is the same for all worker types. In a related point, a simple linear projection as in Abowd et al. (1999) generally provides a good approximation to the estimates of $\mu_{k\ell}$. Specifically, estimate $\hat{\mu}_{k\ell}(x) = a_k(x) + b_\ell(x) + \varepsilon_{k\ell}$, which has an average R^2 of ?? (.85) across the stratifications and controls.

Figures 4-6 show the $\gamma_{k\ell}$ estimates for the same 3 time windows as Figures 1-3 stratified by education and controlled for gender. The mobility parameters do not control for age and within time window time variation. Consistently, male workers prefer higher wage firm types to lower types in the sense of Bradley-Terry competition. Furthermore, there is agreement across male worker types about the ranking of firm types according to γ. There is also agreement across female worker types about the ranking of firm types, but there are for women significant exceptions to the result that high wage firm types are also high γ types. The γ estimates suggest a solid firm ladder structure in the data, and that this ladder is global across worker types. Furthermore, the ladder structure is strongly related to wages, more so for men than women.

To explore the ladder structure further Figures 7-9 present the expected firm type destination conditional on the type of current firm and a job-to-job move. As can be seen, the expected firm type is solidly increasing in current firm type for men, as one would expect in a classic random search job ladder model. On the whole, the same is true for women, but in a reflection of the more complex patterns in γ, there are some exceptions.

The estimation estimates the initial allocation of worker types to firm types separately from the mobility patterns. Table 1 compares compares the stationary distribution associated with the mobility parameters with the initial distribution for the case of low education men in the 1985-1994 time window. By construction, the first spell in a worker history is an employment spell. Hence, we condition the stationary distribution on employment and correct for the bias that unemployed workers enter with employment drawn from unemployment.

Table 1: Comparison of initial distribution with stationary distribution. 1985-1994, High education. Men.

$\pi_{k\ell}(male)$					
$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 4$	$\ell = 5$	$\ell = 6$
0.0253	0.0754	0.0358	0.0570	0.0683	0.0278
0.0108	0.0463	0.0153	0.0202	0.0377	0.0206
0.0157	0.0678	0.0240	0.0406	0.0522	0.0286
0.0115	0.1251	0.0216	0.0409	0.0820	0.0495
$\pi^{ss}_{k\ell}$					
$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 4$	$\ell = 5$	$\ell = 6$
0.0309	0.1046	0.0362	0.0752	0.0581	0.0220
0.0125	0.0588	0.0154	0.0268	0.0248	0.0110
0.0124	0.0487	0.0173	0.0236	0.0781	0.0614
0.0131	0.1341	0.0218	0.0407	0.0483	0.0244

As can be seen, the stationary distribution associated with the mobility patterns is quite close to the initial match distribution. We see the same close match for other education levels, gender and time windows. This is a reassuring consistency check between the mobility part of the data and the observed cross section allocations.

To further illustrate the allocation patterns, Figures 10-12 show contour plots for the initial distribution of worker types to firm types for the same time windows as the previous tables. The figures also show the correlation coefficient between types. As can be seen, if there is sorting between worker and firm types, it is modest. Figure 13 shows the evolution of the correlation coefficient between worker and firm types over time from 1985 to 2003 in both the initial match distribution and the stationary distribution based on the mobility parameters, conditional on employment. The figure is another piece of evidence between the observed cross section match distribution and the stationary distribution from the mobility patterns. It is also seen that there is a tendency toward increased positive sorting over time. This is particularly strong for women, who early in the sample tend to be negatively sorted but come to have much the same sorting pattern as men later on. Regardless of the time period, sorting remains quite modest. This is consistent with the sorting estimates in Bagger and Lentz (2013) where sorting in the Danish private sector for the 1992-2002 period is estimated at a correlation coefficient of about 6%.

6.3.2 Gender wage gap

Define the average log wage of a type x worker by,

$$\overline{\mu}(x) = \sum_{k=1}^{K} \sum_{\ell=1}^{L} \mu_{k\ell}(x)\, p_{k\ell}(x)\,.$$

Specifically, denote by $\overline{\mu}_m(x)$ and $\overline{\mu}_f(x)$ the average male and female log wage where x in this case does not include gender. We will discuss the gender gap conditional on x characteristics by $gap(x) = \exp\left(\overline{\mu}_m(x) - \overline{\mu}_f(x)\right)$. Our framework allows us to decompose this difference into an allocation effect and an a wage effect. For this purpose, define the counter factual female average log wage for the case where the allocation is set according to that of the corresponding male match allocation,

$$\overline{\mu}_f^{cf}(x) = \sum_{k=1}^{K} \sum_{\ell=1}^{L} \mu_{k\ell}(f,x)\, p_k(f,x)\, \frac{p_{k\ell}(m,x)}{p_k(m,x)},$$

where $p_{k\ell}(m,x)$ is the allocation of a male type x worker. Furthermore, $p_k(g,x) = \sum_{\ell=1}^{L} p_{k\ell}(g,x)$ for $g = \{m, f\}$. The fraction of a type x wage gap that is explained by mobility pattern differences is then $\left(\overline{\mu}_f^{cf}(x) - \overline{\mu}_f(x)\right) / \left(\overline{\mu}_m(x) - \overline{\mu}_f(x)\right)$.

In Figure 14, we show the evolution of the wage gap from the period 1985-2002, which is calculated by use of the first time interval in each of 10 year windows from 1985 to 2011. As seen, the wage gap is trending down during this time period for all age groups with the exception of older high education workers. It is also seen that women tend to move to high type firms less intensely than men. Consequently, mobility explains a significant part the wage gap. The relative contribution from mobility is the strongest for low education women.

7 Conclusion

TBC

Figures and Tables

Table 2: Parameter estimates of men, aged 30-34 with all education levels

	$\mu_{k\ell}$				$\sigma_{k\ell}$				
	$\ell=1$	$\ell=2$	$\ell=3$	$\ell=4$	$\ell=0$	$\ell=1$	$\ell=2$	$\ell=3$	$\ell=4$
k=1	3.730	4.100	4.388	4.904		1.397	0.838	0.771	0.870
$k=2$	4.100	4.256	4.421	4.592		0.555	0.253	0.258	0.273
$k=3$	4.422	4.376	4.499	4.644		0.150	0.110	0.126	0.141
$k=4$	4.714	4.617	4.784	4.972		0.245	0.200	0.210	0.228
	λ_ℓ	ν_ℓ	η_k	η_ℓ	$\gamma_{k\ell}$				
$\ell=0$	0.157	0.583							
$\ell=1$	0.022	0.128	0.146	0.233	0.100	0.069	0.042	0.124	0.666
$\ell=2$	0.018	0.132	0.311	0.302	0.097	0.062	0.056	0.164	0.621
$\ell=3$	0.033	0.110	0.313	0.256	0.005	0.047	0.028	0.143	0.777
$\ell=4$	0.090	0.047	0.231	0.209	0.018	0.034	0.014	0.113	0.821

	$\pi_{k\ell}$				
	$\ell=0$	$\ell=1$	$\ell=2$	$\ell=3$	$\ell=4$
$k=1$	0.029	0.032	0.024	0.033	0.028
$k=2$	0.056	0.056	0.063	0.085	0.050
$k=3$	0.004	0.062	0.047	0.108	0.092
$k=4$	0.013	0.038	0.020	0.074	0.085

Table 3: CEM wage parameter estimates

CEM $E[\mu^s_{k\ell}]$				CEM $Std[\mu^s_{k\ell}]$				EM $E[\mu^s_{k\ell}]$			
3.729	4.098	4.387	4.903	0.0018	0.0012	0.0009	0.0013	3.711	4.120	4.366	4.811
4.100	4.255	4.421	4.592	0.0006	0.0002	0.0002	0.0003	4.033	4.195	4.398	4.638
4.422	4.376	4.499	4.644	0.0002	0.0001	0.0001	0.0001	4.269	4.331	4.470	4.638
4.714	4.616	4.784	4.972	0.0003	0.0004	0.0002	0.0002	4.556	4.577	4.702	4.890

CEM $E\left[\sigma^s_{k\ell}\right]$				CEM $Std\left[\sigma^s_{k\ell}\right]$				EM $E\left[\sigma^s_{k\ell}\right]$			
1.396	0.837	0.771	0.869	0.0013	0.0009	0.0007	0.0008	1.200	1.016	0.753	0.788
0.555	0.253	0.258	0.273	0.0003	0.0002	0.0001	0.0002	0.644	0.538	0.379	0.392
0.150	0.110	0.126	0.141	0.0001	0.0001	0.0001	0.0001	0.355	0.283	0.211	0.233
0.245	0.200	0.210	0.228	0.0002	0.0003	0.0001	0.0002	0.261	0.243	0.226	0.246

Table 4: CEM mobility parameter estimates

CEM $E[\gamma^s_{k\ell}]$					CEM $Std[\gamma^s_{k\ell}]$				
0.102	0.075	0.044	0.126	0.652	0.0020	0.0007	0.0011	0.0018	0.0036
0.099	0.068	0.059	0.167	0.607	0.0019	0.0007	0.0015	0.0022	0.0039
0.005	0.051	0.031	0.146	0.767	0.0001	0.0004	0.0007	0.0022	0.0026
0.019	0.037	0.015	0.116	0.814	0.0004	0.0003	0.0004	0.0018	0.0023

