Decomposing Changes in the Gender Wage Gap over Worker Careers

Keyon Vafa
Harvard University
(incoming postdoctoral fellow)

Susan Athey
Stanford University
Graduate School of Business

David Blei
Columbia University
Dept. of Computer Science

Gender wage gaps

In the United States, women earn ~80% the male hourly wage.

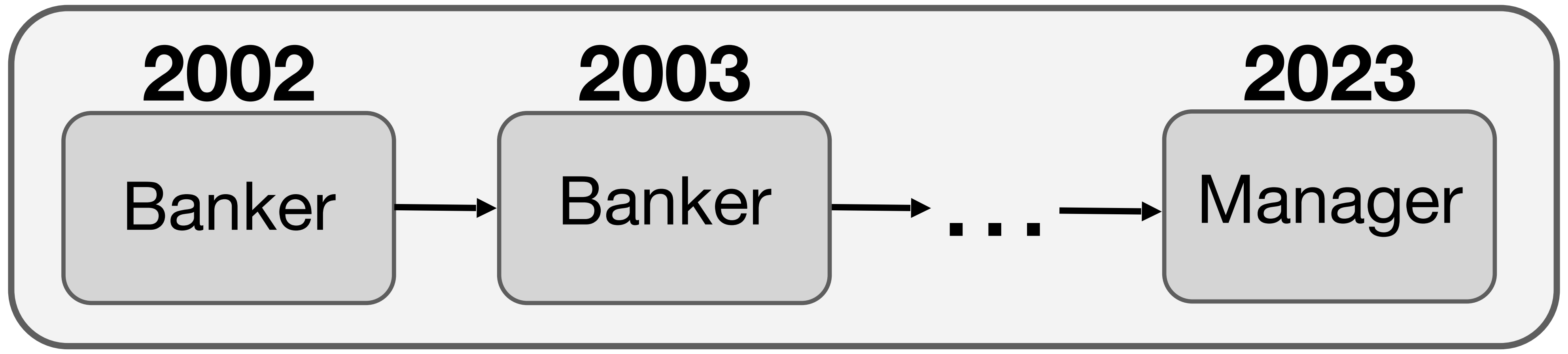
A large body of work studies the explained wage gap (EWG), the portion of the gender wage gap (GWG) that can be accounted for by differences in male and female characteristics.

Blinder (1973); Oaxaca (1973); Blau and Kahn (2017)

Note: Decompositions are descriptive, not causal; we are not attempting to explain why characteristics are different or how they relate to unobservables, or to measure discrimination.

This talk

History captures information about human capital that is missing from simpler variables that is typically included in wage gap decompositions:

$H =$ 

Our goal is explain wage gap with full worker histories:

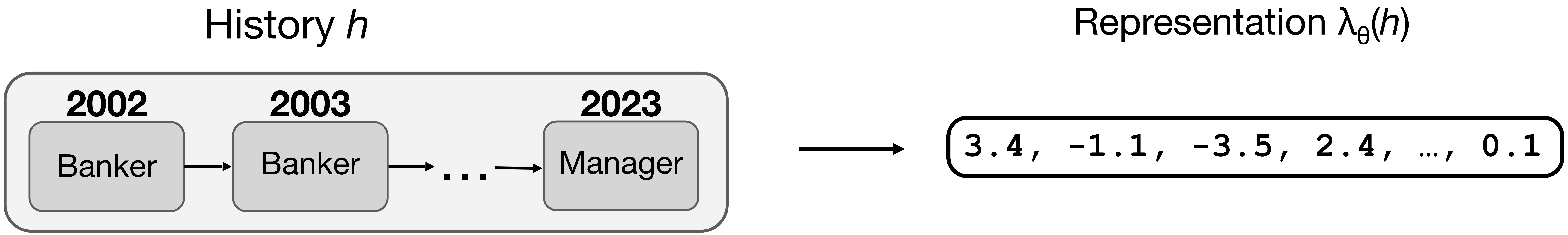
$$\text{EWG} = \mathbb{E}_{p(h|{\color{blue}m})}\left[\mu_m(H)\right] - \mathbb{E}_{p(h|{\color{red}f})}\left[\mu_m(H)\right]$$

where μ_m is the (male) expected wage.

