Essays on Firm and Labor Market

Xuanli Zhu
University of Tokyo

September 13, 2023

Overview

- 4 empirical papers on how firm and labor market interact to affect firm activities and labor market inequalities
- Ch.1: How labor market supply and frictions affects firm dynamics
- Ch.2: How worker and firm differences jointly affect wage inequality
- Ch.3: How firm affects non-wage compensation provision inequality
- Ch.4: How labor market liquidity affects firm training and worker learning
- Contribution: new data; new method; new fact; new theory

Outline

Research Questions

- How do firm/establishment dynamics (entry, exit, lifecycle growth) evolve over long time periods?
  - Important for economics growth (e.g. creative destruction) and market efficiency (e.g. resource reallocation/misallocation)
- What are the main drivers of the evolution of long-run market dynamics?
  - Various explanations in the literature: entry cost; frictions; labor supply; ...
  - Less clear on short-run vs. long-run determinants
  - Potential long-run dynamic effects and history dependency

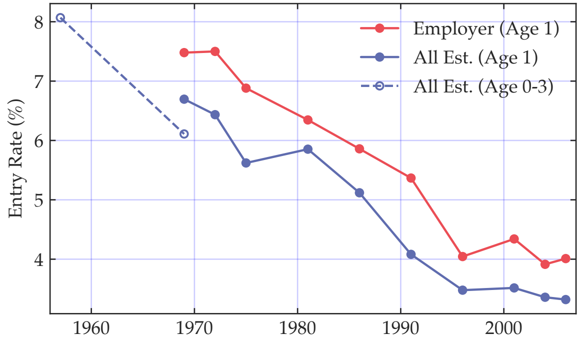
This Paper

- Use newly collected historical statistics to study the long-run establishment dynamics of post-war Japan (1950s-2000s)
  - More works since lost decades, but less known for earlier periods
  - Identify cohort-specific lifecycle growth from repeated cross-sectional data
- Calibrate a typical firm dynamics model to test various theories on the observed evolution of post-war Japan establishment dynamics
  - Test if a newly found driver—labor supply (Karahan et al., 2019; Hopenhayn et al., 2020)—or traditional drivers can explain the entry rate trends in Japan
  - Check if labor market distortions can explain changes in avg. size and lifecycle growth

Data Source and Definition

- Establishment Census in Japan:
  - All private establishments in non-agriculture sectors since 1951 to 2006, conducted every 3 or 5 years (manually collected for before 1980s)
  - Aggregate statistics on establishment number and employment in various categories
  - Age of an establishment is defined as the years passed since it operated its present business in its present physical location
- Focus on incorporated establishments (”Employers”)
  - Excluding individual proprietorship (”Nonemployers”) given its different nature
  - Consistent with the literature
- Establishments are different from firms, but not too much
  - Over 80% of the firms are single-establishment firm
  - Market dynamism more simple and natural at establishment level

Fact 1: Long-Run Decline in Entry Rate

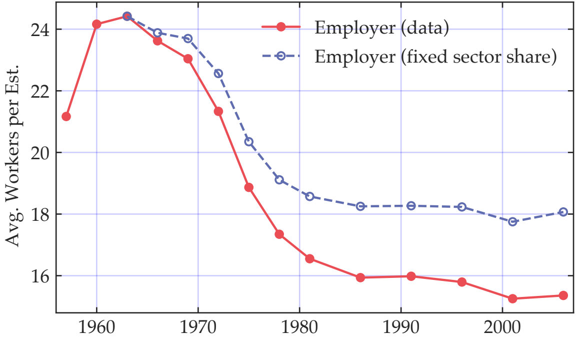


- Entry rate (= Age 1 est. mass / total mass) declined about 3.5 percent points from 1969 to 2006
- This steady decline starts since around late 1950s if not earlier!

◂ similar across industry ◂ stagnated exit rate
◂ aging est. demographics

$\rightsquigarrow$ Long-run driver?

