It Ain't Where You're From, It's Where You're At..

Firm Effects, State Dependence, and the Gender Wage Gap

Patrick Kline, Raffaele Saggio, Mikkel Sølvsten

October 12, 2019

Abowd, Kramarz, and Margolis (1999)

The AKM model of wage determination

$$y_{it} = \alpha_i + \psi_{j(i,t)} + X'_{it}\beta + \varepsilon_{it}$$

where $j(i,t)$ gives identity of worker i's current employer

Partitions log-wages y_{it} at time t into:

- Time varying covariates X_{it} (age, year, etc)
- Perfectly transferable "person effect"α_i
- Non-transferable "firm effect" $\psi_{j(i,t)}$

Can it be so simple?

AKM loosely motivated by wage posting models (e.g., Burdett and Mortensen, 1998) that feature a stable *wage ladder*

- Wages depend on worker type α and the "rung" ψ of the ladder
- Irrelevant how one gets to that rung

Is there a stable firm wage ladder? Reasons to be skeptical:

- Workers may be willing to take a wage *cut* to move to more productive firms (Postel-Vinay and Robin, 2002)
- Longer run effects of initial conditions (Beaudry and Dinardo, 1991; Von Wachter and Bender, 2006; Oreopolous et al., 2012)

Policy relevance: recent bans on inquiring about past salary

- Concern that gender differences in mobility / bargaining have cumulative effects

A state-dependent ladder

Consider simple dynamic generalization of AKM model (DAKM)

$$y_{it} = \alpha_i + \underbrace{\psi_{j(i,t)}}_{\text{“where you're at”}} + \underbrace{\lambda_{\ell(i,t)}}_{\text{“where you're from”}} + X_{it}'\beta + \varepsilon_{it}$$

where $\ell(i,t)$ gives identity of worker i's *previous* employer or labor force state (e.g., unemployment, NILF)

- $\lambda_{\ell(i,t)}$ is *partially* transferable component
- Interpretations of λ:
  - Reputation effects (Gibbons and Katz, 1992; Gibbons et al., 2005)
  - Implicit contracts (Beaudry and Dinardo, 1991, 1995)
  - Counter-offers / sequential wage bargaining (Postel-Vinay and Robin, 2002; Cahuc et al., 2006)

Some prior estimates of state-dependent models

Random effects: Postel-Vinay and Robin (PVR, 2002); Cahuc et al. (2006); Bagger et al. (2014)

- Joint restrictions across mobility and wage equations
- Assumes 1-factor model:

$$(\psi_j, \lambda_j) = (\psi(p_j), \lambda(p_j))$$

Group fixed effects: Bonhomme, Lamadon, and Manresa (2019)

- 10 firm types identified from clustering on cross-section
- Non-separable contemporaneous + separable lagged effects
- Allow endogenous mobility (including future firm effects)
- Result 1: statistically significant lagged effects
- Result 2: Modest improvement in fit when moving from static model ($R^2 = 74.9\%$) to dynamic model ($R^2 = 77.9\%$)

Today

Assess importance of $\lambda_{\ell(i,t)}$ for "poaching" wages

- Allow each firm to be its own 2-D (ψ, λ) type
- Avoid modeling within-match wage dynamics
- Treat non-employment and N(Y)ILF as separate lagged "firms"
- Discuss conditions under which $\left(\psi_{j(i,t)}, \lambda_{\ell(i,t)}\right)$ separately identified
- Unbiased variance decompositions ala KSS (2018)
- Relate poaching wage from non-employment to Urate

Decomposition of gender gap in hiring wages:

- How much would gender gap shrink if women came *from* the same firms as men?
- How much would gender gap shrink if women were *at* the same firms as men?

Empirical conclusion: It ain't where you're from

It's where you're at..

A motivating framework: PVR (2002)

- Postel-Vinay and Robin (2002, IER) introduced the sequential auction model of worker poaching
- Empirical adaptation in PVR (2002, ECTA)

Model primitives:

- Workers have flow utility over wages $U(w)$
- Worker productivity type ε
- Firm productivity type $p \sim \Gamma(\cdot)$
- Sampling dist is $F(\cdot)$
- Marginal productivity of a match is εp

Market Structure

- Random on the job search ala BM (1998)
- Firms make take it or leave it offers of piece-rate contracts (price per unit of output εp)
- Complete information: firm knows the worker's res wage
  - Reservation wage depends on prod of current employer (if any)
- Incumbent employer can respond which leads to 2nd price auction
  - Worker goes to whichever firm is more productive
- Without heterogeneity, wage would follow a 2-pt distribution:
  - a wage for workers hired from unemployment
  - a wage for those hired from other jobs

Poaching wages

- Value of unemployment: $V_0(\varepsilon)$
- Value of employment: $V(\varepsilon, w, p)$ (p influences wage *growth*)
- Unemployed worker of type ε upon making contact with firm of type p, will be offered monopsony wage $\phi_0(\varepsilon, p)$ obeying:

$$V(\varepsilon, \phi_0(\varepsilon, p), p) = V_0(\varepsilon)$$

- When contacted by outside firm with productivity $p' > p$ will move to new firm and be paid "poaching wage" $\phi\left(\varepsilon, p, p'\right)$ obeying:

$$V\left(\varepsilon, \phi\left(\varepsilon, p, p'\right), p'\right) = V(\varepsilon, \varepsilon p, p)$$

- When contacted by outside firm with productivity $p' < p$ will be paid "retention wage" $\phi\left(\varepsilon, p', p\right)$ provided this does not involve a wage cut

PVR show that:

$$U\left(\phi\left(\varepsilon,p,p'\right)\right)=U\left(\varepsilon p\right)-\kappa\int_{p}^{p'}\bar{F}\left(x\right)\varepsilon U'\left(\varepsilon x\right)dx$$

where $\bar{F}\left(x\right)=1-F\left(x\right)$ and $\kappa=\frac{\lambda_1}{\rho+\delta+\mu}$ is fn of offer arrival, discount rate, etc. If $U\left(x\right)=\ln x$ then poaching wage can be written:

$$\ln\phi\left(\varepsilon,p,p'\right)=\underbrace{\ln\varepsilon}_{\text{person type}}+\underbrace{\ln p}_{\text{poached firm type}}\underbrace{-\kappa\int_{p}^{p'}\bar{F}\left(x\right)\frac{dx}{x}}_{\text{type upgrade}}$$