We find that history explains more of the GWG and helps us understand the role of transitions, i.e., how workers move from job to job.

Modeling histories with machine learning

We develop machine learning methods to include occupational histories in GWG decompositions by learning **low-dimensional *representations*** of history



We analyze a dataset of 24M resumes to help learn these representations.

On PSID, our out-of-sample wage R^2 improves from 0.46 to 0.53, when additionally conditioning on representations of history.

Representations of history explain ~25% of GWG that is unexplained with standard covariates.

Understanding impacts of transitions

Males and females have different early-career characteristics, evolve differently.

Loprest (1992); Manning and Swaffield (2008); Bertrand et al. (2010), Goldin et al. (2017)

Which difference is driving EWG to **change between two periods of career**?

- Gender differences in **initial histories** (at period 0):

$$p(h^0 \mid m) \text{ vs. } p(h^0 \mid f)$$

- Gender differences in **how histories transition** between periods 0 and 1:

$$p(h^1 \mid h^0, m) \text{ vs. } p(h^1 \mid h^0, f)$$

We propose a **decomposition** of wage gap change into these two sources.

Using representations of history, we find that gender differences in **transitions** exacerbate GWGs while gender differences in **initial histories** close them.

Transition effect dominates early career; **initial histories** effect dominates later.

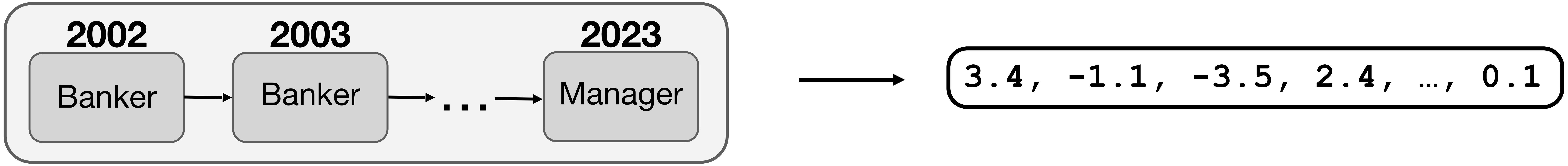
Methodology

Representations of history

The first step is to learn a mapping from full histories to **representations**:

$$\lambda_\theta : \mathcal{H} \rightarrow \mathbb{R}^D$$

History h Representation $\lambda_\theta(h)$



Our goal is to find a representation that is **predictive of expected wage**.

We use CAREER, a *transformer neural network* that involves the same idea behind large language models (LLMs).

Vaswani et al. (2017), Brown et al. (2020), Vafa et al. (2022)