CEM $E\left[\lambda^s_\ell\right]$					CEM $Std\left[\lambda^s_\ell\right]$				
0.156	0.022	0.018	0.032	0.086	0.0002	0.0002	0.0001	0.0002	0.0013

CEM $E\left[\nu^s_\ell\right]$					CEM $Std\left[\nu^s_\ell\right]$				
0.588	0.124	0.128	0.112	0.048	0.0027	0.0016	0.0018	0.0007	0.0004

EM $E[\gamma^s_{k\ell}]$				
0.136	0.140	0.093	0.145	0.486
0.119	0.118	0.102	0.162	0.500
0.059	0.099	0.101	0.175	0.567
0.037	0.095	0.093	0.153	0.622

EM $E\left[\lambda^s_\ell\right]$				
0.152	0.025	0.023	0.030	0.060

EM $E\left[\nu^s_\ell\right]$				
0.580	0.103	0.099	0.109	0.109

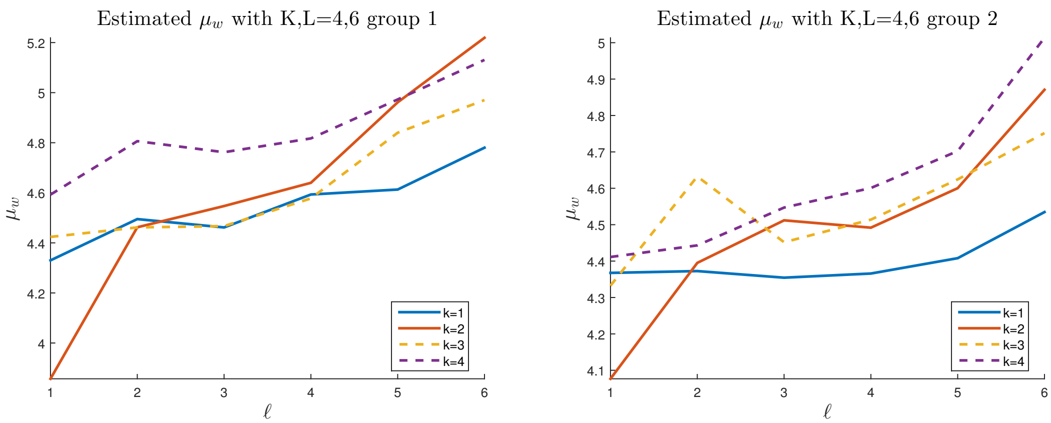
Table 5: CEM proportions of firm and worker types

CEM				EM	
$E\left[(\eta_k)^s\right]$	$Std\left[(\eta_k)^s\right]$	$E\left[(\eta_\ell)^s\right]$	$Std\left[(\eta_\ell)^s\right]$	$E\left[(\eta_k)^s\right]$	$E\left[(\eta_\ell)^s\right]$
0.146	0.0001	0.234	0.001	0.135	0.250
0.311	0.0001	0.295	0.001	0.208	0.250
0.313	0.0001	0.261	0.001	0.322	0.250
0.231	0.0001	0.209	0.000	0.335	0.250

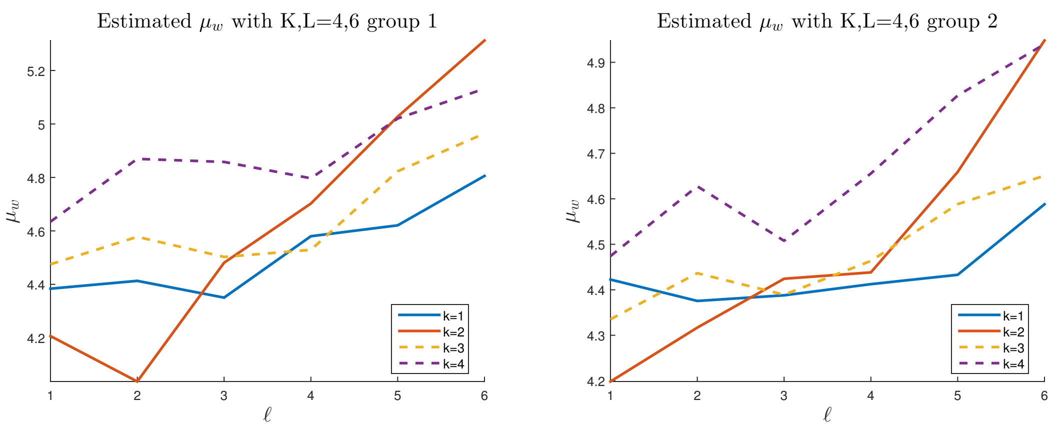
	CEM $E\left[\pi_{k\ell}\right]$					EM $E\left[\pi_{k\ell}\right]$				
	$\ell=0$	$\ell=1$	$\ell=2$	$\ell=3$	$\ell=4$	$\ell=0$	$\ell=1$	$\ell=2$	$\ell=3$	$\ell=4$
$k=1$	0.029	0.032	0.024	0.033	0.028	0.027	0.029	0.021	0.029	0.030
$k=2$	0.056	0.056	0.062	0.086	0.050	0.034	0.037	0.034	0.053	0.050
$k=3$	0.004	0.062	0.046	0.108	0.092	0.024	0.052	0.053	0.094	0.098
$k=4$	0.013	0.038	0.020	0.074	0.085	0.016	0.050	0.049	0.095	0.124

Figure 1: $\mu_{k\ell}$, High Ed Workers (men left, women right), ages 30-36

(a) 1985-1986



(b) 1993-1994



(c) 1999-2000

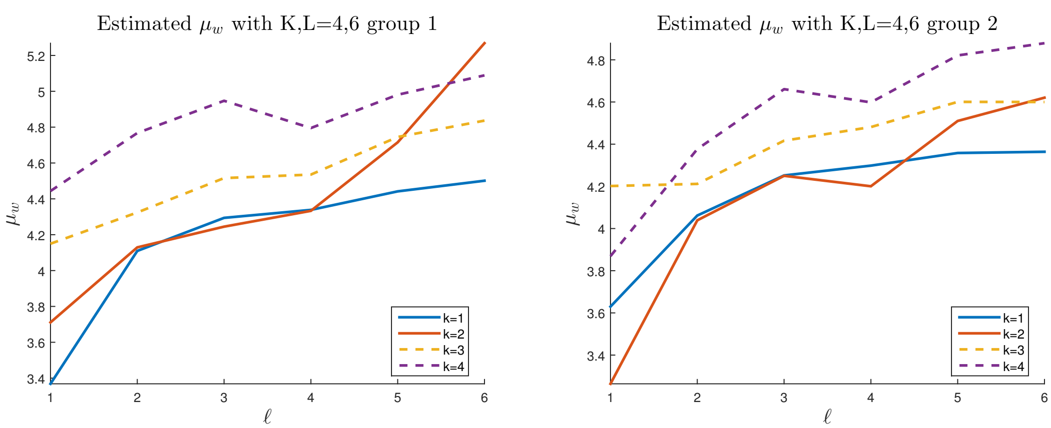
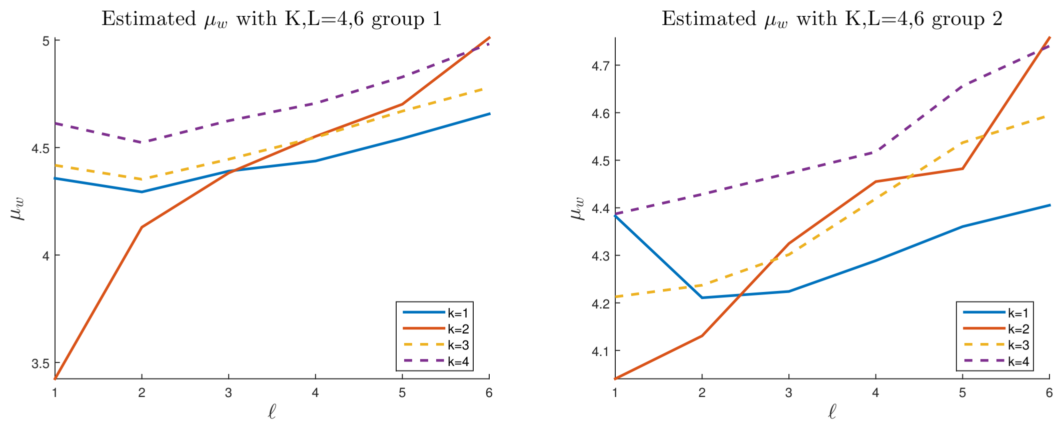
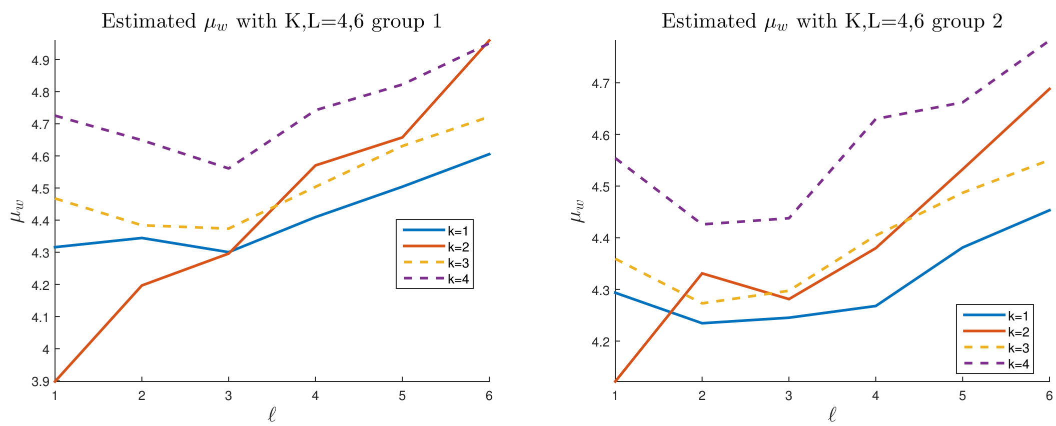


