The Anatomy of the Wage Distribution

How do Gender and Immigration Matter?

Rasmus Lentz Suphanit Piyapromdee Jean-Marc Robin

UW-Madison, UCL&IFS, Sciences Po

- Matched employer-employee data provide wage outcomes in matches for different combinations of workers and firms.
- What are the sources of wage inequality between (i) men and women (ii) immigrants and natives?
- What are the determinants of wages and mobility?

Introduction

Wage models with double sided unobserved heterogeneity for MEE data.

- Abowd, Kramarz, Margolis (ECMA, 1999):

$$\ln w_{it} = x_{it}^{\top}\beta + \alpha_i + \psi_{j(i,t)} + u_{it}.$$

- Restrictive: Additive effects.
  - OLS for fixed-effects estimation: small-T biases. Addressed by Andrews, Gil, Schank, and Upward (JRSS, 2008).
- Bonhomme, Lamadon, Manresa (2015): Version of the AKM model with discrete worker and firm types.
  - Nonparametrically identified with $T = 2$.
  - Firm classification in first step by k-means. Then identify discrete mixture over worker side by EM algorithm.

Our contribution

- We adopt a discrete mixture approach as in BLM.
- Contrary to BLM we use a parametric model of wage distributions and job-to-job mobility
  - Do women have lower wages because they have different offer arrival probability or because they rank jobs differently?
- These transition probabilities are nonlinear but we develop an MM-algorithm.
- Our estimation algorithm uses both wage and mobility data to *update firm classification.*
- We present Monte Carlo analyses and estimation on Danish match employer-employee data.

Data observation

- Workers: $i \in \{1, ..., I\}$. Firms: $j \in \{1, ..., J\}$. $j = 0$ denotes non-employment.
- Observation for worker i in week t, (w_{it}, j_{it}, x_{it}), $t = 1, 2, \ldots, T_i$.
  - $j_{it} \equiv j(i, t) \in \{0, 1, ..., J\}$ is the ID of the firm employing worker i in week t.
  - x_{it} are observed worker controls.
  - w_{it} is the workers wage rate at time t.

Classification

- Firms clustered into L different groups indexed by $\ell \in \{0, 1, \ldots, L\}$, where non-employment is $\ell = 0$.
- Workers clustered into K different groups indexed by $k \in \{1, ..., K\}$.
- Unobserved firm types ℓ are treated as fixed effects (parameters).
- Unobserved worker types k are treated as random effects, drawn from a distribution.
- Let $\pi = (\pi_1, ..., \pi_K)$ denote the proportions of workers' types in the population.

Wages

- Let $f_{k\ell}(w)$ denote the wage density, conditional on worker type k and employer type ℓ.
- Wages are assumed lognormal,

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

- We assume w_{it} and $w_{it'}$ of a type k worker are independent conditional on $\left(\ell_{j_{it}}, \ell_{j_{it'}}\right)$.
- Straightforward to include controls in $f_{kl}\left(w\right)$.

Mobility

- $M_{k\ell\ell'}$ is probability that a type k worker moves from a type ℓ to type ℓ' firm.
- $\overline{M}_{k\ell} = 1 - \sum_{\ell'=0}^{L} M_{k\ell\ell'}$ is probability of staying with same employer.
- $m_{k\ell}$ is probability that a type k worker is matched with a type ℓ firm.
- Assume stationarity,

$$m_{k\ell} = m_{k\ell}\overline{M}_{k\ell} + \sum_{\ell'=0}^{L} m_{k\ell'} M_{k\ell'\ell}, \ \forall (k,\ell),$$

  This allows us to make inference about mobility parameters from worker flows and the initial allocation of workers to jobs.

Empirical specification

Transition probabilities

- The probability that a type k transitions from a type ℓ to a type ℓ' firm is

$$M_{k\ell\ell'} = \lambda_\ell \nu_{\ell'} P_{k\ell\ell'}, \quad \lambda_\ell, \nu_{\ell'} \in [0,1], \quad \sum_{\ell'=0}^{L} \nu_{\ell'} = 1$$

- λ_ℓ probability of a meeting with an outside employer when the current firm type is ℓ.
- $\nu_{\ell'}$ probability that the outside draw is of type ℓ'.
- $P_{k\ell\ell'}$ probability that the transition from ℓ to ℓ' becomes effective where $\gamma_{k\ell}$ measures the quality of the match (k,ℓ). We assume a Bradley-Terry specification (Hunter's ,2004 MM-algorithm) with $P_{k00} = 0$ and

$$P_{k\ell\ell'} = \frac{\gamma_{k\ell'}}{\gamma_{k\ell} + \gamma_{k\ell'}}, \quad \sum_{\ell=0}^{L} \gamma_{k\ell} = 1.$$