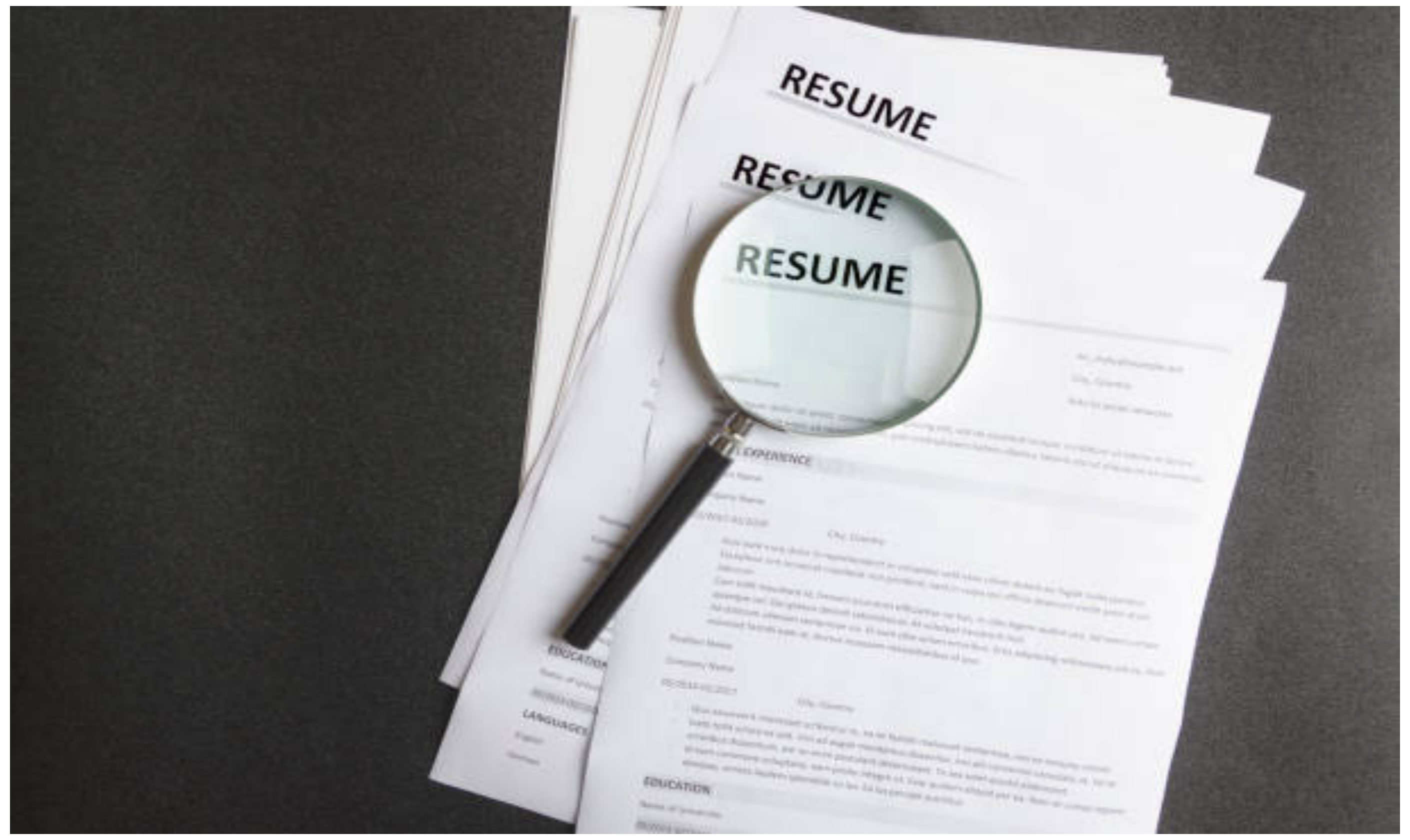
Representing histories with transformers