Figure 2: $\mu_{k\ell}$, Med Ed Workers (men left, women right), ages 30-36

(a) 1985-1986



(b) 1993-1994



(c) 1999-2000

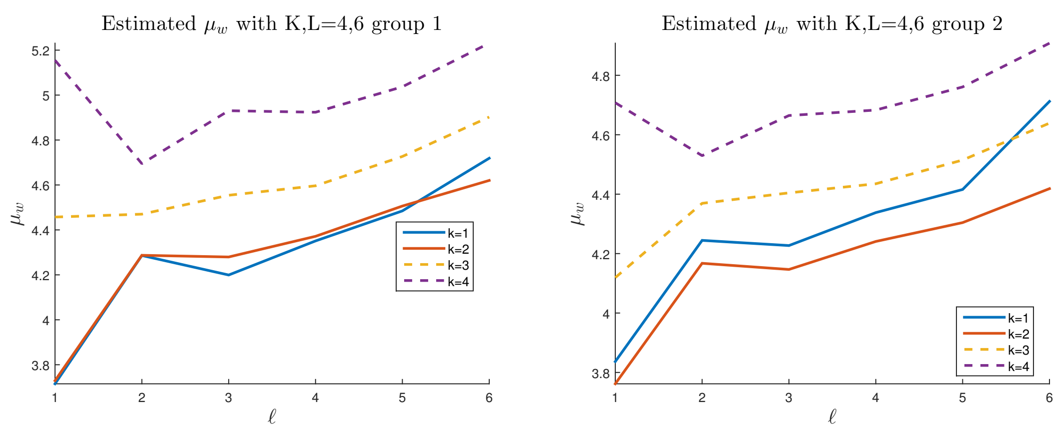
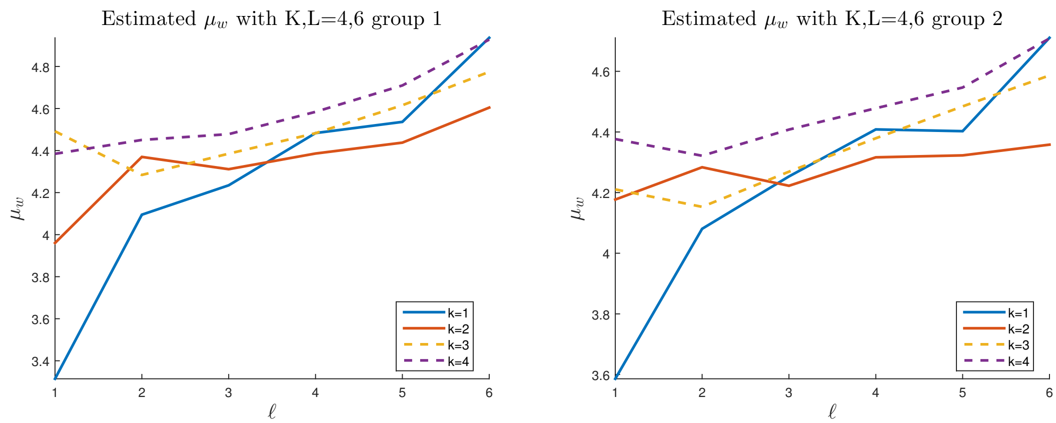
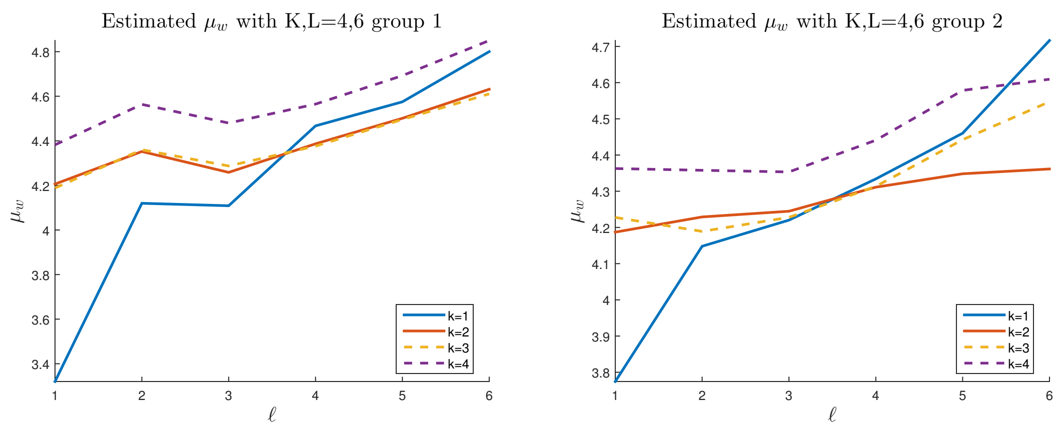


Figure 3: $\mu_{k\ell}$, Low Ed Workers (men left, women right), ages 30-36

(a) 1985-1986



(b) 1993-1994



(c) 1999-2000

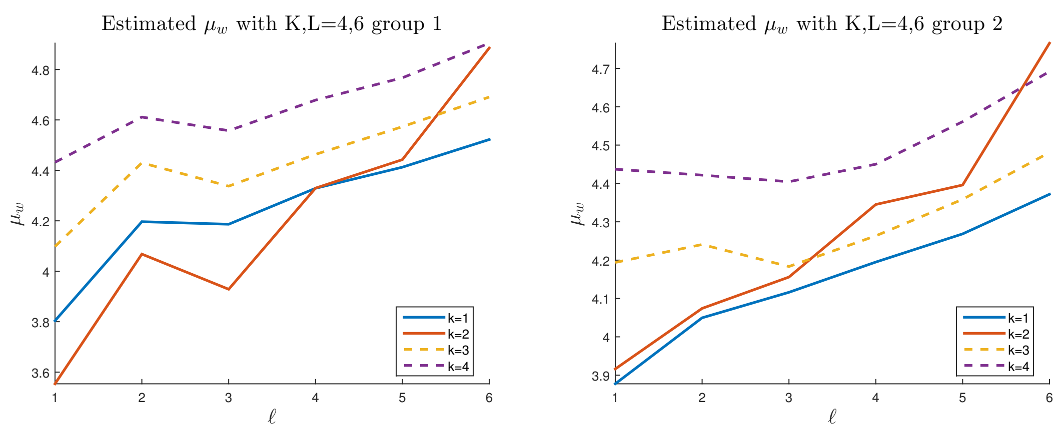
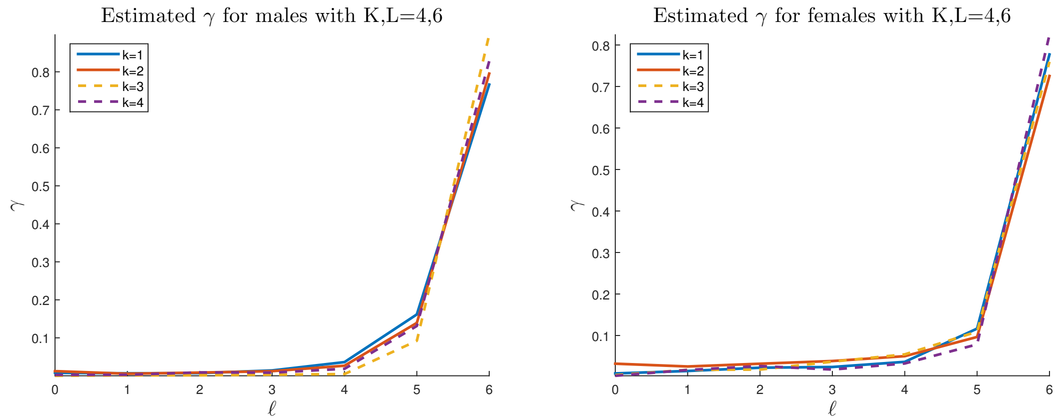
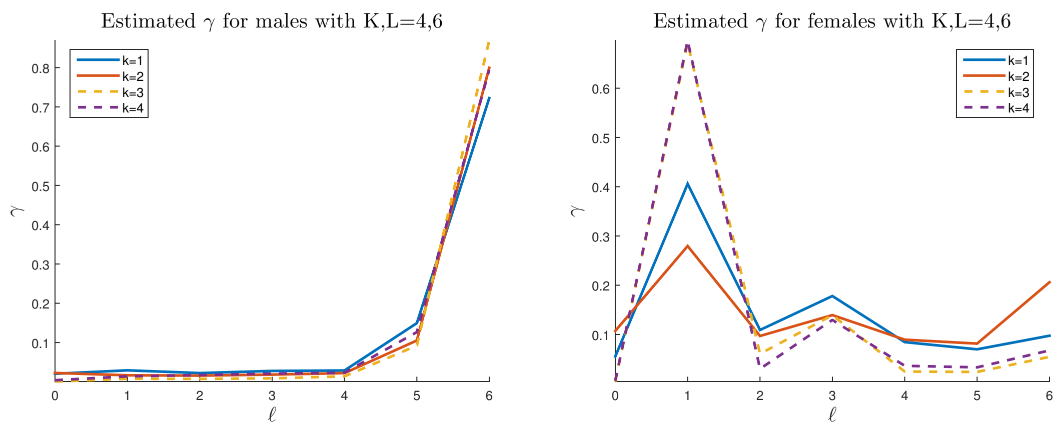