- Poaching wage is decreasing in the productivity gap between poaching and poached firms (compensating diff)
- Same eq for hires from unemp which is just a “firm” w/ prod b

Link to DAKM

By fundamental theorem of calculus $\int_p^{p'} \bar{F}(x) \frac{dx}{x} = H(p') - H(p)$, where $H(p) = \int_0^p \bar{F}(x) \frac{dx}{x}$. Thus we can write the reduced form

$$\ln \phi\left(\varepsilon, p, p'\right) = \underbrace{\ln \varepsilon}_{=\alpha} + \underbrace{\left(-\kappa H(p')\right)}_{=\psi(p')} + \underbrace{\ln p + \kappa H(p)}_{=\lambda(p)}$$

- Evolution of poaching wages driven by movement along 1-dimensional *productivity ladder*
- Limiting case: $\kappa = 0$ (workers are myopic) starting wages only depend on productivity of previous firm
- Contemporaneous and lagged effects of given firm are *negatively* dependent: $\frac{d\psi(p)}{dp} < 0$ while $\frac{d\lambda(p)}{dp} > 0$.

A Simplified Example

Suppose uniform dist of productivity, so that $\bar{F}(x) = 1 - x$. Hence,

$$\int_p^{p'} \bar{F}(x) \frac{dx}{x} = \ln p' - \ln p - \left(p' - p\right).$$

- Resulting poaching wage is:

$$\ln \phi\left(\varepsilon, p, p'\right) = \underbrace{\ln \varepsilon}_{\alpha} + \underbrace{\kappa \left(p' - \ln p'\right)}_{\psi(p')} + \underbrace{(1+\kappa) \ln p - \kappa p}_{\lambda(p)}$$

- Uniform $\Gamma(\cdot)$ implies $Corr(\psi(p), \lambda(p)) < -.98$!

Dynamics of hiring wages

Suppose we have wage data for n workers who switch jobs at least once

- Let $t \in \{1, 2, 3\}$ index job #
- Y_{it} gives first full-time wage ("poaching wage") at t'th job
- Suppressing covariates X_{it}, DAKM model is:

$$y_{it} = \alpha_i + \psi_{j(i,t)} + \lambda_{\ell(i,t)} + \varepsilon_{it} \qquad (i = 1, \ldots, n,\ t = 1, 2, 3)$$

- Difference across jobs to eliminate worker effects:

$$\begin{aligned} y_{it} - y_{it-1} = \psi_{j(i,t)} - \psi_{j(i,t-1)} + \lambda_{\ell(i,t)} - \lambda_{\ell(i,t-1)} \\ + \varepsilon_{it} - \varepsilon_{it-1} \end{aligned}$$

Three career paths

Let N denote "not yet in LF," J a new full-time job, and U an intervening spell of non-employment. We consider three labor market histories and associated wage changes between last two jobs:

- NJJJ (2 consecutive job-2-job transitions after entry)

$$y_{i3} - y_{i2} = \psi_{j(i,3)} - \psi_{j(i,2)} + \lambda_{j(i,2)} - \lambda_{j(i,1)} + \varepsilon_{i3} - \varepsilon_{i2}$$

Three career paths

Let N denote "not yet in LF," J a new full-time job, and U an intervening spell of non-employment. We consider three labor market histories and associated wage changes between last two jobs:

- NJJJ (2 consecutive job-2-job transitions after entry)

$$y_{i3} - y_{i2} = \psi_{j(i,3)} - \psi_{j(i,2)} + \lambda_{j(i,2)} - \lambda_{j(i,1)} + \varepsilon_{i3} - \varepsilon_{i2}$$

- NJUJJ (1 JUJ + 1 J2J)

$$y_{i3} - y_{i2} = \psi_{j(i,3)} - \psi_{j(i,2)} + \lambda_{j(i,2)} - \lambda_{U} + \varepsilon_{i3} - \varepsilon_{i2}$$

Three career paths

Let N denote "not yet in LF," J a new full-time job, and U an intervening spell of non-employment. We consider three labor market histories and associated wage changes between last two jobs:

- NJJJ (2 consecutive job-2-job transitions after entry)

$$y_{i3} - y_{i2} = \psi_{j(i,3)} - \psi_{j(i,2)} + \lambda_{j(i,2)} - \lambda_{j(i,1)} + \varepsilon_{i3} - \varepsilon_{i2}$$

- NJUJJ (1 JUJ + 1 J2J)

$$y_{i3} - y_{i2} = \psi_{j(i,3)} - \psi_{j(i,2)} + \lambda_{j(i,2)} - \lambda_{U} + \varepsilon_{i3} - \varepsilon_{i2}$$

- NJJ- (2 lifetime jobs, 1 J2J)

$$y_{i2} - y_{i1} = \psi_{j(i,2)} - \psi_{j(i,1)} + \lambda_{j(i,1)} - \lambda_{N} + \varepsilon_{i2} - \varepsilon_{i1}$$

Pooled model

Pooling these three cases together, we have a single wage change per individual (growth in hiring wage between last two jobs):

$$\begin{aligned}\Delta y_i &= \Delta\psi_i + \Delta\lambda_i + \Delta\varepsilon_i \\ &= \Delta F_i'\psi + \Delta L_i'\lambda + \Delta\varepsilon_i\end{aligned}$$

Conditioning on employment history $\mathcal{F} = \{(\Delta F_i, \Delta L_i)\}_{i=1}^n$, all uncertainty derives from the independent errors $\{\Delta\varepsilon_i\}_{i=1}^n$ which obey

$$\mathbb{E}[\Delta\varepsilon_i \mid \mathcal{F}] = 0 \qquad \text{(exogenous mobility)}$$

$$\mathbb{E}[\Delta\varepsilon_i^2 \mid \mathcal{F}] = \sigma_i^2 \qquad \text{(heteroscedasticity)}$$