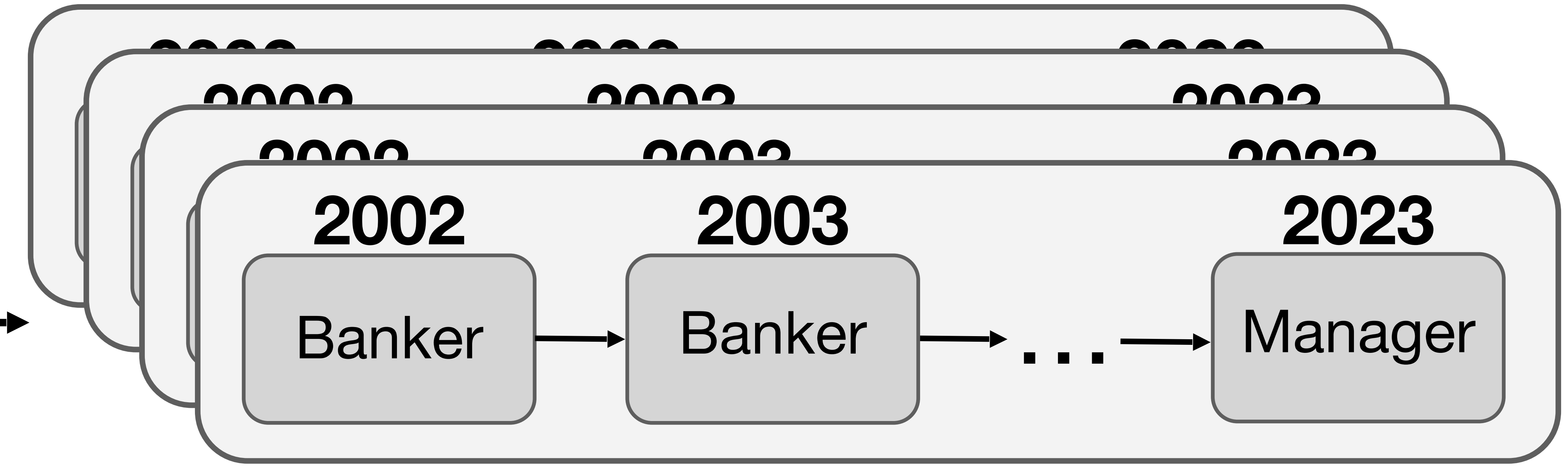
Transformers require large amounts of data to learn high-quality representations.

Kaplan et al. (2020)

But longitudinal surveys for wage gap estimation are small.

CAREER is first trained to a dataset of 23.7M resumes to learn representations:





These representations are then adjusted to predict wage.

Fitting CAREER's representation

	Input	**Goal**
CAREER pretraining on large-scale resumes:	Passively-collected resumes **2002** Banker → **2003** Banker → … → **2022** Manager	Representation that predicts next job Manager
CAREER fine-tuning on longitudinal surveys:	Survey data **2002** Tutor → **2003** Graduate student → … → **2023** Teacher + covariates and gender	Predict wage $24/hour

Pretraining: Representations for next-job

Resumes do not contain wages, but they contain many career trajectories.

Sequence from resume

$H =$ 2002 Banker → 2003 Analyst → … → 2023 Manager

Modeling objectives

$p(H_1 = \text{Banker})$

$p(H_2 = \text{Analyst} \mid H_1 = \text{Banker})$

$\vdots$

$p(H_{21} = \text{Manager} \mid h_1, \ldots, h_{20})$

Model uses representations ($\alpha_j \in \mathbb{R}^D$)

$$p(H_{t+1} = j | h_1, \ldots, h_t) = \frac{\exp\left\{\alpha_j \cdot \lambda_\theta(h_1, \ldots, h_t)\right\}}{\sum_{j'} \exp\left\{\alpha_{j'} \cdot \lambda_\theta(h_1, \ldots, h_t)\right\}}$$

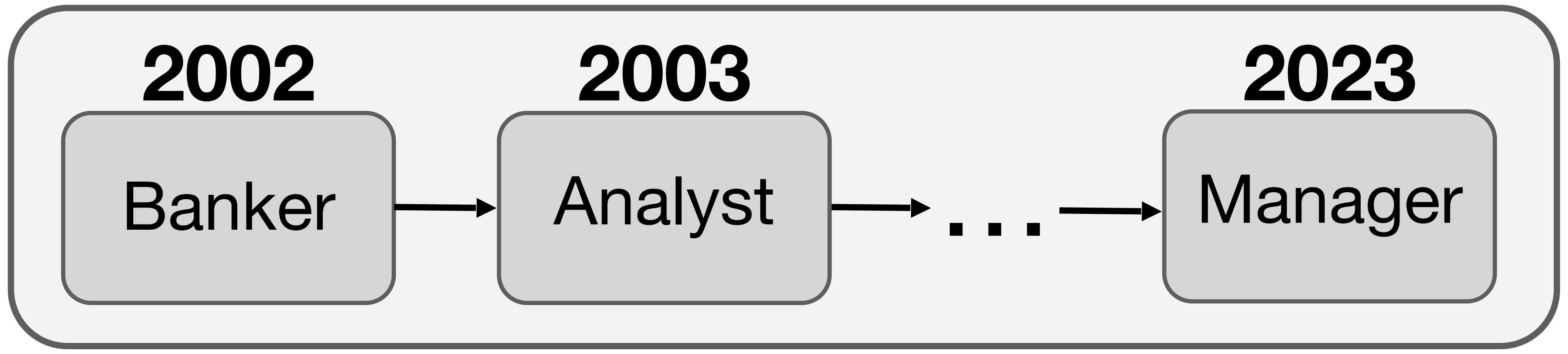
Representation is fit using resume sequences, where it acquires features that are also predictive of wage.