Fact 2: Shrinking Establishment Size



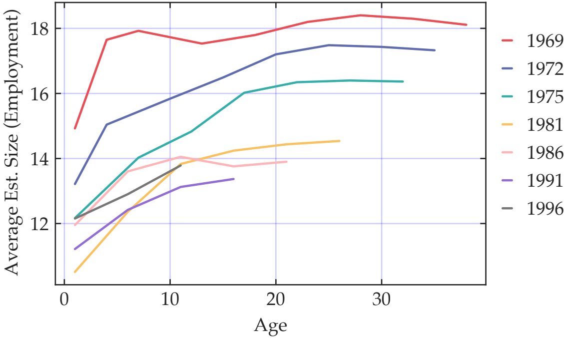
- Average establishment size declined over 30% in 1960s and 1970s
- Structural transformation from Manufacturing to Service? $\longrightarrow$ About 68% of the decline is within (2-digit) industry

◂ diverge across industry

◂ different trends for different age groups

$\rightsquigarrow$ Puzzling?

Fact 3: Birth Cohort Effect on Life Cycle Growth



(Imputed from ▸ *avg. size-age correlation)*

- Growth of a cohort mainly occurs when young, and nearly stops after around age 20 (▸ cross-country comparison)
- Parallel-like shifts in life-cycle growth across entry cohorts (cond. on surviving)
- Early-life growth is higher in early years and flattened since 1990s

◂ different across industry ◂ early cohorts also parallel

⇝ History matters

Benchmark Model

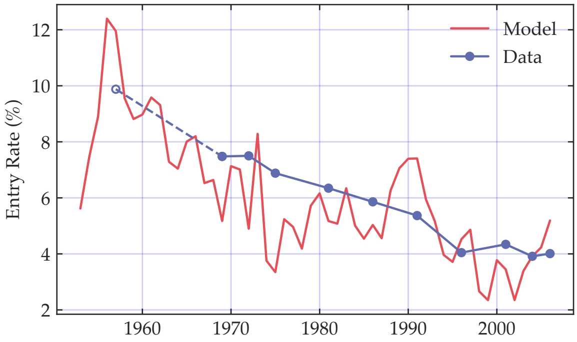
- We use the canonical Hopenhayn firm dynamics model as our benchmark
  - $\pi_t(s_t, n_t, w_t) = s_t n_t^\theta - w_t n_t - w_t c_f$
  - $V(s_t, \mathbf{w}_t) = \max_{n_t} \pi_t(s_t, n_t, w_t) + \beta \max_{X \in \{0,1\}} \{\mathbb{E} V(s_{t+1}, \mathbf{w}_{t+1} | s_t), 0\}$
  - $\bar{s}_t = \inf\{s | \mathbb{E} V(s_{t+1}, \mathbf{w}_{t+1} | s_t) \geqslant V^x\}$ (optimal exit)
  - $V^e(\mathbf{w}_t) = \int V(s, \mathbf{w}_t) \, dG_t(s) - c_e$ (free entry)
- Putting the model to Balanced Growth Path (L grow at rate η):
  - Labor market clearing: $\int \{n(s, w^*) + c_f\} \, d\tilde{\mu}^*(s) = 1$
  - Law of motion on the productivity distribution:
    $\tilde{\mu}^*(A) = \frac{1}{1+\eta} \int\int_{s' \in A, s \geq \bar{s}^*} dF(s'|s) d\tilde{\mu}^*(s) + \tilde{m}^* \int_{s' \in A} dG(s')$
  - Both total est. measure and entrant measure m grow at η
- During Transitional Path (due to η changes), aggregate states w^* and $\bar{s}^*$ keep invariant, and entrants work as a labor-absorbing wedge

Calibrating to Period Average

Moments	Data	Model	
Entry rate, %	5.76	5.62	Target
Exit rate, %	2.56	3.62	
Average establishment size	17.57	16.82	Target
Average entrant size	12.63	13.57	Target
Average life-cycle growth rate, % (conditional on survival)			
Age 1-10	21.65	21.88	Target
Age 1-20	30.17	29.72	Target
Age 1-26	31.98	32.32	
Number share by size, %			
Employment 1-9	61.64	63.86	Target
Employment 10-29	27.14	25.13	
Employment 30-99	9.03	8.76	
Employment 100+	2.16	2.25	
Number share of entrant by size, %			
Employment 1-9	67.98	67.40	Target
Employment 10-29	24.21	23.66	
Employment 30-99	6.55	7.53	
Employment 100+	1.19	1.41	