Figure 4: $\gamma_{k\ell}$, High Ed Workers

(a) 1985-1994



(b) 1993-2002



(c) 1999-2008

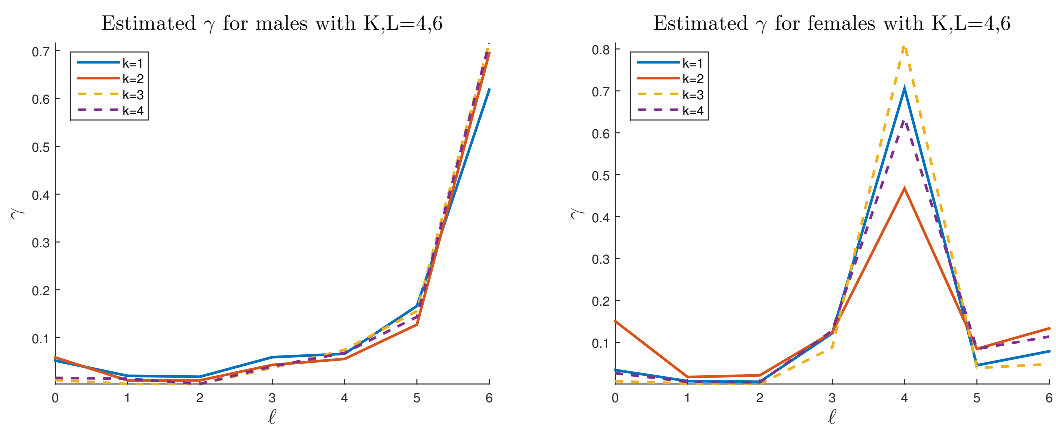
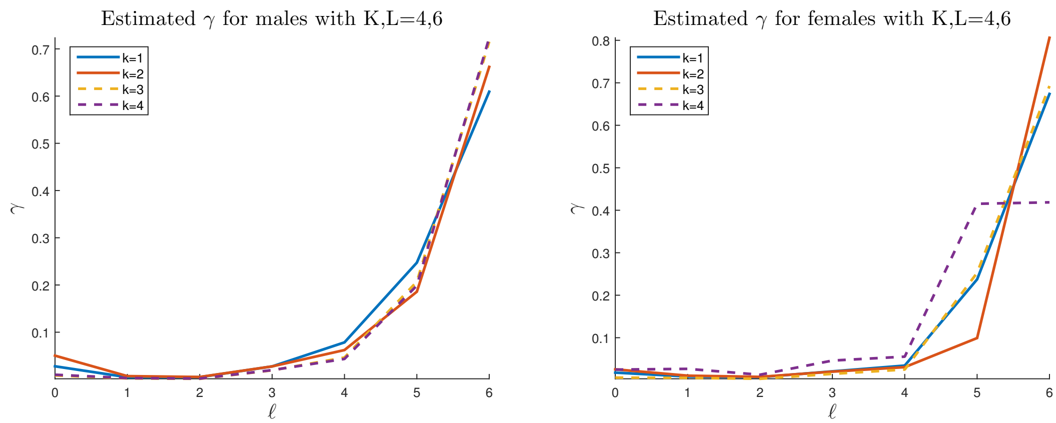
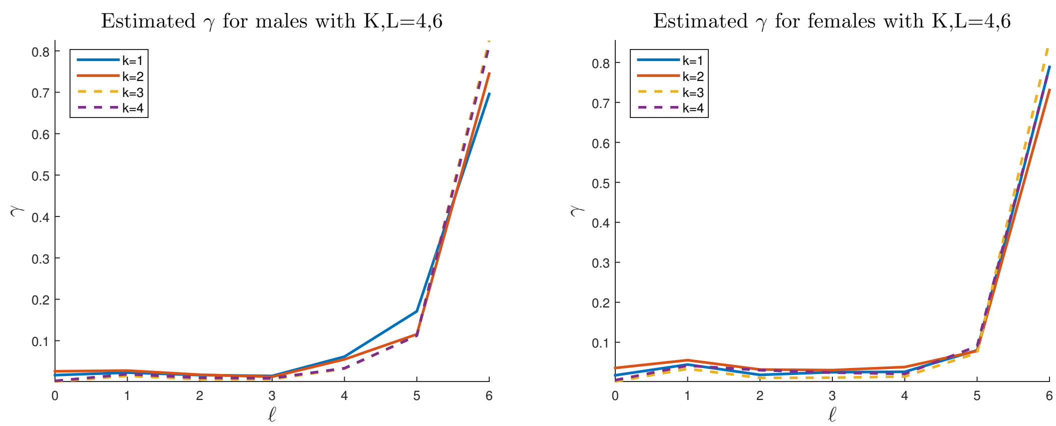


Figure 5: $\gamma_{k\ell}$, Med Ed Workers

(a) 1985-1994



(b) 1993-2002



(c) 1999-2008

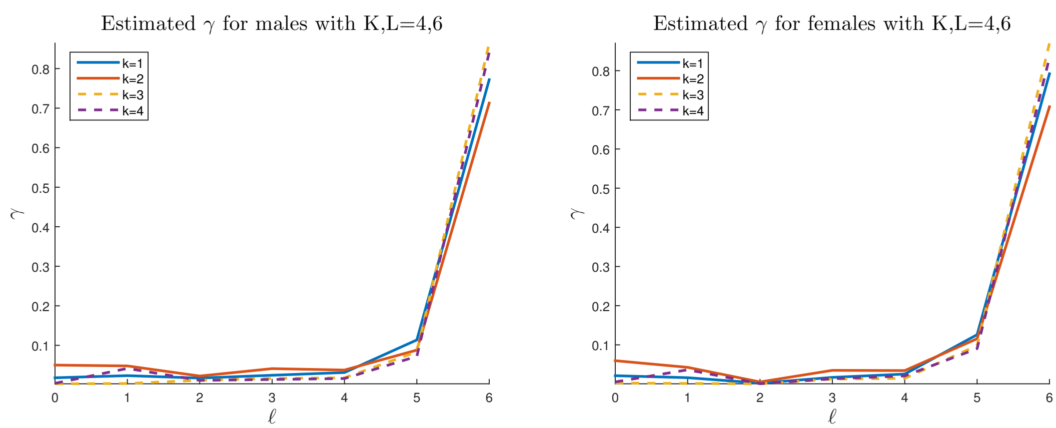
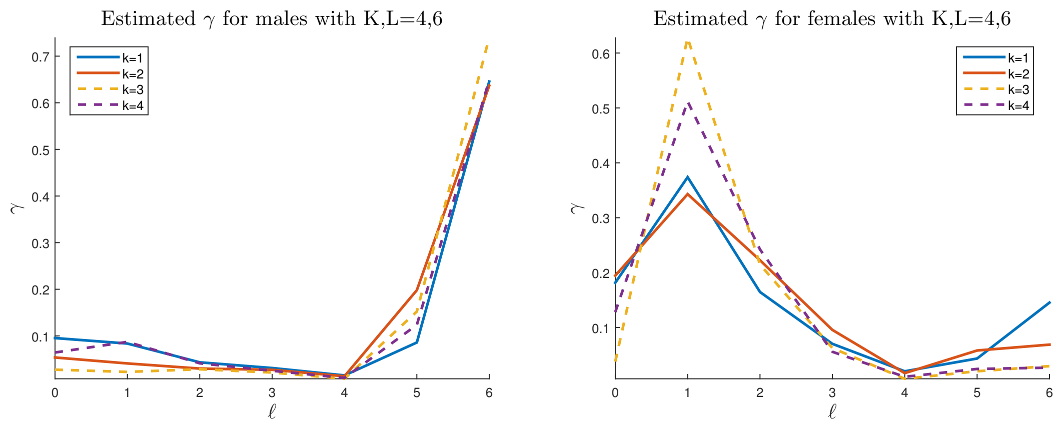
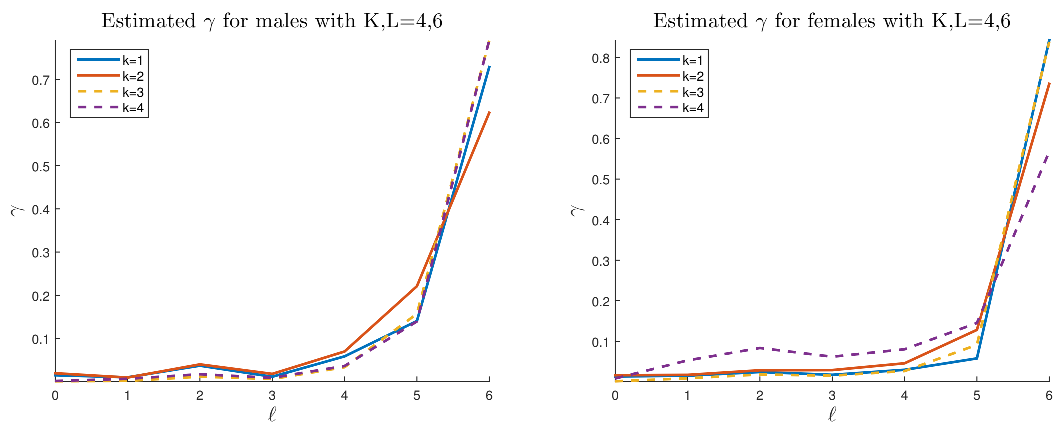