Fine-tuning: Representations for wage

On survey data, we adjust the representation to predict wages:

Survey observables

$H =$



Modeling objective

$$\mathbb{E}[Y|G = g, X, H]$$

Log-wage $y \in \mathbb{R}$, covariates $X \in \mathbb{R}^p$, gender $G \in \{m, f\}$

Approximate conditional expected wage function with

$$\mu_g(x, \lambda_\theta(h)) = \beta_g \cdot x + \rho_g(\lambda_\theta(h))$$

$\rho_g : \mathbb{R}^D \to \mathbb{R}$ is small neural network, $\beta_g \in \mathbb{R}^P$ is regression coefficients.

Optimize parameters to predict wage, initializing with λ_θ fit to resumes.

Minimizing Omitted Variable Bias

We've described a method to learn representations that are predictive of wage.

But representations discard information. What if representations discard important aspect of history for explaining wage gap? (OVB)

We provide a condition under which representations do not omit important variables, which is satisfied when representations are **sufficient**:

$$\underbrace{G}_{\text{gender}} \perp\!\!\!\perp \underbrace{H}_{\text{history}} | \underbrace{\lambda_\theta(H)}_{\text{representation}}, \underbrace{X}_{\text{standard covariates}}$$

We propose an inference algorithm to encourage representations that are both sufficient and predictive of wage.

Data

Resumes: 23.7 million resumes collected by Zippia.

PSID Sample: Full-time, non-farm and non-military wage and salary workers aged 25-64 between 1989-2018 (91,391 total observations).

Construct worker histories:

- Use **occ1990dd** codes for occupations (330 total).
- Use special occupations when not working (e.g. "unemployed", "student").

Trimming:

- Trim data to ensure overlap: $0.01 < P(G = F|X, \lambda_\theta(H)) < 0.99$

Two classes of models

1) Linear models **based on summary statistics**:

$$\mu_g(x, \lambda_\theta(h)) = \beta_g \cdot x$$

We consider:

- coarse-grained occupations (21 categories) fine-grained occupations (330 categories)
- OLS vs LASSO

Summary statistics:

- experience
- education
- race/ethnicity
- region
- union status
- industry
- occupation
- year interactions

2) Models that **include history** with CAREER:

$$\mu_g(x, \lambda_\theta(h)) = \beta_g \cdot x + \rho_g(\lambda_\theta(h))$$

Predictive performance

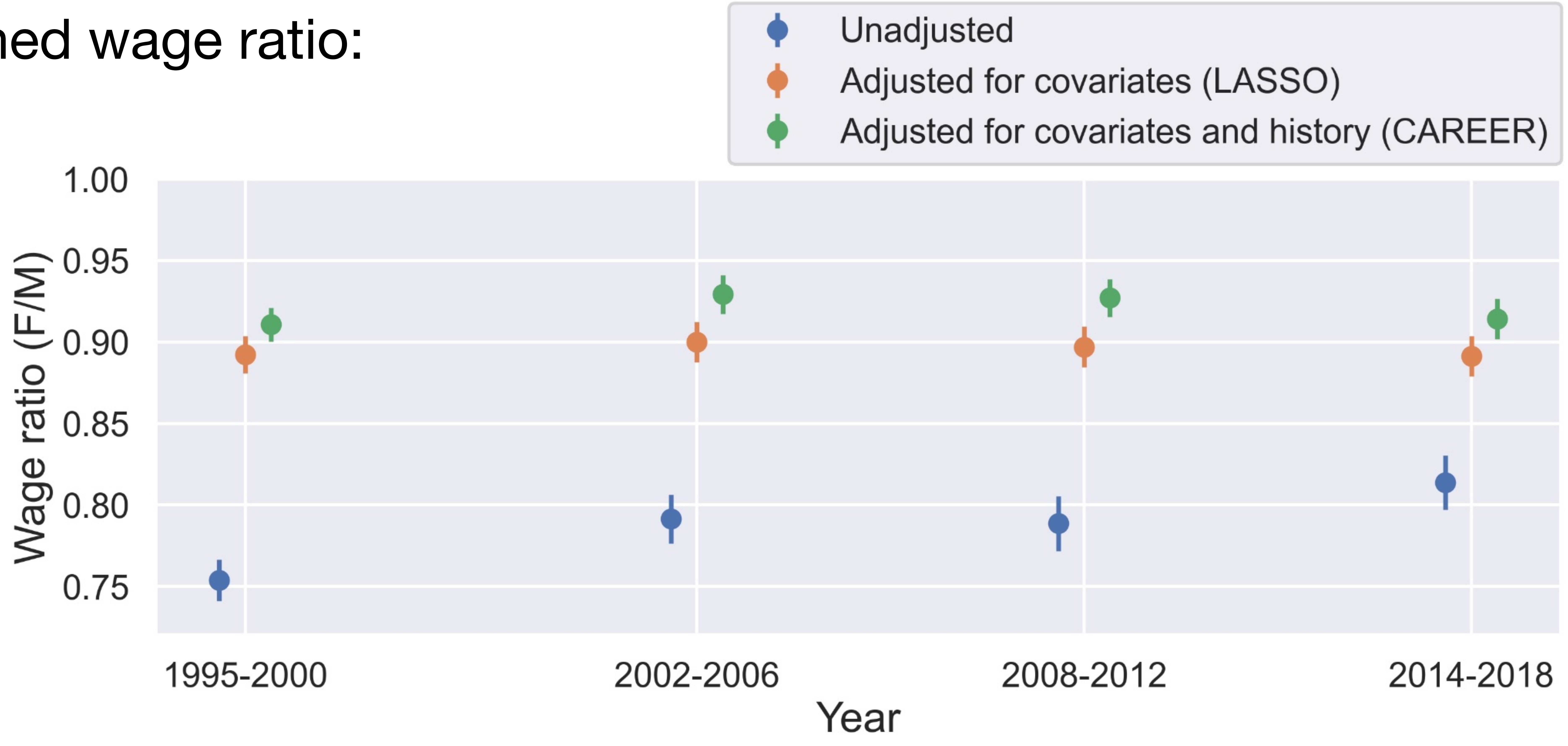
All models are fit with cross-fitting; all reported values are **out-of-sample**.

	Overall R^2	Male R^2	Female R^2
Coarse-grained regression	0.417 (0.010)	0.479 (0.007)	0.404 (0.010)
Coarse-grained LASSO	0.430 (0.010)	0.492 (0.006)	0.419 (0.009)
Fine-grained LASSO	0.456 (0.010)	0.522 (0.006)	0.454 (0.008)
CAREER (current job only)	0.458 (0.009)	0.524 (0.006)	0.456 (0.007)
CAREER (participation and current job only)	0.475 (0.009)	0.543 (0.007)	0.478 (0.007)
CAREER (no pretraining on resumes)	0.491 (0.008)	0.561 (0.006)	0.499 (0.007)
CAREER (pretraining on resumes)	0.527 (0.007)	0.591 (0.006)	0.533 (0.005)

Improvement is not only due to better functional form of current occupation or capturing workforce participation spells.

Decomposing wage gap

Unexplained wage ratio:



CAREER's representation of history explains ~25% of remaining wage gap when history is not included between 1995-2018.