- Calibrate our benchmark model parameters to the average firm statistics over 1969-2006 and average life cycle growth over 1969-1981 model lifecycle growth
- We assign an average labor supply growth rate of 2% thus the model exit rate in BGP is higher than data
- The avg. size derived from model is a little deviated from the data, but the life cycle growth and share distribution is well matched

Declining Labor Force Can Drive Long-Run Entry Decline



- By feeding the labor force growth into our benchmark model, it generate entry rate decline in the transition path qualitatively similar to data in the long-run

- The medium/short-run fluctuations come from labor supply trends

◂ with HP filters

▸ weak feedback effects

▸ cannot explain exit rate and avg. size changes

Traditional Explanations Fail to Explain Large Entry Decline W/O Generating Inconsistent Trends

To produces 2.2 percent points entry rate decline

	Benchmark	Labor Growth	Entry Cost	Exit Value	Fixed Cost
η, %	2.00	0.00	-	-	-
c_e	76.05	-	136.05	-	-
V^x	0.00	-	-	-20.79	-
c_f	2.12	-	-	-	0.86
w^*	0.98	0.98	0.78	0.95	1.09
$\bar{x}^*$	1.32	1.32	0.82	0.82	0.82
Entry Rate, %	5.62	3.43	3.41	3.41	3.41
Exit Rate, %	3.62	3.43	1.41	1.41	1.41
Avg. Entry Size	13.57	13.57	23.49	14.68	9.46
Avg. Entry Size (after exit)	14.89	14.89	23.84	14.89	9.61
Avg. Est. Size	16.82	17.31	21.61	13.58	8.71
LifeCycle Growth Rate 10y, %	21.88	21.88	-2.51	-2.36	-2.51
LifeCycle Growth Rate 20y, %	29.72	29.72	-7.71	-7.25	-7.71

- Entry cost increase: price effect dominates and raises more entry size than avg. size
- Exit value decline: weakened selection effects lower avg. size for incumbents but has less effect on entry size
- Fixed cost decline: a combination of higher wage and weakened selection

Size-Correlated Labor Tax

Assume a labor wage tax $(1+\tau_i^w) = s^{\gamma_i}$ that depends on productivity s

	Benchmark	γ=0.04	γ=0.07	γ=0.12	γ=0.20
w^*	0.98	0.92	0.87	0.80	0.71
w min	0.98	0.83	0.72	0.58	0.43
w max	0.98	1.04	1.09	1.17	1.30
w max / w min	1.00	1.25	1.50	2.00	3.00
$\bar{x}^*$	1.32	1.26	1.20	1.12	0.99
Entry Rate, %	5.62	5.29	4.99	4.60	4.06
Exit Rate, %	3.62	3.29	2.99	2.60	2.06
Avg. Entry Size	13.57	13.59	13.60	13.63	13.67
Avg. Est. Size	16.82	16.07	15.44	14.62	13.55
LifeCycle Growth Rate 10y, %	21.88	16.77	12.53	7.11	0.47
LifeCycle Growth Rate 20y, %	29.72	22.31	16.20	8.40	-1.18

- A tax generating a 2-fold wage gap btw. the smallest and largest establishments reduces avg. est. size for about 13%
- It has cancelled effects (lower cost for entry but higher cost for growth) for entry size and thus decrease the life cycle growth.

Labor Adjustment Costs

Assume a adjustment costs $\Phi\left(n_t, n_{t-1}\right) = \tau^a \cdot \max\left\{0, n_{t-1} - n_t\right\}$

	Benchmark	Firing		Full	
	τ^a=0.00	τ^a=0.25	τ^a=0.50	τ^a=0.25	τ^a=0.50
w^*	0.98	0.95	0.93	0.92	0.88
$\bar{x}^*$ (mean)	1.32	1.29	1.26	1.26	1.20
Entry Rate, %	5.62	5.45	5.29	5.29	5.01
Exit Rate, %	3.62	3.45	3.29	3.29	3.01
Avg. Entry Size	13.67	11.56	10.70	10.64	9.71
Avg. Est. Size	16.93	16.51	16.34	16.28	15.92
LifeCycle Growth Rate 10y, %	21.85	42.10	53.53	53.74	67.03
LifeCycle Growth Rate 20y, %	29.66	50.47	61.82	62.07	75.62
Job Turnover Rate, %	0.47	0.29	0.24	0.24	0.18