Figure 6: $\gamma_{k\ell}$, Low Ed Workers

(a) 1985-1994



(b) 1993-2002



(c) 1999-2008

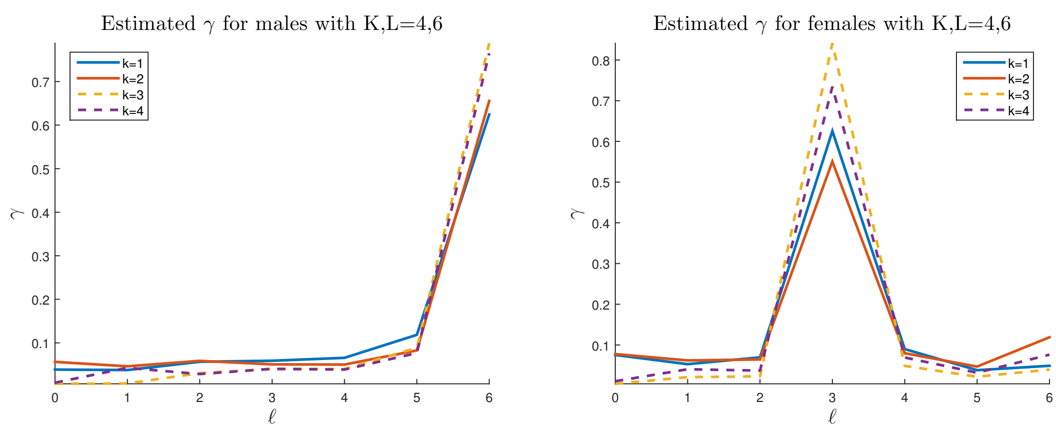
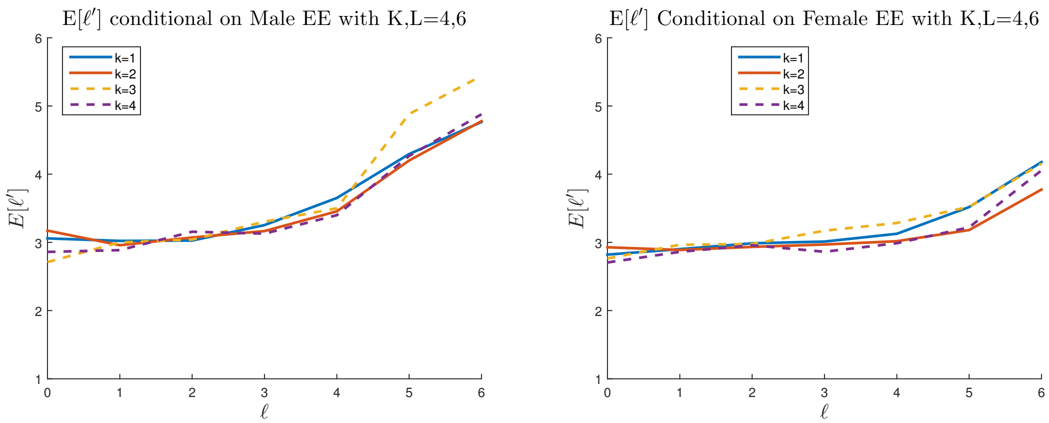
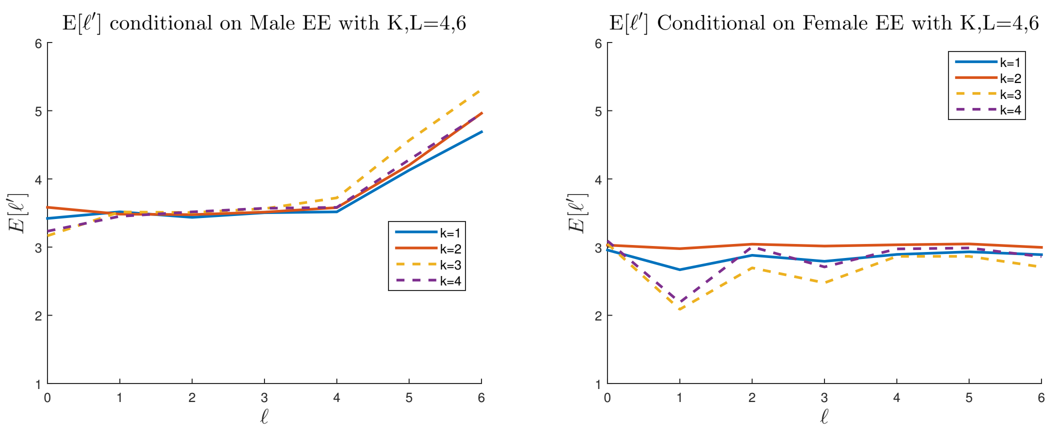


Figure 7: High Ed Workers

(a) 1985-1994



(b) 1993-2002



(c) 1999-2008

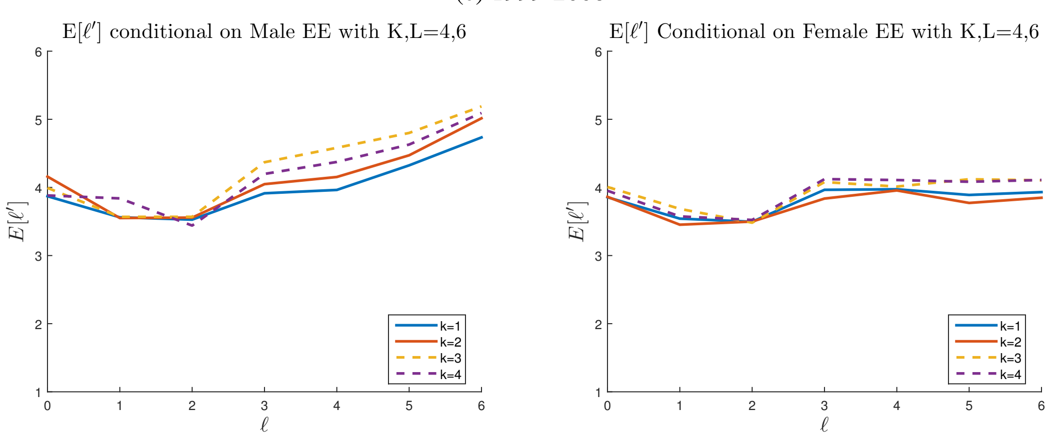
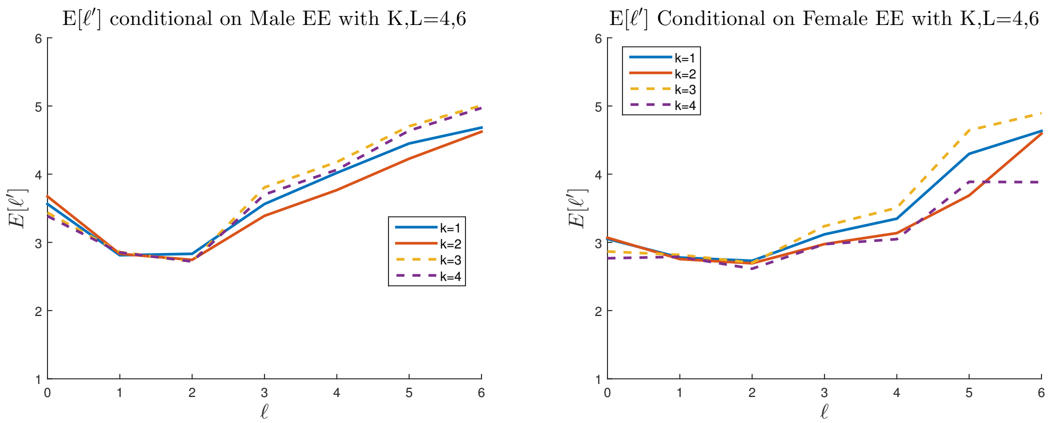
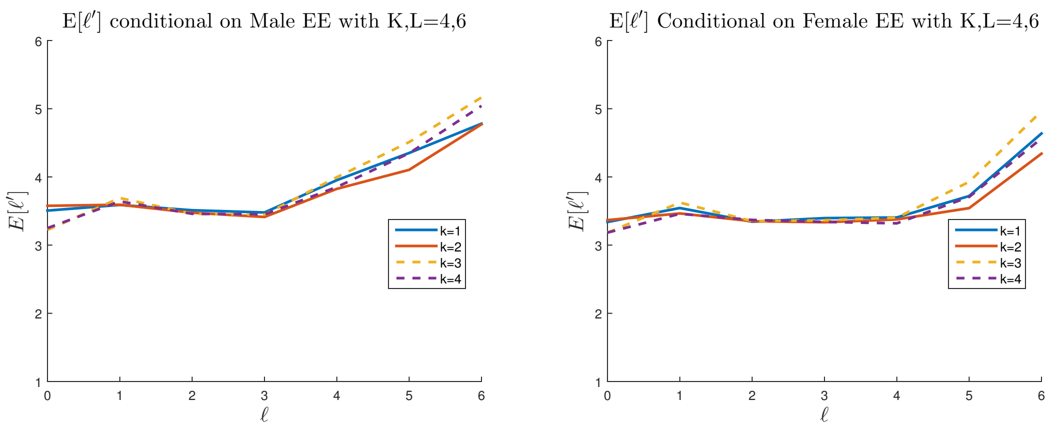