Decomposing wage growth

Under exogenous mobility, a variance decomposition of wage growth is

$$\mathbb{E}\left[\mathbb{V}_n\left[\Delta y_i\right] \mid \mathcal{F}\right] = \underbrace{\mathbb{V}_n[\Delta\psi_i]}_{\Delta\text{dest eff}} + \underbrace{\mathbb{V}_n[\Delta\lambda_i]}_{\Delta\text{origin eff}} + 2\,\underbrace{\mathbb{C}_n[\Delta\psi_i, \Delta\lambda_i]}_{\text{trajectory}} + \mathbb{E}\left[\mathbb{V}_n\left[\Delta\varepsilon_i\right] \mid \mathcal{F}\right]$$

where $\mathbb{V}_n\left[\cdot\right]$ denotes sample variation across *workers*

Decomposing wage growth

Under exogenous mobility, a variance decomposition of wage growth is

$$\mathbb{E}\left[\mathbb{V}_n\left[\Delta y_i\right] \mid \mathcal{F}\right] = \underbrace{\mathbb{V}_n[\Delta\psi_i]}_{\Delta\text{dest eff}} + \underbrace{\mathbb{V}_n[\Delta\lambda_i]}_{\Delta\text{origin eff}} + 2\underbrace{\mathbb{C}_n[\Delta\psi_i, \Delta\lambda_i]}_{\text{trajectory}}$$
$$+\,\mathbb{E}\left[\mathbb{V}_n\left[\Delta\varepsilon_i\right] \mid \mathcal{F}\right]$$

where $\mathbb{V}_n\left[\cdot\right]$ denotes sample variation across *workers*

Also of interest are *firm*-level components:

$$\mathbb{V}_n[\psi_j], \quad \mathbb{V}_n[\lambda_j], \quad \mathbb{C}_n[\psi_j, \lambda_j]$$

- PVR implies $\mathbb{C}_n[\psi_j, \lambda_j] \ll 0$ due to forward looking behavior / compensating diffs
- We weight these variances by firm size

Identification

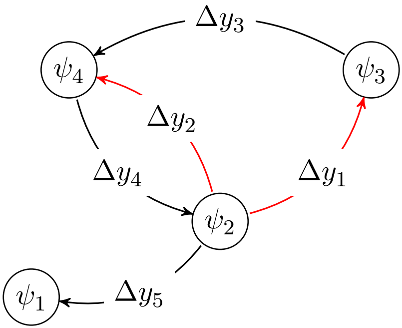
Model of wage growth

$$\Delta y_i = \Delta\psi_i + \Delta\lambda_i + \Delta\varepsilon_i$$

Identification of the variance decomposition requires that $\Delta\psi_i$ and $\Delta\lambda_i$ are separately identified

- Recall that AKM is identified by worker mobility that forms a network of paths between firms
- DAKM involves both contemporaneous and lagged mobility networks
- Identification of DAKM from worker mobility that forms paths on one network and cycles on the other
- Generalization of classic "pairwise differencing" arguments (e.g., Ahn and Powell, 1993)

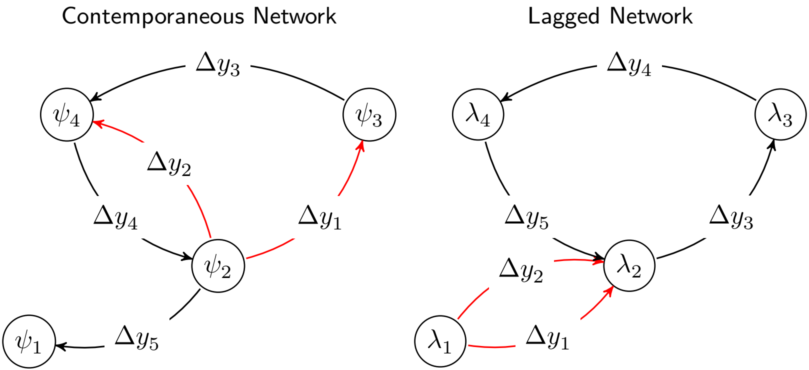
Identification of AKM



The pairwise difference $\Delta y_1 - \Delta y_2$ forms a path from ψ_4 to ψ_3:

$$\begin{aligned} \mathbb{E}[\Delta y_1 - \Delta y_2 \mid \mathcal{F}] &= \psi_3 - \psi_2 - (\psi_4 - \psi_2) \\ &= \psi_3 - \psi_4 \end{aligned}$$

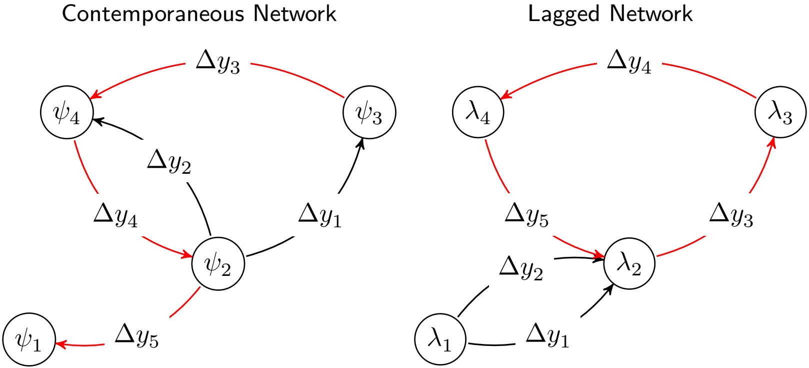
Identification of DAKM



Here, $\Delta y_1 - \Delta y_2$ forms a cycle on the lagged network

$$\begin{aligned}\mathbb{E}[\Delta y_1 - \Delta y_2 \mid \mathcal{F}] &= \underbrace{\psi_3 - \psi_2 - (\psi_4 - \psi_2)}_{\text{Contemporanous diff}} + \underbrace{\lambda_2 - \lambda_1 - (\lambda_2 - \lambda_1)}_{\text{Lagged diff}} \\ &= \psi_3 - \psi_4\end{aligned}$$