- Entrants have avg. size decline because they would hire less to avoid an additional firing cost when exit
- However, incumbent avg. size does not decline much and the life cycle growth thus increase substantially

Summarizing Main Results

1. Persistent decline in market dynamism in post-war Japan
   - Can be large explained by decline in labor supply growth
   - Other traditional explainers would generate inconsistent changes in est. avg. sizes
2. Establishment size decline and Lifecycle growth downward shifting in 1960s & 1970s
   - Labor market distortions such as size-correlated labor tax and labor adjustment costs fail to generate such declines
   - Alternative mechanisms such as initial investment channel might be required

Outline

Motivation

- What's the determinants of wage dispersion in the labor market?
  $\rightarrow$ Worker heterogeneity + Firm heterogeneity + W-F sorting + ...

- Major econometric problem: unobserved worker/firm characteristics
  $\rightarrow$ common approach: TWFE + linked EE panel data (AKM1999)

- Results from the literature:
  1. 50+% worker effect $\rightarrow$ unobserved skill & task variations
  2. 5-15% firm effect $\rightarrow$ variations in firm wage premiums
  3. 5-15% sorting $\rightarrow$ important to correct for limited mobility bias

$\rightarrow$ *Q1: Only available for a limited set of developed countries. Other countries? Alternative ways?*

$\rightarrow$ *Q2: Do we fully understand any of these components? Deep drivers? Heterogeneity?*

This Paper - New Method

- A new way to study wage determination taking advantage of
    1. Online job vacancy/ads data
    2. Machine learning algorithms

- Key idea: worker $\sim$ job

  As firms document all the job characteristics to attract their ideal candidates, and post wage based on their valuation ▸ vacancy sample

  Implicit presumptions: directed search & perfect matching

- Advantage:
    1. Vacancy data is more accessible & up-to-date
       $\rightarrow$ EE data is not always available, e.g. China
    2. Not only alternative but also ideal environment for studying firm effect & sorting
       $\rightarrow$ Pre-bargaining; Pre-mismatch
    3. Estimation is more flexible & parsimonious
       $\rightarrow$ No restriction on connected set or exogenous mobility, less limited mobility bias
    4. Open the black box of worker effect in a data-driven way
       $\rightarrow$ See what are the important skills/tasks contributing to wage differential & sorting

What Exactly We Do

0. Use 4m vacancy data from a Chinese job board (2013-2020) with full job description texts & posted wages
1. ML part: Use basic supervised & unsupervised ML methods to explore the high-dimensional job-text data and to generate proxy variables for various skills&tasks
   1.1 Feature Selection
   1.2 Feature Clustering
   1.3 Dimensional Reduction

   two methods (w/ & w/o human knowledge)

   (Why basic? Interpretation + Performance)
2. Econometrics part: Embed these proxy variables into the typical wage regression & variance decomposition and examine different wage components
3. Extensive analysis: Examine potential heterogeneity of skill prices & firm wage premium and the driver of inequality trend

Data: Basic Info

`Lagou.com`: the largest IT-centered online job board in China (mostly ”cognitive jobs”)

- Over 6 million vacancies between 2013 and 2020 ▸ vacancy trend
- Mainly jobs in all occupations demanded by IT-producing/using firms: Computer, Design & Media, Business Operation, Financial & Law, Sales, Admin ▸ occupation classification
- Like other vacancy data, biased to young/low-experienced and high education workers/jobs in large cities ▸ details & reliefs
- Vacancy information: job name, posted wage, location, requirements on education and experience, job task or skill description, job benefits, firm name, ... ▸ vacancy sample
- Final Sample after cleaning: 4 million vacancies ▸ sample cleaning ▸ summary statistics

Potential concerns: various data/sample representativeness issues ▸ details & reliefs

Posted Wage Regression

- Baseline: $\ln w_i = X_i\beta + \psi_j + \iota_t + \epsilon_i$
  - w_i is the mean of the posted wage scope
  - X_i is a vector of job characteristics, denote $\theta_i \equiv X_i\beta$
  - ψ_j is the firm effects
  - ι_t is the year effects