Likelihood with observed firm types and stationarity

- Let ℓ_{it} be the type of worker i's employer in period t.
- Indicate an employer change between t and $t+1$ by,

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

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

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

where $L_i(k; \beta, \mathcal{L})$ is the worker type k individual likelihood,

$$L_i(k; \beta, \mathcal{L}) = \pi_k m_{k\ell_{i1}} \prod_{t=1}^{T_i - 1} f_{k\ell_{it}}(w_{it}) \prod_{t=1}^{T_i - 1} \overline{M}_{k\ell_{it}}^{1-D_{it}} M_{k\ell_{it}\ell_{i(,t+1)}}^{D_{it}},$$

EM elements

- Given $(\beta, \mathcal{L})$, the posterior probability that worker i is type k is

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

- Worker i's expected wage log-likelihood,

$$Q_i(f; \beta, \mathcal{L}) = \sum_{k=1}^{K} p_i(k; \beta, \mathcal{L}) \left[\sum_{t=1}^{T_i} \ln f_{k\ell_{it}}(w_{it}) \right]$$

and worker i's expected mobility log-likelihood,

$$H_i(M; \beta, \mathcal{L}) = \sum_{k=1}^{K} p_i(k; \beta, \mathcal{L}) \left[\ln m_{k\ell_{i1}} + \sum_{t=1}^{T_i - 1} \left\{ (1 - D_{it}) \ln \overline{M}_{k\ell_{it}} + D_{it} \ln M_{k\ell_{it}\ell_{i(t+1)}} \right\} \right].$$

A CEM algorithm

C-step First, rank firms by average wage divide the firms equi-proportionately into L groups.

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

M-step Update $\beta^{(m)}$ by maximizing $\sum_i \sum_k p_i\left(k; \beta^{(m)}, \mathcal{L}\right) \ln L_i(k; \beta, \mathcal{L})$ subject to $\sum_k \pi_k = 1$, that is

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

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

$$\pi_k^{(m+1)} = \frac{1}{I} \sum_{i=1}^{I} p_i(k; \beta^{(m)}, \mathcal{L}).$$

Firm re-classification (C-step)

- Disregarding the part of the likelihood that involves mobility between firms, we suggest to update $\mathcal{L}^{(m)}$ separately for each firm j as,

$$
\begin{aligned}
\ell_j^{(m+1)} = \arg\max_{\ell_j} \Bigg\{ & \sum_{i=1}^{I} \sum_{k=1}^{K} p_i\left(k; \beta^{(m)}, \mathcal{L}^{(m)}\right) \sum_{\substack{t=1 \\ (j(i,t)=j)}}^{T_i} \ln f_{k\ell_j}^{(m+1)}(w_{it}) \\
& + \sum_{i=1}^{I} \sum_{k=1}^{K} p_i\left(k; \beta^{(m)}, \mathcal{L}^{(m)}\right) \sum_{\substack{t=1 \\ (j(i,t)=j)}}^{T_i-1} \Big\{ (1 - D_{it}) \ln \overline{M}_{k\ell_j}^{(m+1)} + \\
& D_{it} \left[\mathbb{1}\{\ell_{j(i,t+1)} = 0\} \ln M_{k\ell_j 0}^{(m+1)} + \mathbb{1}\{\ell_{j(i,t+1)} \neq 0\} \ln \sum_{\ell'=1}^{L} M_{k\ell_j\ell'}^{(m+1)} \right] \Big\} \Bigg\}.
\end{aligned}
$$

- Iterate between C-E-M until we have convergence.