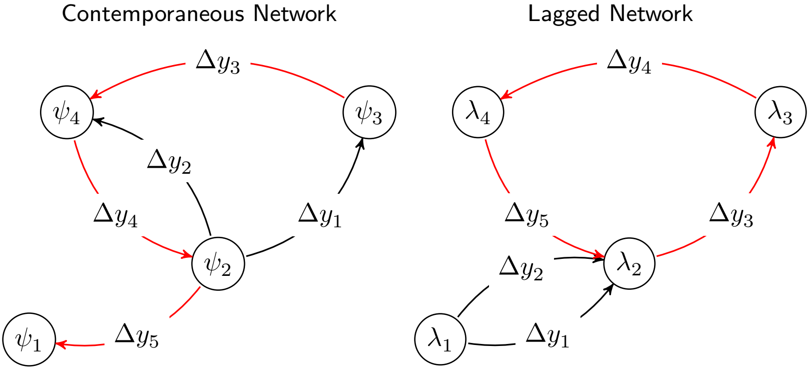
Identification of DAKM



Similarly, $\Delta y_3 + \Delta y_4 + \Delta y_5$ forms a path on the contemporaneous network and a cycle on the lagged one:

$$\mathbb{E}[\Delta y_3 + \Delta y_4 + \Delta y_5 \mid \mathcal{F}] = \psi_1 - \psi_3$$

Identification of DAKM

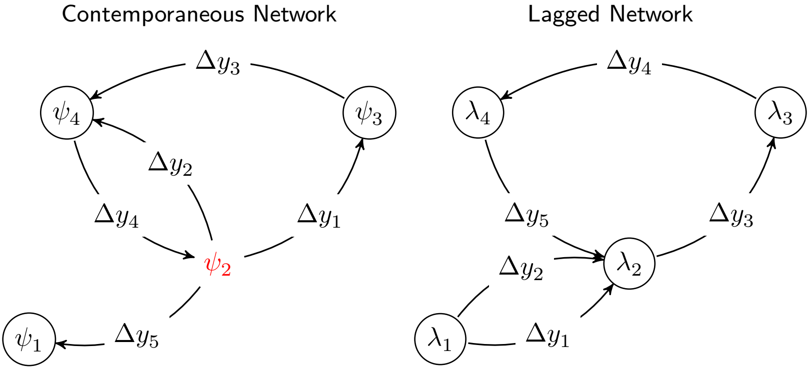


Similarly, $\Delta y_3 + \Delta y_4 + \Delta y_5$ forms a path on the contemporaneous network and a cycle on the lagged one:

$$\mathbb{E}[\Delta y_3 + \Delta y_4 + \Delta y_5 \mid \mathcal{F}] = \psi_1 - \psi_3$$

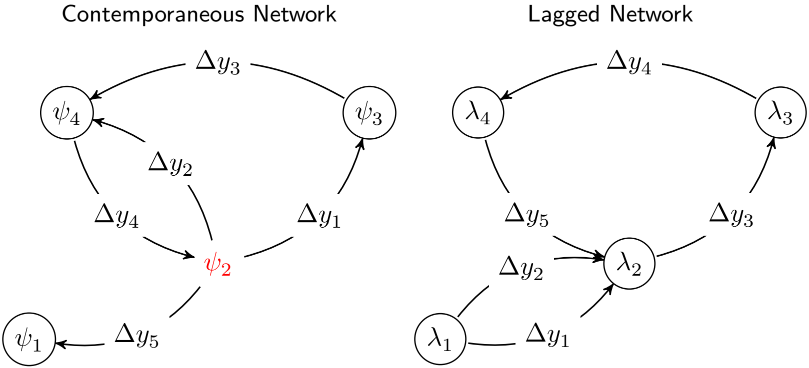
Thus the differences $\{\psi_4 - \psi_1,\ \psi_4 - \psi_3,\ \psi_3 - \psi_1\}$ are identified

Identification of DAKM



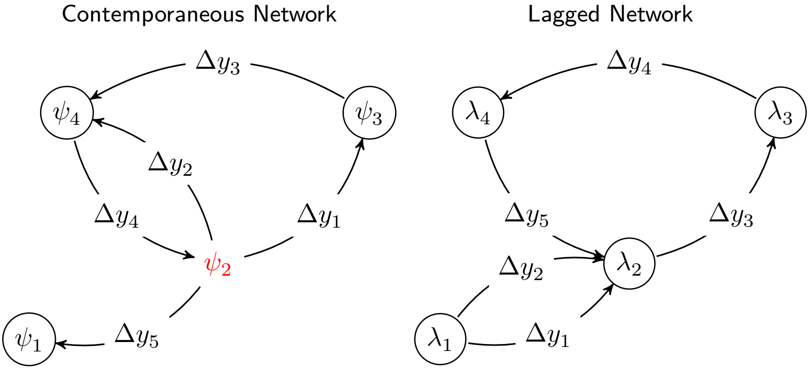
- $\psi_2 - \psi_j$ is not identified for $j \in \{1, 3, 4\}$

Identification of DAKM



- $\psi_2 - \psi_j$ is not identified for $j \in \{1, 3, 4\}$
- We refer to $\{\psi_1, \psi_3, \psi_4\}$ as the *largest identified set*
- Same argument for lagged effects

Identification of DAKM



- $\psi_2 - \psi_j$ is not identified for $j \in \{1, 3, 4\}$
- We refer to $\{\psi_1, \psi_3, \psi_4\}$ as the *largest identified set*
- Same argument for lagged effects
- Observations with $\Delta\psi_i$ and $\Delta\lambda_i$ in largest identified sets are referred to as the *largest identified sample*

Leave-one-out identification for variance components

KSS (2018) established necessity of *leave-i-out* identification for unbiased variance decomposition under heteroscedasticity:

- Decomposing Δy_i requires identification of unexplained variation:

$$\begin{aligned}\mathbb{E}\left[\mathbb{V}_n\left[\Delta y_i\right] \mid \mathcal{F}\right] = & \mathbb{V}_n[\Delta\psi_i] + \mathbb{V}_n[\Delta\lambda_i] + 2\mathbb{C}_n[\Delta\psi_i, \Delta\lambda_i] \\ & + \underbrace{\mathbb{E}\left[\mathbb{V}_n\left[\Delta\varepsilon_i\right] \mid \mathcal{F}\right]}_{=\frac{n-1}{n^2}\sum_{i=1}^n \sigma_i^2}\end{aligned}$$

- σ_i^2 is identified iff $\Delta\psi_i + \Delta\lambda_i$ is leave-i-out identified

Observations where σ_i^2 is identified are referred to as the *leave-one-out sample*. We use this sample for estimation.