- Estimated β will be the market average prices of the job characteristics

- Estimated ψ_j will be the firm-specific wage premiums/discounts for any reasons

- $\hat{\beta}$ and $\hat{\psi}_j$ would be biased if $\text{cov}\left(X_i, \epsilon_i\right) \neq 0$ and $\text{cov}\left(\psi_j, \epsilon_i\right) \neq 0$

- $\text{var}\left(\ln w_i\right) = \underbrace{\text{var}\left(\theta_i\right)}_{\text{Job Effect}} + \underbrace{\text{var}\left(\psi_j\right)}_{\text{Firm Effect}} + \underbrace{2\,\text{cov}\left(\theta_i, \psi_j\right)}_{\text{Firm-Job Sorting}} + \text{var}\left(\epsilon_i\right)$

Overview of ML Procedures

▸ Jump to Results

1. Feature Selection: *110,000+* $\rightarrow$ *3100+*
   Transform vacancy documents $\mathbf{D}$ to an indicator matrix $\mathbf{C}$ ($N \times K$), where $K = |V|$;
   Run Lasso regression of $\ln w$ on $\mathbf{C}$ to shrink the entire vacancy text vocabulary set V V to a vocabulary subset V' (and $\mathbf{C}$ to $\mathbf{C}'$)

   ▸ Lasso detail ▸ Lasso turning by BIC ▸ Lasso inference & sanity check

2. Feature Clustering: *3100+* $\rightarrow$ *8 groups*
   Train a word embedding model (Word2Vec) on vacancy text $\mathbf{D}$ to obtain the embedding space representation for selected features: $\mathbf{U}' \equiv \{\mathbf{u}_k\}$ where $k \in V'$;
   Apply K-Means classifier to $\mathbf{U}'$ generate P $(= 8)$ clusters $\{V'_p\}_{p=1}^{P}$

   ▸ word embedding detail ▸ K-Means detail ▸ a data driven skill & task space ▸ a data driven skill & task space

3. Dimensional Reduction: *3100+* $\rightarrow$ *8* $\times$ *3 = 24*
   Use PLS to transform each $\mathbf{C}'_p \equiv \{\mathbf{c}_k\}$, $k \in V'_p$ into a low dimensional representation Ξ_p ($N \times Q$; $Q = 3$) and obtain $\{\Xi_p\}_{p=1}^{P}$

   ▸ dimensional reduction detail

Feature Clustering: Skill/Task Structure

▸ Overview

A data-driven skill/task structure shows layers of specificity ▸ specificity measure

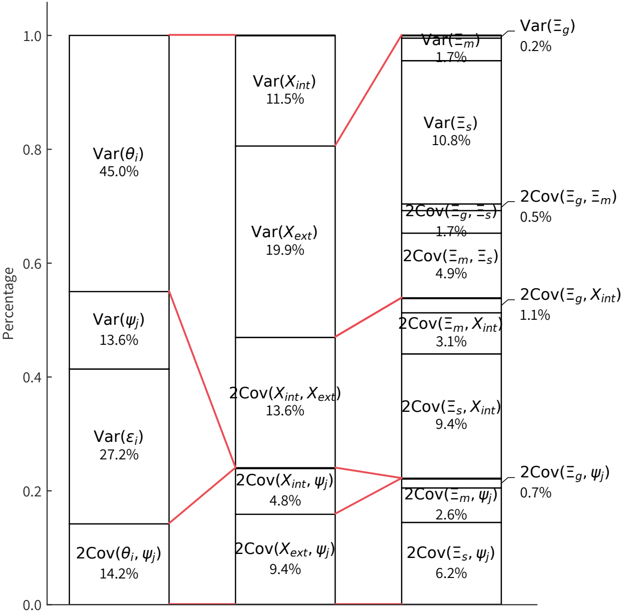
0. Compensation (V'_c)
1. General skills (V'_g)
    - Cognitive: e.g. logic, self-learning
    - Interpersonal: e.g. communication, extrovert
    - Non-cognitive: e.g. hard working, responsibility
2. Education-related or -extensive skills (V'_e)
    - e.g. education level, college majors, certificates, fundamental occupational skills, basic field experience
3. Occupation-specific skills and tasks ($V'_{s1}, \ldots, V'_{s5}$)
    - e.g. c++, python, graphic design, logistic management, audit, business negotiation, client responding, ...