Which histories are improving predictions?

Clustering representations allows us to interpret the aspects of history that improve wage predictions.

These clusters (omitted by standard models) are also predictive of gender:

Decomposing Explained Gaps over Careers

Decomposing changes in explained wage gap over careers

Consider a fixed set of workers observed at two periods of their career, e.g. at age 30 (t=0) and age 42 (t=1).

- EWG_0 is explained wage gap at initial period.
- EWG_1 is explained wage gap at end period.
- EWG_1 - EWG_0 describes how the explained wage gap changes over the workers' careers.

We propose a decomposition of EWG_1 - EWG_0.

Decomposing wage gaps over careers

We show EWG_1 - EWG_0 can be written as the sum of two terms:

1) Effect of gender differences in **transition distributions** between periods:

fixed initial history distribution

$$\underbrace{\mathbb{E}_{p(h^0|f)}\left[\mathbb{E}_{p(h^1|h^0,f)}\left[\mu_m(H^1)|H^0\right]\right]}_{\substack{\text{expected end wage for} \\ \text{female initial histories} \\ \text{and female transitions}}} - \underbrace{\mathbb{E}_{p(h^0|f)}\left[\mathbb{E}_{p(h^1|h^0,m)}\left[\mu_m(H^1)|H^0\right]\right]}_{\substack{\text{expected end wage for} \\ \text{female initial histories} \\ \text{and male transitions}}}$$

Difference in expected period-1 wage due to difference in male and female **transitions** (with same initial histories).

Decomposing wage gaps over careers

We show EWG_1 - EWG_0 can be written as the sum of two terms:

2) Effect of gender differences in **initial histories**:

$$\overbrace{\mathbb{E}_{p(h^0|f)}\left[\mathbb{E}_{p(h^1|h^0,m)}\left[\mu_m(H^1)|H^0\right]\right]}^{\text{expected end wage for female initial histories and male transitions}} - \overbrace{\mathbb{E}_{p(h^0|m)}\left[\mathbb{E}_{p(h^1|h^0,m)}\left[\mu_m(H^1)|H^0\right]\right]}^{\text{expected end wage for male initial histories and male transitions}}$$
$$- \underbrace{\left(\mathbb{E}_{p(h^0|f)}\left[\mu_m(H^0)\right] - \mathbb{E}_{p(h^0|m)}\left[\mu_m(H^0)\right]\right)}_{\text{initial period explained wage gap}}$$

Difference in expected wage growth due to difference in male and female **initial histories** (with same transitions).