Estimator

$$\Delta y_i = \Delta\psi_i + \Delta\lambda_i + \underbrace{\Delta X_i'\beta}_{\text{age + yr effects}} + \Delta\varepsilon_i$$
$$= Z_i'\gamma + \Delta\varepsilon_i$$

where $Z_i = (\Delta F_i', \Delta L_i', \Delta X_i')'$ and $\gamma = (\psi', \lambda', \beta')$

- Write any variance component θ of interest as quadratic form:

$$\theta = \gamma' A \gamma$$

- When decomposing $\mathbb{E}\left[\mathbb{V}_n\left[\Delta y_i\right] \mid \mathcal{F}\right]$, A considers individuals where both $\Delta F_i'\psi$ and $\Delta L_i'\lambda$ are in largest identified set
- For employer-centric variance components, A considers firms where both ψ_j and λ_j are in largest identified set

Plug-in Estimator (aka OLS)

Letting $S_{zz} = \sum_{i=1}^{n} Z_i Z_i'$, the OLS estimator of γ is:

$$\hat{\gamma} = S_{zz}^{\dagger} \sum_{i=1}^{n} Z_i' \Delta y_i$$

where $\dagger$ is MP-inverse (spit out by CG-routine when singular)

- "Plug-in" estimator of variance component is:

$$\hat{\theta}_{\mathsf{PI}} = \hat{\gamma}' A \hat{\gamma}.$$

- Easy to show that $\hat{\theta}_{\mathsf{PI}}$ is biased:

$$\mathbb{E}[\hat{\theta}_{\mathsf{PI}} \mid \mathcal{F}] = \theta + \operatorname{trace}\left(A \mathbb{V}[\hat{\gamma} \mid \mathcal{F}]\right) = \theta + \sum_{i=1}^{n} B_{ii} \sigma_i^2$$

  where $B_{ii} = Z_i' S_{zz}^{\dagger} A S_{zz}^{\dagger} Z_i$

KSS (2018) Estimator

$$\mathbb{E}[\hat{\theta}_{\mathsf{PI}} \mid \mathcal{F}] = \theta + \sum_{i=1}^{n} B_{ii}\sigma_i^2$$

- Let leave-i-out estimator of γ be $\hat{\gamma}_{-i} = \left(S_{zz} - Z_i Z_i'\right)^{\dagger} \sum_{l \neq i} Z_l \Delta y_l$
- Key idea in KSS: unbiased estimator of σ_i^2 is

$$\hat{\sigma}_i^2 = \Delta y_i \left(\Delta y_i - Z_i' \hat{\gamma}_{-i}\right)$$

- KSS estimator of variance component is:

$$\hat{\theta}_{\mathsf{KSS}} = \hat{\gamma}' A \hat{\gamma} - \sum_{i=1}^{n} B_{ii}\hat{\sigma}_i^2$$

Data: Veneto Work History File

Administrative social security records for Italian region of Veneto years 1984–2001

- Panel of annual earnings, weeks worked, employer ids in that year
- Extract individuals w/ career paths: NJJJ, NJUJJ, or NJJ-
- Non-employment "gap" (U) when worked < 37 weeks in year
- NILF (N) before first observed job
- "Poaching wage" is average daily earnings in first year w/ ≥ 37 weeks that employer earnings record is dominant
- For each worker, extract a single *change* in poaching wages between last two jobs

Restricting the Sample

Iterative algorithm restricts to workers in leave-one-out sample and characterizes largest identified set for contemporaneous and lagged firm effects

- Drop observations where σ_i^2 is not identified: Drop observations with statistical leverage $P_{ii} = Z_i' S_{zz}^{\dagger} Z_i$ equal to 1
- Characterize largest identified set for ψ: For $\tilde{\gamma} = S_{zz}^{\dagger} S_{zz} \gamma$, $\mathcal{J}_\psi$ is the largest set of firms such that $j \in \mathcal{J}_\psi$ if and only if

  $$\tilde{\psi}_j - \tilde{\psi}_{j'} = \psi_j - \psi_{j'}$$

  for all $j' \in \mathcal{J}_\psi$ and any γ. Same for λ yields $\mathcal{J}_\lambda$.
- Largest identified sample are individuals whose employment history is contained in $\mathcal{J}_\psi$ and $\mathcal{J}_\lambda$.

Table 1: Summary Statistics

	All Workers	Male Workers	Female Workers
Starting Sample			
Number of individuals	$572,421$	$366,810$	$205,611$
Number of firms	$167,453$	$117,632$	$104,044$
Average wage growth	0.1031	0.1177	0.0771
Variance of wage growth	0.0670	0.0611	0.0765
Leave-one-out Sample			
Number of individuals	$465,336$	$288,677$	$130,970$
Number of firms	$86,104$	$57,093$	$35,623$
Average wage growth	0.1088	0.1263	0.0766
Variance of wage growth	0.0636	0.0594	0.0662
Largest Identified Sample			
Number of individuals	$392,731$	$241,501$	$89,841$
Number of firms	$58,527$	$38,945$	$18,515$
Average wage growth	0.1091	0.1280	0.0676
Variance of wage growth	0.0640	0.0605	0.0664

DAKM explains 2% more wage growth variance than AKM

(But coefficients vary substantially by gender..)