M-step details

Wage distribution

- Given normality assumption, the M-step update takes the following form for wage parameters:

$$\mu_{k\ell}^{(m+1)} = \frac{\sum_{i=1}^{I} p_i(k;\beta^{(m)}) \sum_{t=1}^{T_i} \mathbb{1}\{j(i,t)=\ell\} w_{it}}{\sum_{i=1}^{I} p_i(k;\beta^{(m)}) \sum_{t=1}^{T_i} \mathbb{1}\{j(i,t)=\ell\}}$$

$$\sigma_{k\ell}^{(m+1)} = \frac{\sum_{i=1}^{I} p_i(k;\beta^{(m)}) \sum_{t=1}^{T_i} \mathbb{1}\{j(i,t)=\ell\} [w_{it} - \mu_{k\ell}^{(m+1)}]^2}{\sum_{i=1}^{I} p_i(k;\beta^{(m)}) \sum_{t=1}^{T_i} \mathbb{1}\{j(i,t)=\ell\}}.$$

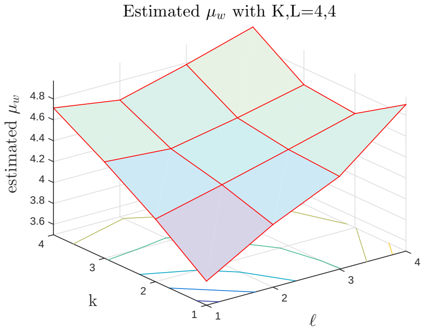
- Mobility parameters are estimated using the Minorize-maximization (MM) algorithm
  - update λ, ν, γ using formulas and maximize the mobility likelihood until convergence.

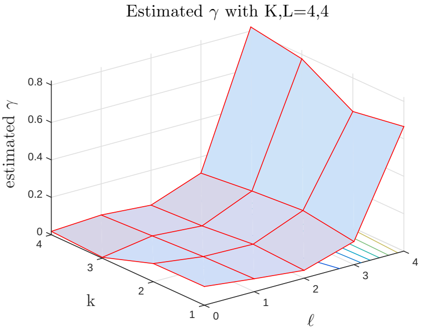
Preliminary estimates
Data

- 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 frequency.

Preliminary estimates

All men, aged 30-34, 5 years



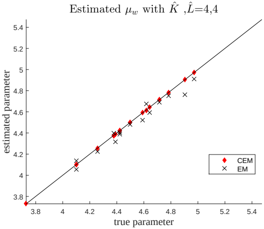


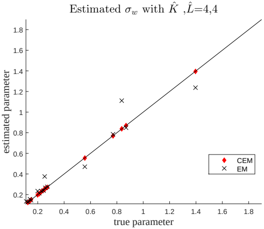
- 1,089,764 workers and 253,150 firms
- In this case, clear evidence of a firm ladder.

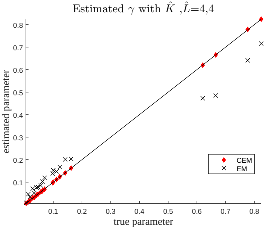
Monte Carlo simulations

- We illustrate the performance of the estimator by its ability to recover the estimated parameters of men aged 30-34.
- The sample has 1,089,764 workers and 253,150 firms, 5 years/260 weeks. Not a toy simulation!
- Each estimation takes about 25 minutes (FORTRAN, some parallelization).
- Compare iterated CEM to EM based on initial firm classification.

Simulation with firm proportions $\eta_\ell = \{0.1, 0.5, 0.3, 0.1\}$

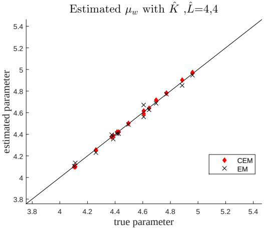


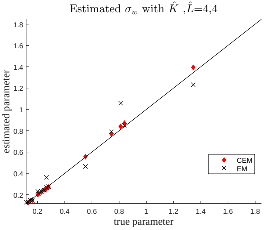


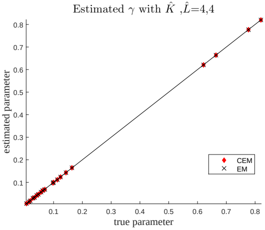


After about 4 iterations of the C-step, the proportion of misclassified firms reduces from **46 to 3%**.