Putting decomposition together

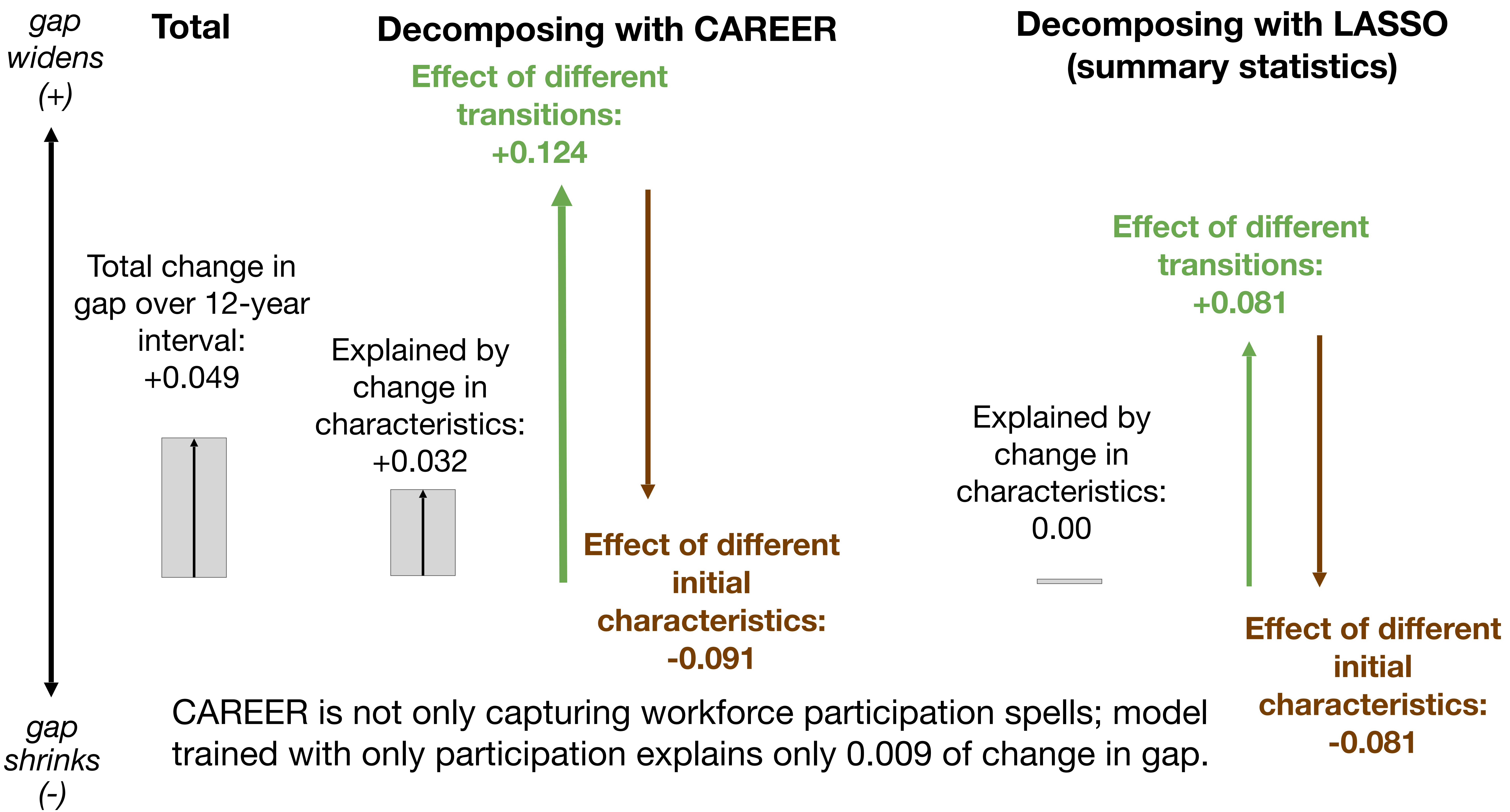
We show:

EWG_1 - EWG_0 = Effect of gender differences in **history transitions**
+ Effect of gender differences in **initial histories**.

We use this decomposition to study individuals who are full-time workers at both endpoints of 12-year interval.

Note: Decomposition includes individuals who are more committed to labor force.

Individuals aged 25-35 at beginning of 12-year interval



Findings

Age at beginning of 12-year interval:	**25-35**	**40-50**
After 12-years, GWG:	Expands: +0.049 log points	Shrinks: -0.052 log points
[1] Fixing initial characteristics, do female transitions lead to higher-value characteristics than males?	No: +0.124	No: +0.046
[2] Fixing transitions, do female initial characteristics set up females for more wage growth than males?	Yes: -0.091	Yes: -0.086
Net effect:	EWG expands |**[1]**| > |**[2]**|	EWG shrinks |**[1]**| < |**[2]**|

Conclusion

ML methods capture omitted variables relevant to wage gap decompositions.

On PSID, full worker history explains ~25% of the wage gap that is unexplained by traditional models.

We propose a decomposition for the **change** in explained gaps over careers.

Results about change in wage gaps over 12-year intervals in PSID

- Younger workers: **widening** wage gap driven by **gender differences in transitions**, keeping initial characteristics fixed
- Older workers: **shrinking** wage gap over 12-year interval is driven by **gender differences in initial wage growth potential**, keeping transitions fixed.