Table 2: Explained Variance of Wage Growth

	All Workers	Male Workers	Female Workers
Variance of wage growth	0.0636	0.0594	0.0662
Explained variance, R^2 (Plug-in)			
Lagged model, $\Delta\lambda_i$	0.3894	0.4547	0.3781
AKM model, $\Delta\psi_i$	0.5190	0.6029	0.4860
DAKM, $\Delta\psi_i + \Delta\lambda_i$	0.5926	0.6692	0.5642
DAKM by gender, $\Delta\psi_i^g + \Delta\lambda_i^g$	0.6741		
Explained variance, R^2 (KSS)			
Lagged model, $\Delta\lambda_i$	0.2908	0.3620	0.2353
AKM model, $\Delta\psi_i$	0.4215	0.5197	0.3303
DAKM, $\Delta\psi_i + \Delta\lambda_i$	0.4460	0.5417	0.3520
DAKM by gender, $\Delta\psi_i^g + \Delta\lambda_i^g$	0.5275		

"Where you're at" explains 4–$5\times$ "where you're from"

Table 3: Variance Decomposition of Wage Growth

	All Workers	Male Workers	Female Workers
Variance of wage growth	0.0640	0.0605	0.0664
Variance decomposition (Plug-in)			
$\mathbb{V}_n[\Delta\lambda_i]$	0.0176	0.0160	0.0270
$\mathbb{V}_n[\Delta\psi_i]$	0.0317	0.0310	0.0413
$2\mathbb{C}_n[\Delta\psi_i, \Delta\lambda_i]$	-0.0209	-0.0189	-0.0366

$$\mathbb{E}\left[\mathbb{V}_n\left[\Delta y_i\right] \mid \mathcal{F}\right] = \mathbb{V}_n[\Delta\psi_i] + \mathbb{V}_n[\Delta\lambda_i] + 2\mathbb{C}_n[\Delta\psi_i, \Delta\lambda_i] + \mathbb{E}\left[\mathbb{V}_n\left[\Delta\varepsilon_i\right] \mid \mathcal{F}\right]$$

“Where you’re at” explains 4–$5\times$ “where you’re from”

Table 3: Variance Decomposition of Wage Growth

	All Workers	Male Workers	Female Workers
Variance of wage growth	0.0640	0.0605	0.0664
Variance decomposition (Plug-in)			
$\mathbb{V}_n[\Delta\lambda_i]$	0.0176	0.0160	0.0270
$\mathbb{V}_n[\Delta\psi_i]$	0.0317	0.0310	0.0413
$2\mathbb{C}_n[\Delta\psi_i, \Delta\lambda_i]$	-0.0209	-0.0189	-0.0366
Variance decomposition (KSS)			
$\mathbb{V}_n[\Delta\lambda_i]$	0.0039	0.0039	0.0046
$\mathbb{V}_n[\Delta\psi_i]$	0.0165	0.0177	0.0165
$2\mathbb{C}_n[\Delta\psi_i, \Delta\lambda_i]$	-0.0019	-0.0022	-0.0028

$$\mathbb{E}\left[\mathbb{V}_n\left[\Delta y_i\right] \mid \mathcal{F}\right] = \mathbb{V}_n[\Delta\psi_i] + \mathbb{V}_n[\Delta\lambda_i] + 2\mathbb{C}_n[\Delta\psi_i, \Delta\lambda_i] + \mathbb{E}\left[\mathbb{V}_n\left[\Delta\varepsilon_i\right] \mid \mathcal{F}\right]$$

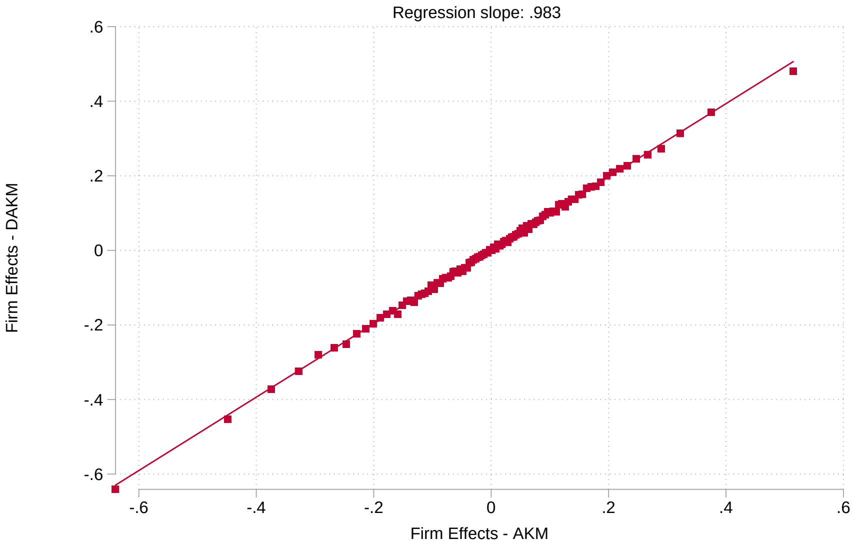
Lagged and contemporaneous effects positively correlated!

Table 4: Moments of Firm Effects (Person Weighted)

	All Workers	Male Workers	Female Workers
Firms in leave-one-out sample	$86,104$	$57,093$	$35,623$
Firms with identified ψ_j and λ_j	$40,946$	$26,477$	$11,777$
Covariances (KSS)			
$\mathbb{V}_n[\lambda_j]$	0.0039	0.0037	0.0053
$\mathbb{V}_n[\psi_j]$	0.0172	0.0172	0.0212
$\mathbb{C}_n[\psi_j, \lambda_j]$	0.0036	0.0027	0.0045
Correlation	0.4433	0.3430	0.4249

Ignoring lags has no effect on estimates of ψ

Figure 1: Estimated Contemporaneous Effects in DAKM and AKM



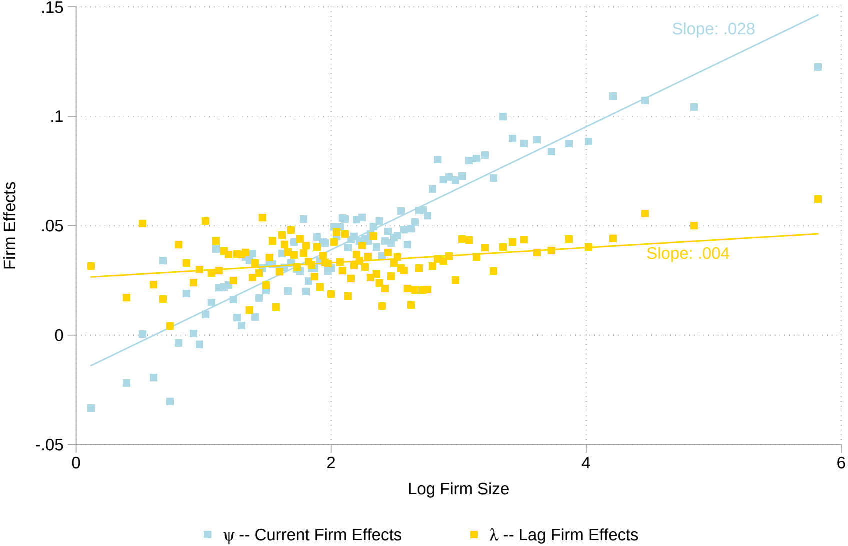
Some Normalizations

Useful to impose some normalizations on (ψ_j, λ_ℓ) to study differences in mean firm effects across groups:

1. Impose $\lambda_N = 0$
   - Effect of lagged state measured relative to N(Y)ILF
2. Set $\mathbb{E}_n[\psi_j|\text{firm size in bottom vingtile}] = 0$
   - Firm size measured as average number of workers for whom firm is dominant employer across years the firm is alive
   - For gender-specific estimates impose normalization among firms in each gender's leave-out-sample ala Card, Cardoso, and Kline (2016)

Lags more weakly correlated with firm size

Figure 2: Contemporaneous and lagged firm effects by size



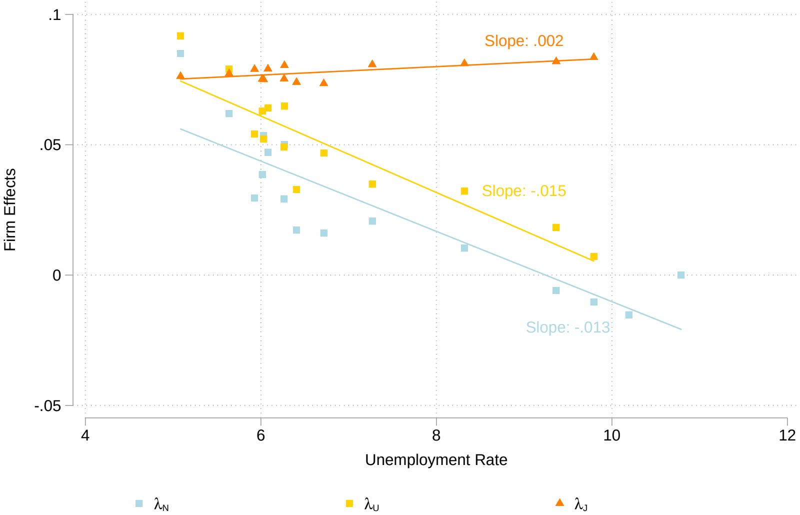
Time varying state-dependence

Figure 3: Urate, λ_U, and λ_N by year

Note: This model allows (λ_U, λ_N) to vary by calendar year. λ_N is normalized to 0 in 1984.

Better to enter LF or be hired from U when urate is low

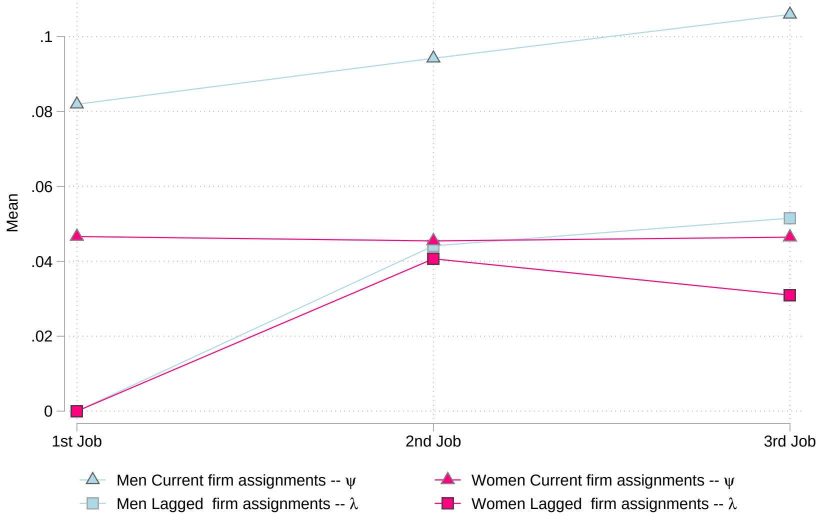
Figure 4: Employment weighted mean of ψ and (λ_N, λ_U) by Urate



Note: This model allows (λ_U, λ_N) to vary by calendar year. λ_N is normalized to 0 in 1984.

Men start out ahead, and climb, while women fall behind..

Figure 5: Mean of firm effects and lagged states by job # (NJJJ sample)



Note: ψ_j normalized to have mean zero among firms in bottom firm size vingtile. λ_j normalized relative to λ_N.

Gender Wage Decomposition (Pooled Model)

Hiring wage equation at job $t \in \{1, 2, 3\}$:

$$y_{it} - X_{it}'\beta = \alpha_i + F_{it}'\psi + L_{it}'\lambda + \varepsilon_{it}$$

Mean wages by gender $G_i \in \{m, f\}$ at job j:

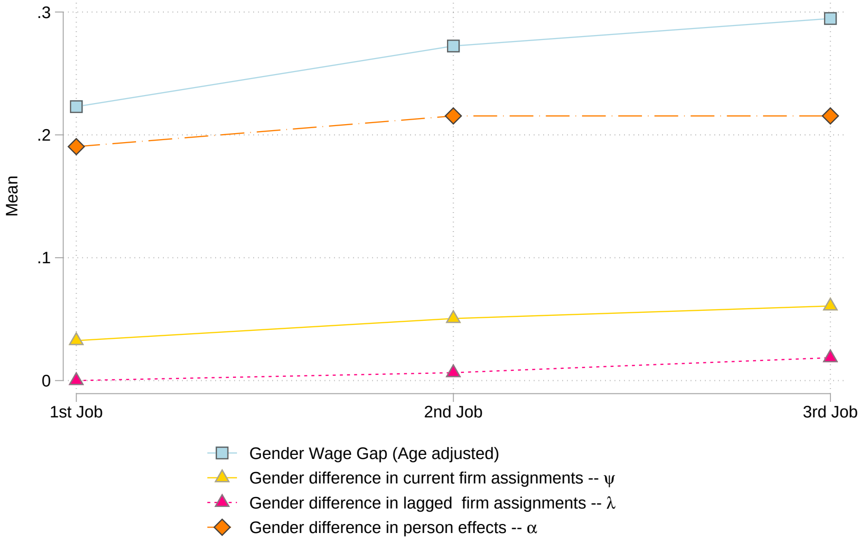
$$\mathbb{E}_n\left[y_{it} - X_{it}'\beta \mid G_i = g\right] = \bar{\alpha}_g + \bar{F}_{gt}'\psi + \bar{L}_{gt}'\lambda$$