Simulation with firm proportions $\eta_\ell = \{0.23, 0.30, 0.26, 0.21\}$

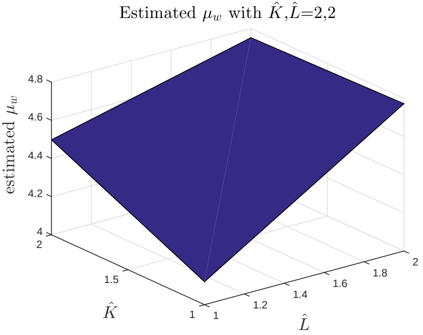


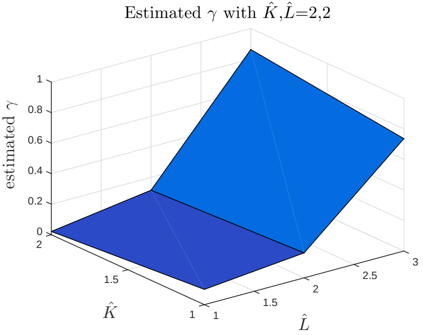




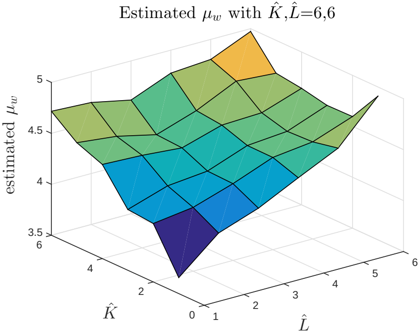
- The initial wage clustering results in 24% of misclassified firm types.
- After about 4 iterations of the CEM algorithm, this proportion of misclassified firm types reduces to 1% along with an improvement in the likelihood function.

Wrong number of types $K = L = 2$ (1 simulation)





Wrong number of types $K = L = 6$ (1 simulation)



When $\hat{K} = \hat{L} = 6$, the algorithm assigns very small probabilities to type 4 workers and type 2 firms.

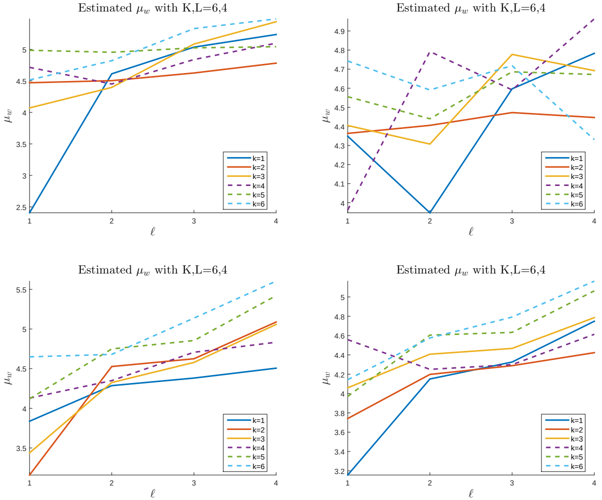
Stratified estimates

- Model estimation done flexibly with respect to observable worker characteristics:
  - Focus on working age (30-50)
  - Stratified by gender and education: High (>12 yrs), Medium (=12 yrs), Low (<12 yrs).
  - Split data into five-year periods 1985-1989, 1990-1994, ...

Mean wage of high ed workers

Top panel: 1985-1989 and bottom panel: 2004-2009

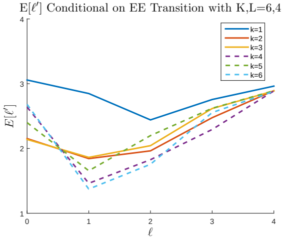
Male | Female

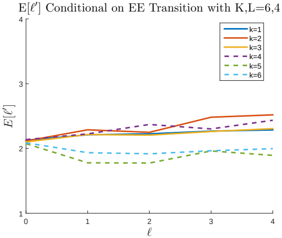


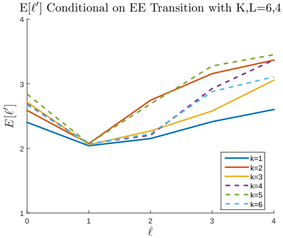
Mobility of high ed workers

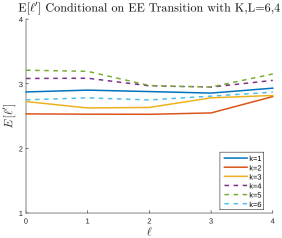
Top panel: 1985-1989 and bottom panel: 2004-2009