Figure 8: Med Ed Workers

(a) 1985-1994



(b) 1993-2002



(c) 1999-2008

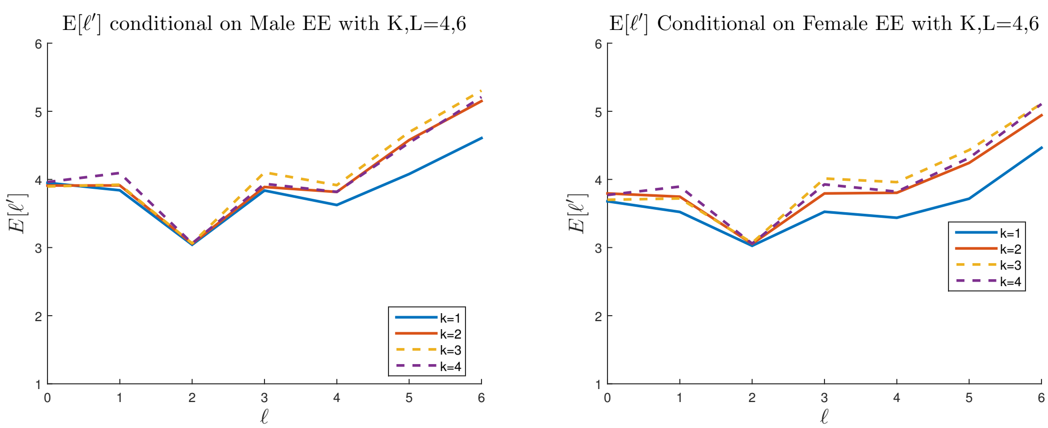
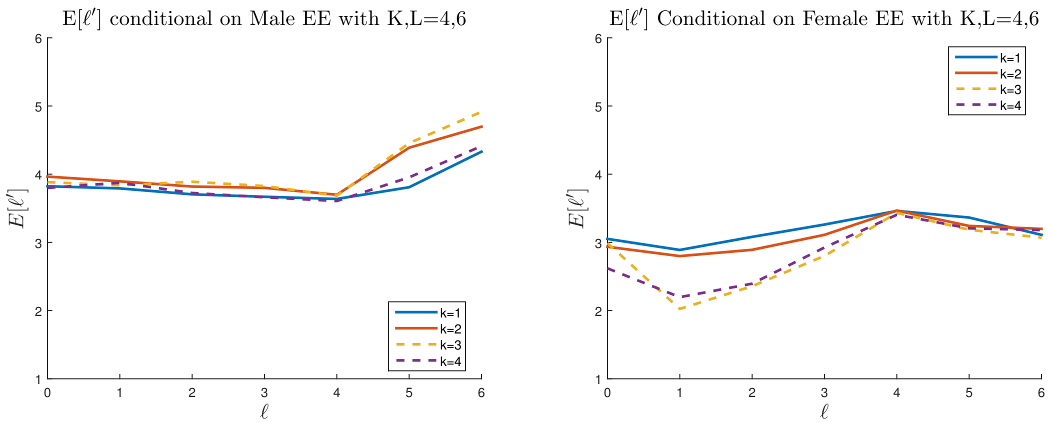
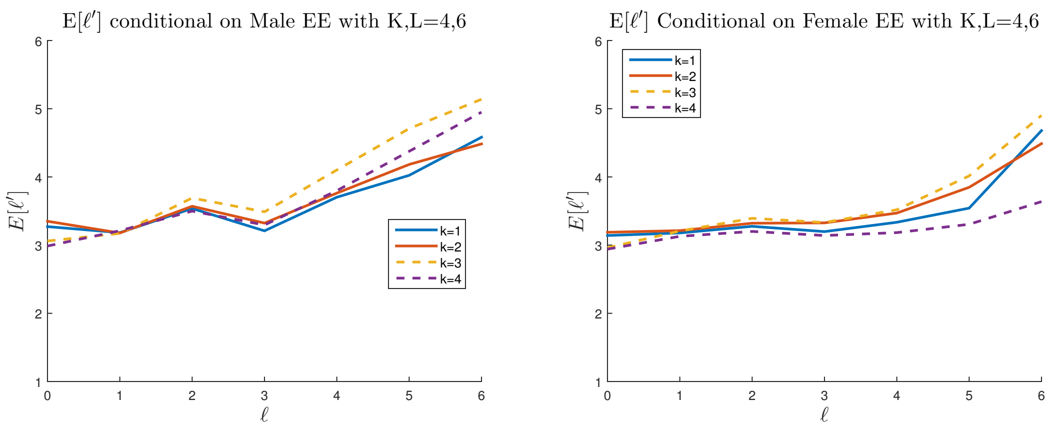


Figure 9: Low Ed Workers

(a) 1985-1994



(b) 1993-2002



(c) 1999-2008

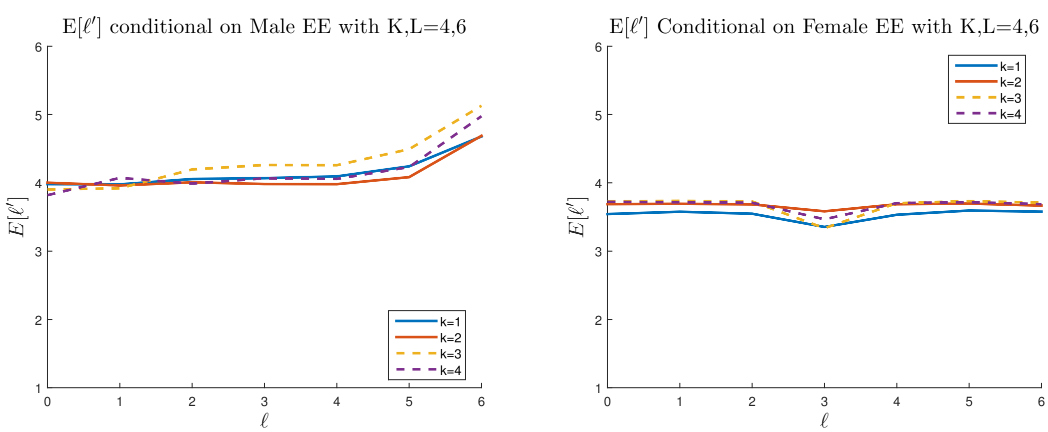
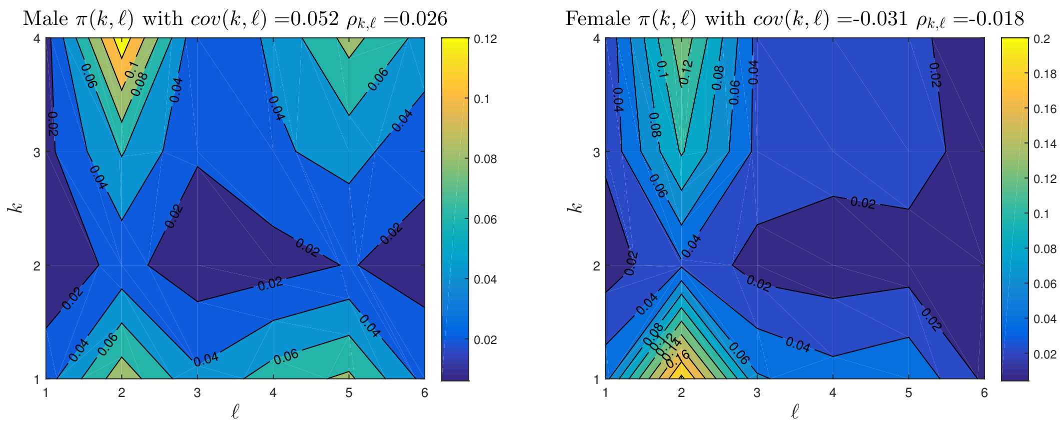
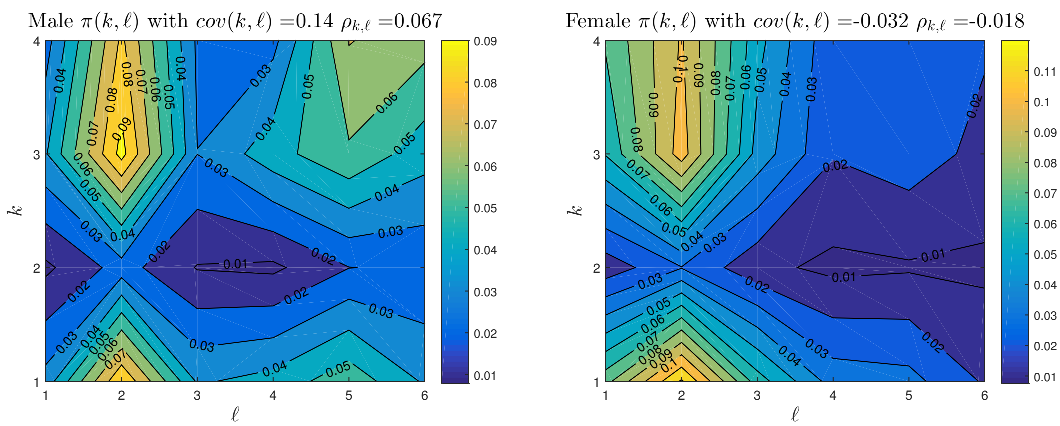