(way more granular than cognitive/social/... dimension or traditional occ dimension)

Proxy Variables on Skills & Tasks

- Under our construction, $\{\Xi_g, \Xi_e, \Xi_{s1}, \dots, \Xi_{s5}\}$ proximate to a full set of skills/tasks required in the vacancy that are predictive for posted wage
- Our final specification of job controls: $X = \{X_{ext}, X_{int}\}$
  - $X_{ext} \equiv \{\text{EDU}, \Xi_g, \Xi_e, \Xi_{s1}, \dots, \Xi_{s5}\}$, (extensive margin)
  - $X_{int} \equiv \{\text{EXP}\}$ (intensive margin) compare R2
- We further split X_{ext} into three groups:
  - Most general group: Ξ_g
  - Medium specific group: $\Xi_m \equiv \{\text{EDU}, \Xi_e\}$
  - Most specific group: $\Xi_s \equiv \{\Xi_{s1}, \dots, \Xi_{s5}\}$

Variance Decomposition



Variance Decomposition: Robustness

- Limited mobility bias is limited as long as firms have enough number of vacancies
  ▸ bias correction

- Education or Experience composition does not drive our results ▸ conditional on EXP & EDU

- Switching Ξ_4 from Ξ_s to Ξ_m has strongest impact on Admin sample ▸ $\Xi_m \equiv \{\text{EDU}, \Xi_4\}$

- Can still largely replicate the results in Deming and Kahn (2018) ▸ replicate DK ▸ app

- Non-wage compensation terms selected by Lasso largely because they can predict job and firm effects ▸ add Ξ_0 into regression

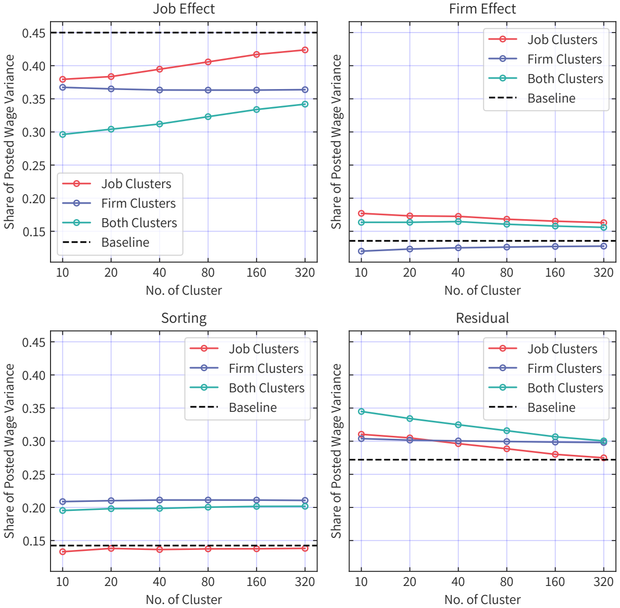
- Estimated firm wage premium are positively correlated with firm size (conditional on sorting) and accounted by firm location, consistent with the literature ▸ firm FE regression

- Mean residuals by firm-job cells show that the linear (additive separability) assumption seems to be a worse approximation in pooled sample ▸ mean residual distribution

A Shortcut

- Occupation is itself a concept born from skill/task specificity, though too coarse
- Bonhomme et al. (2019) suggests another way to solve the finite sample bias: estimating latent firm groups: $\min_{\mathfrak{k}_1,\ldots,\mathfrak{k}_J,H_1,\ldots,H_{\mathfrak{K}}} \sum_{j=1}^{J} n_j \int \left(\widehat{F}_j(y) - H_{\mathfrak{k}_j}(y)\right)^2 d\mu(y)$
- Here we can also use our embedding space representation to classify latent job groups:
  - First, for each vacancy: $\mathbf{z}_i = \sum_{k \in V_i} \mathbf{u}_k = (z_{i1}, \ldots, z_{iH})$
  - Then, $\min_{\{\mathfrak{l}_1,\ldots,\mathfrak{l}_I,G_1,\ldots,G_{\mathfrak{L}}\}} \sum_{i=1}^{I} \sum_{h=1}^{H} \left(z_{ih} - G_{\mathfrak{l}_i}(h)\right)^2$
  - This can be seen as a way to generate occupations with arbitrary number $\mathfrak{L}$