Male | Female





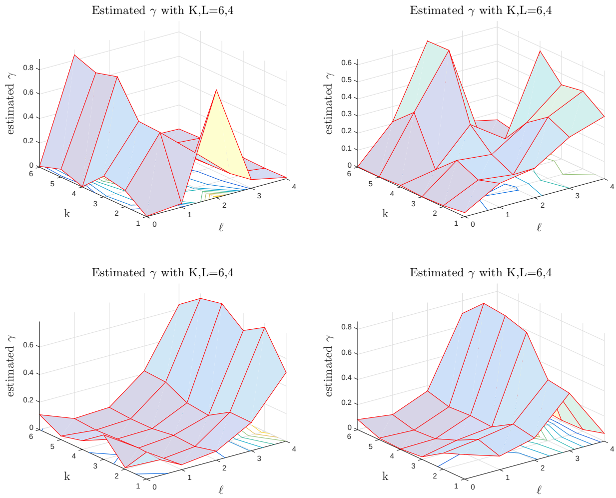




Values of jobs for high ed workers

Top panel: 1985-1989 and bottom panel: 2004-2009

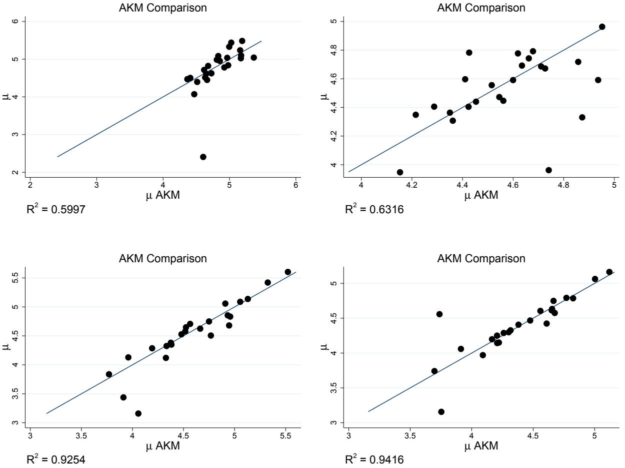
Male | Female



Male workers: regression of μ_w on k and ℓ dummies, no interaction

Top panel: 1985-1989 and bottom panel: 2004-2009

Male | Female



Concluding remarks

- Following discrete mixture approach in BLM, we present a CEM algorithm for the flexible estimation of wage and mobility parameters.
    - Fast estimation of mobility parameters from MM algorithm in M-step.
    - C-step improves performance of estimator.
- Main Findings
    - Some evidence of “memory” coming out of nonemployment.
    - Men more positively sorted over time than women.
    - Complementarity between firm and worker types for both males and females.
    - Mobility patterns show evidence of upward movement on firm ladder for male workers.

The EM algorithm itself is based on minorization of the individual likelihood in the point $\beta^{(m)}$,

$$L_i(\beta, \mathcal{L}) = \ln \sum_{k=1}^{K} L_i(k; \beta, \mathcal{L}) \geq \hat{L}_i(\beta; \beta^{(m)}, \mathcal{L}),$$

with equality for $\beta = \beta^m$, where

$$\hat{L}_i(\beta; \beta^{(m)}, \mathcal{L}) = L_i(\beta^{(m)}, \mathcal{L}) + \sum_{k=1}^{K} p_i(k; \beta^{(m)}, \mathcal{L}) \ln \frac{L_i(k; \beta, \mathcal{L})}{L_i(k; \beta^{(m)}, \mathcal{L})},$$

which provides the algorithm where the log-likelihood of the data always improves from step m to $(m+1)$,

$$\beta^{(m+1)} = \arg\max_{\beta} \sum_{i=1}^{I} \sum_{k=1}^{K} p_i(k; \beta^{(m)}, \mathcal{L}) \ln L_i(k; \beta, \mathcal{L}).$$

back

Minorize-maximization galore

M-step, transition mobilities - 1

- Update mobility parameters based on mobility part of $\ln L_i(k;\beta,\mathcal{L})$ (excluding steady state). Count stays and moves to rewrite as,

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

where

$$\overline{n}_{k\ell}^{(m)} = \sum_i p_i(k;\beta^{(m)})\,\#\left\{t : D_{it}=0, \ell_{it}^{(m)}=\ell\right\},$$

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

back

Minorize-maximization galore

M-step, transition mobilities - 2

- Find minorization that is additive in controls. Using Jensen's inequality, probability of staying is minorized in $M^{(s)}$ by,

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

- This provides log-additivity. Still need to deal with $\ln P_{k\ell'\ell} = \ln \gamma_{k\ell'} - \ln(\gamma_{k\ell} + \gamma_{kl'})$. For this, use Hunter (2004) minorization,

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

back

Minorize-maximization galore

M-step, transition mobilities - 3

- From this $M^{(m+1)}$ is found through an MM-algorithm where updates are,

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

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

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

- Use this as long as log-likelihood is improved. Otherwise, use numerical optimizer to improve on $\sum_{i=1}^{I} H_i(M;\beta^m,\mathcal{L}^m)$.