Figure 10: High Ed Workers

(a) 1985-1994



(b) 1993-2002



(c) 1999-2008

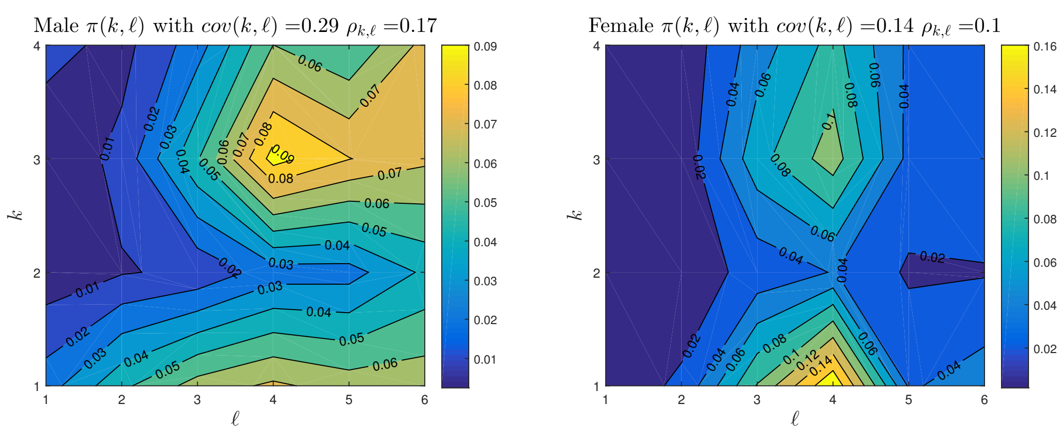
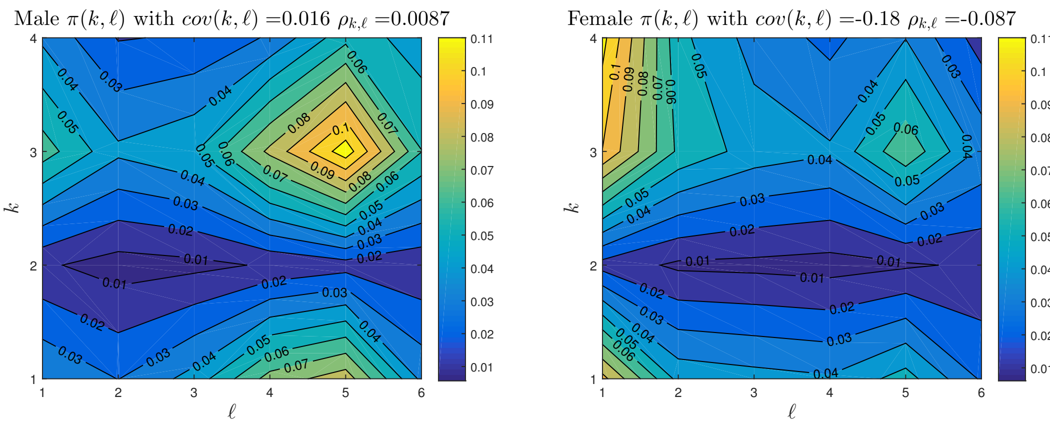
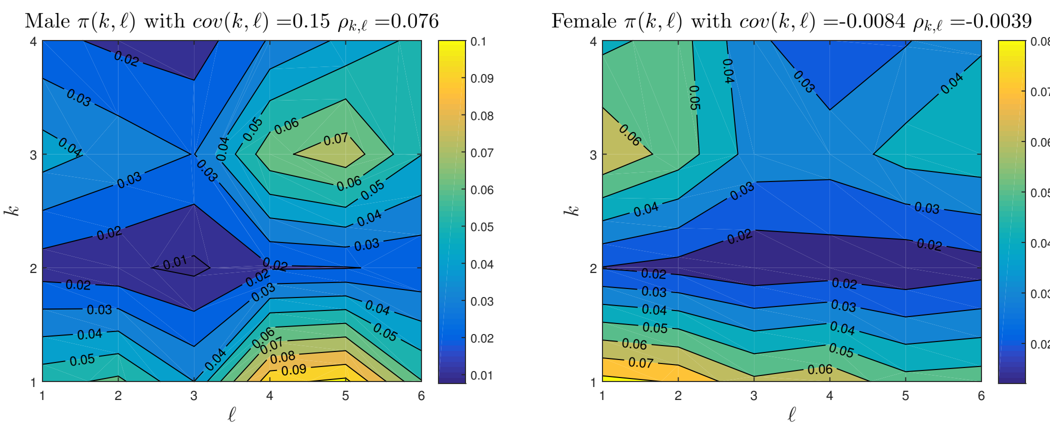


Figure 11: Med Ed Workers

(a) 1985-1994



(b) 1993-2002



(c) 1999-2008

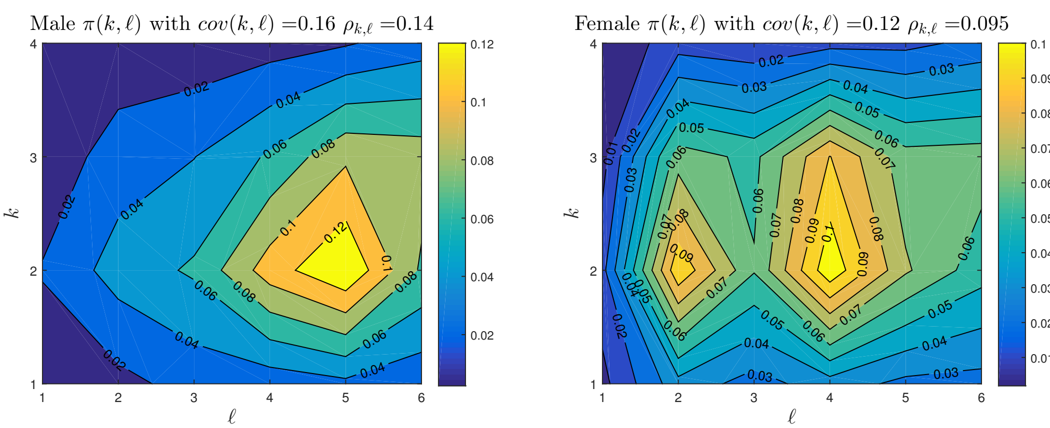
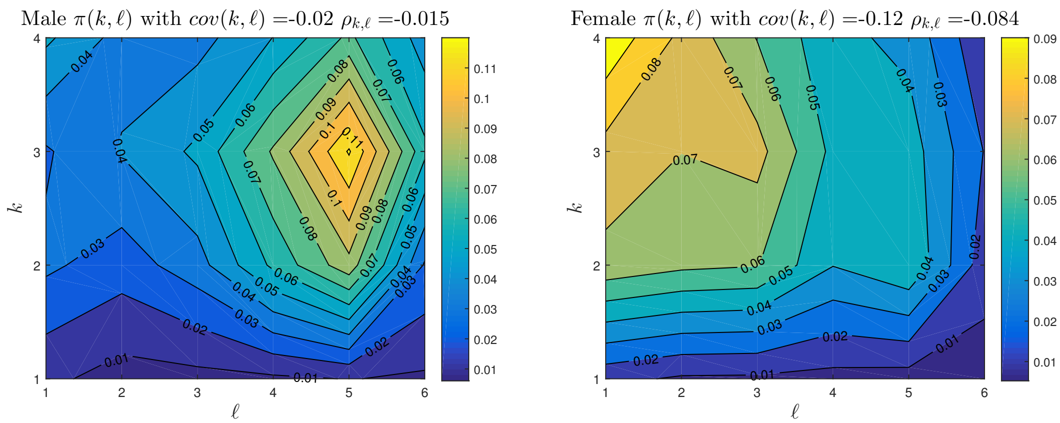
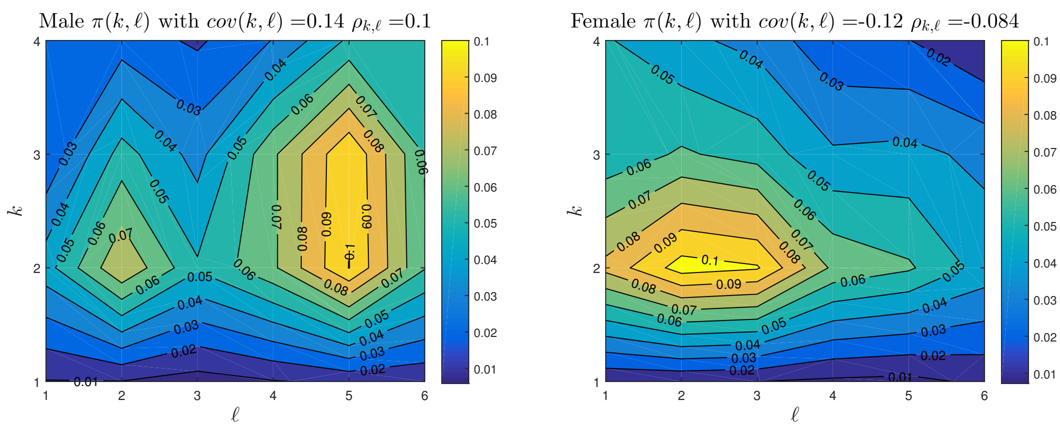