A Shortcut


Job Effect
Firm Effect
Sorting
Residual
Share of Posted Wage Variance
No. of Cluster
0.45
0.40
0.35
0.30
0.25
0.20
0.15
10
20
40
80
160
320
Job Clusters
Firm Clusters
Both Clusters
Baseline

Summarizing Main Results

1. At least for this market, our estimated shares of wage inequality components (45.0% job effect; 13.6% firm effect; 14.2% sorting) are consistent with the literature
2. Our approach shows a data-driven skill/task structure featured by different specificity levels
3. For the posted wage variations from job effect and firm-job sorting
   - Extensive margins account for 2/3; Intensive margin (Exp) accounts 1/3
   - Occupation-specific skills/tasks account for the major shares, esp. in high-skill occ
   - Education-related skills/tasks account for more shares in low-skill occ
   - General skills, whether cognitive, interpersonal, or noncognitive, barely matter (here)
4. Levels of skill prices, firm wage premiums, & sorting vary across occupations
5. Increased posted wage variance in our data is largely driven by increased sorting, esp. from those occupation-specific skills/tasks

Outline

Chapter 1. Establishment Dynamics in Post-War Japan

Chapter 2. Post Wage Inequality

Chapter 3. Post Compensation Inequality

Chapter 4. Japanese Programmer and Technology Adoption

Research Questions

Empirical:

1. What consists non-wage compensations in today's labor market?
2. Do firms distinguish in their provision of amenities/disamenities? How?
3. What are their impact on wage disparity?

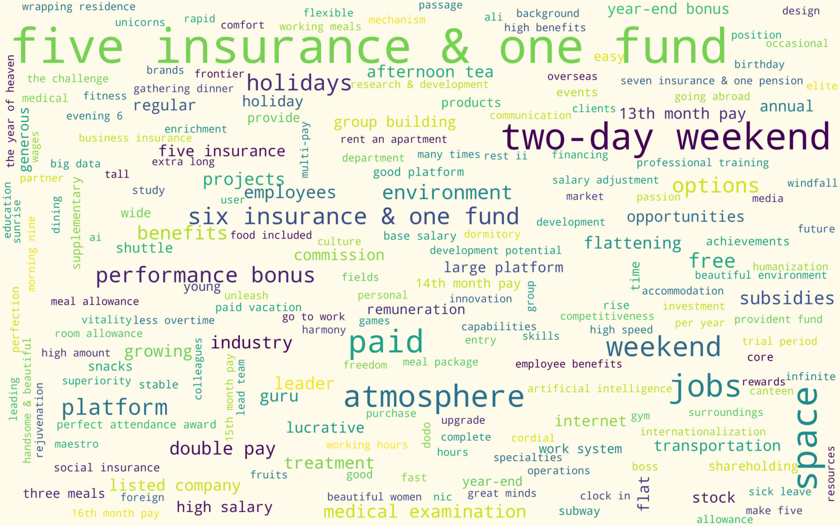
Theoretical:

1. Do observed firms' provision patterns consisting with existing theories?
2. Why empirical tests of compensating differential often fail?
3. What are general implications of non-wage compensations on labor market?

What This Paper Does

1. Investigate the provision patterns & wage effects of non-wage compensation (both pecuniary & nonpecuniary) by using job ads/vacancy data Pros & Cons
   - Difficult to observe in census/survey data
   - Extract info from job texts using (basic) ML methods
   - Find stylized patterns in the data
   - Discuss the inconsistency between findings and existing theories
2. Construct a new & simple theory to rationalize our empirical findings
   - Extend the idea of compensating differential with a new force
   - Reconcile our empirical findings and offer important implications