Hence, we have the decomposition

$$\begin{aligned}
&\mathbb{E}_n\left[y_{it} - X_{it}'\beta \mid G_i = m\right] - \mathbb{E}_n\left[y_{it} - X_{it}'\beta \mid G_i = f\right] \\
&= \bar{\alpha}_m - \bar{\alpha}_f + \underbrace{\left(\bar{F}_{mt} - \bar{F}_{ft}\right)'\psi}_{\text{Move women to same firms as men}} + \underbrace{\left(\bar{L}_{mt} - \bar{L}_{ft}\right)'\lambda}_{\text{Move women to same lagged states as men}}
\end{aligned}$$

“Where you’re from” negligible for gender gap

Figure 6: Mean gender gap and firm effect contribution by job #
(all groups pooled)



Oaxaca Decomposition

Hiring wage equation at job $t \in \{1, 2, 3\}$:

$$y_{it} - X_{it}'\beta = \alpha_i + F_{it}'\psi^{G_i} + L_{it}'\lambda^{G_i} + \varepsilon_{it}$$

Mean wages by gender at job j:

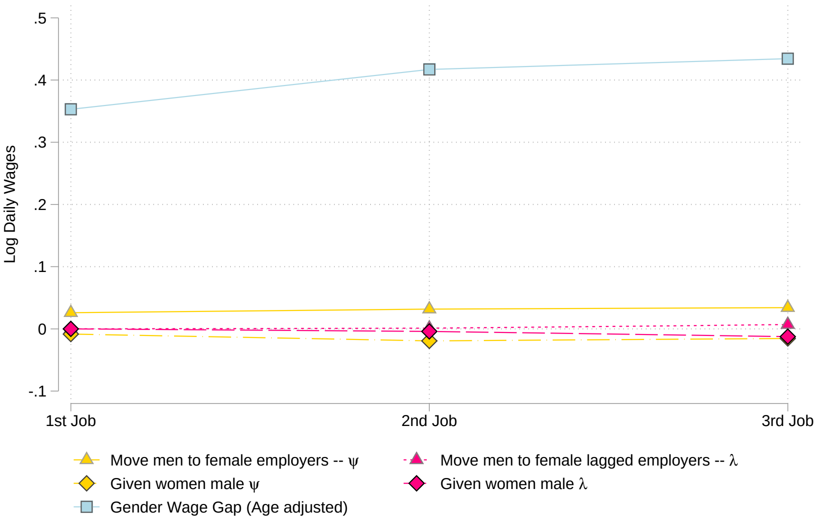
$$\mathbb{E}_n \left[y_{it} - X_{it}'\beta \mid G_i = g \right] = \bar{\alpha}_g + \bar{F}_{gt}'\psi^g + \bar{L}_{gt}'\lambda^g$$

Hence, we have the decomposition:

$$\begin{aligned}
&\mathbb{E}_n \left[y_{it} - X_{it}'\beta \mid G_i = m \right] - \mathbb{E}_n \left[y_{it} - X_{it}'\beta \mid G_i = f \right] \\
&\quad = \bar{\alpha}_m - \bar{\alpha}_f - \underbrace{\left(\bar{F}_{ft} - \bar{F}_{mt} \right)' \psi^m}_{\textbf{Move \textit{men} to same firms}} - \underbrace{\left(\bar{L}_{ft} - \bar{L}_{mt} \right)' \lambda^m}_{\textbf{Move \textit{men} to same states}} \\
&\qquad + \underbrace{\bar{F}_{ft}' \left(\psi^m - \psi^f \right)}_{\textbf{Give \textit{women} male } \psi\textbf{'s}} + \underbrace{\bar{L}_{ft}' \left(\lambda^m - \lambda^f \right)}_{\textbf{Give \textit{women} male } \lambda\textbf{'s}} \qquad \text{(normalization dependent)}
\end{aligned}$$

Oaxaca: it's still where you're at

Figure 7: Mean gender gap and firm effect contribution by job #
(all groups pooled)



Conclusions

Modest state dependence in job ladder (consistent w/ BLM)

- Roughly 4% premium on average for poached vs. N(Y)ILF
- Negligible bias in ψ estimates from omitting lagged states
- Positive correlation between ψ_j and λ_j inconsistent w/ PVR
- Firm size more strongly correlated with ψ_j than λ_j
  - ψ a better proxy of rents?
- Strong correlation of (λ_U, λ_N) with Urate
  - Consistent w/ many models: sequential auctions / Nash Bargaining / implicit contracts.
  - But potentially at odds with reduced form findings of Jäger et al (2019) that hiring wages insensitive to UI benefits

Conclusions

Stark gender differences in sorting

- Men start out at higher paying firms
- Male advantage intensifies by (slowly) climbing the ladder, while women fail to climb when moving
- NJJ- women initially move to firms that are worse to be from!

Conclusions

Stark gender differences in sorting

- Men start out at higher paying firms
- Male advantage intensifies by (slowly) climbing the ladder, while women fail to climb when moving
- NJJ- women initially move to firms that are worse to be from!

But $\lambda_{\ell(i,t)}$ quantitatively unimportant for gender gap in hiring wages..

- 13–20% of gender gap attributable to where workers "are at"
- $\leq 1\%$ due to differences in where "they're from"
- Prediction: salary non-disclosure laws unlikely to substantially alter gender gap (consistent w/ Agan et al, 2019)