Figure 12: Low Ed Workers

(a) 1985-1994



(b) 1993-2002



(c) 1999-2008

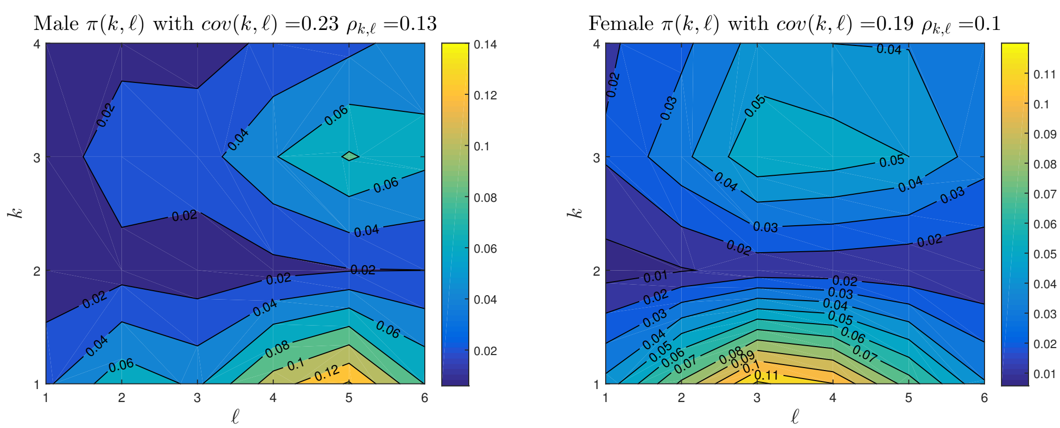
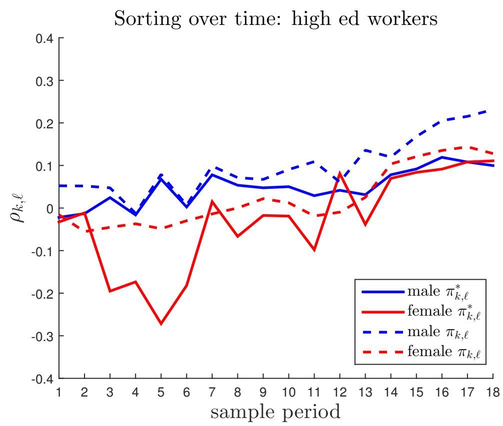
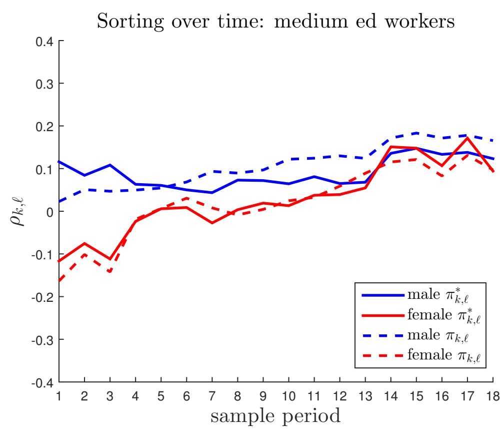


Figure 13: Sorting over time. Correlation coefficient (k, ℓ). Initial and steady state distributions.

(a) High education



(b) Medium education



(c) Low education

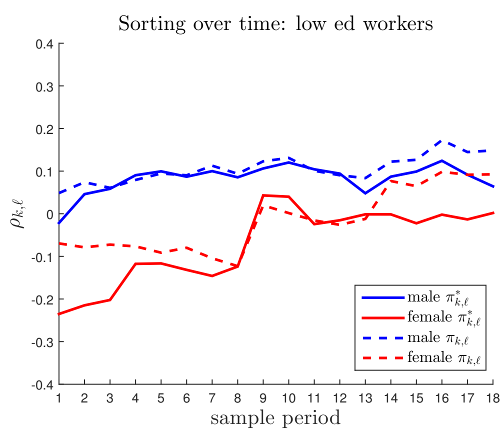
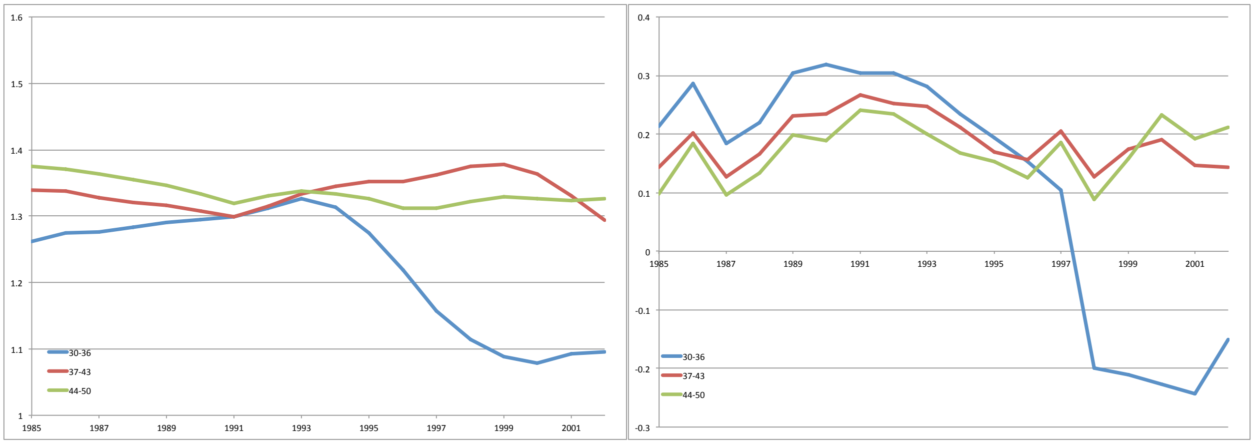


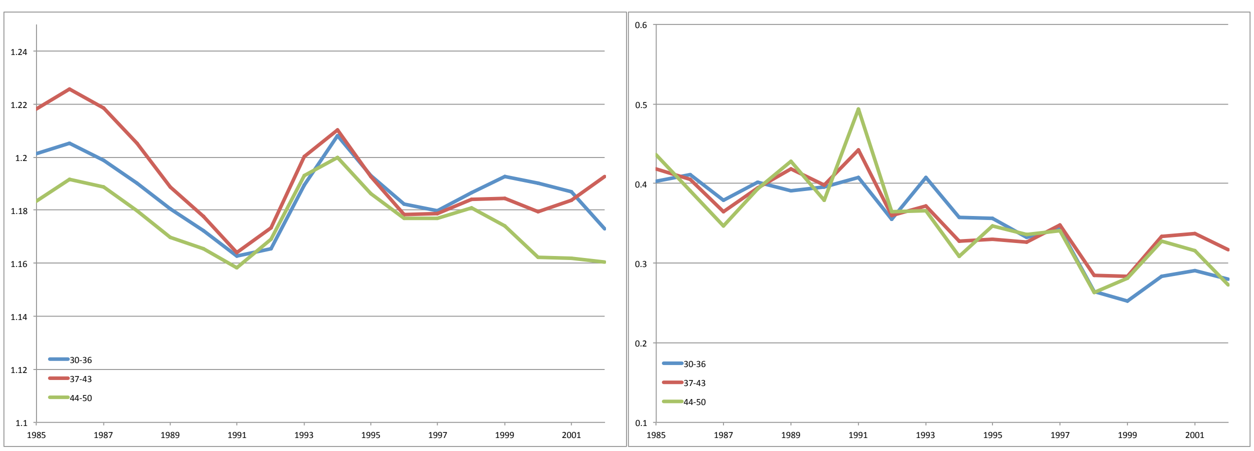
Figure 14: Gender wage gap (left) and contribution from mobility (right).

(a) Hi Education



(b) Medium Education

(c) Low Education



References

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999). High wage workers and high wage firms. *Econometrica 67*(2), 251–334.

Agresti, A. (2003). *Categorical Data Analysis*. Wiley Series in Probability and Statistics. Wiley.

Arcidiacono, P. and J. B. Jones (2003, 05). Finite Mixture Distributions, Sequential Likelihood and the EM Algorithm. *Econometrica 71*(3), 933–946.

Arcidiacono, P. and R. A. Miller (2011). Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. *Econometrica 79*(6), 1823–1867.

Bagger, J. and R. Lentz (2013). An empirical model of wage dispersion with sorting. *University of Wisconsin–Madison Working Paper*.

Bonhomme, S., T. Lamadon, and E. Manresa (2015). A Distributional Framework for Matched Employer Employee Data. mimeo, University of Chicago.

Bonhomme, S. and J.-M. Robin (2009, 01). Assessing the equalizing force of mobility using short panels: France, 1990-2000. *Review of Economic Studies 76*(1), 63–92.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological) 39*(1), pp. 1–38.

Hunter, D. R. (2004, 02). Mm algorithms for generalized bradley-terry models. *Ann. Statist. 32*(1), 384–406.