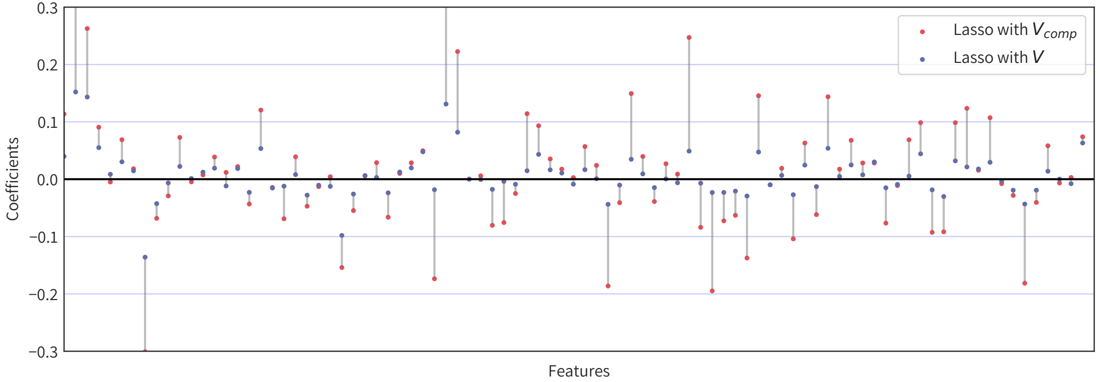
Fact 1: Firms Provide ”Common” Non-wage Compensations ◂ chinese



insurance&fund; leisure; growth potential, bonus, environment, fringe benefits, ...

Fact 2a: Firm Non-wage Compensations Correlated With Job Attributes

◂ Lasso top features using V_{comp} ◂ Lasso top features using V



◂ All V'_{comp}

Fact 2b: Compensations Explain Wage Differentials Through Linkage with (Both Job and) Firm Heterogeneity

posted wage regression details

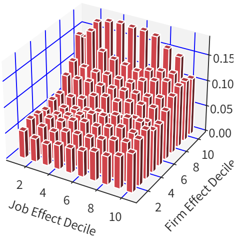
$$\ln w_{i,j,t} = \theta_i + \psi_j + \delta_i + \iota_t + \epsilon_i$$

	With δ		Without δ	
	Comp.	Share	Comp.	Share
$\mathrm{Var}(\ln w)$	.362	-	.362	-
$\mathrm{Var}(\theta_i)$	.158	.437	.163	.450
$\mathrm{Var}(\psi_j)$	.046	.128	.049	.136
$\mathrm{Var}(\delta_i)$	.002	.004		
$\mathrm{Var}(\epsilon_i)$	.097	.269	.098	.272
$2\,\mathrm{Cov}(\theta_i, \psi_j)$	.049	.137	.052	.142
$2\,\mathrm{Cov}(\delta_i, \theta_i)$	.006	.017		
$2\,\mathrm{Cov}(\delta_i, \psi_j)$	.003	.008		
$\mathrm{Corr}(\theta_i, \psi_j)$	.289		.288	
$\mathrm{Corr}(\delta_i, \theta_i)$	.193			
$\mathrm{Corr}(\delta_i, \psi_j)$	.174			
Obs	3998840		3998840	
Firm	86165		86165	

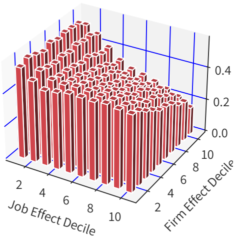
Fact 3: Systematic Differences in Compensation Provision Across Firms and Jobs

procedures more types

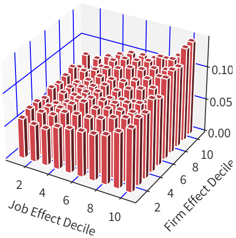
(a) Advanced Insurance



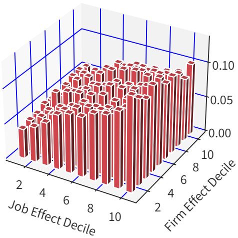
(b) Basic Insurance



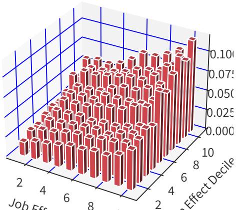
(c) Backloading Wage



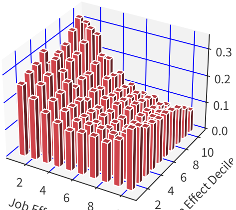
(d) Stock Option



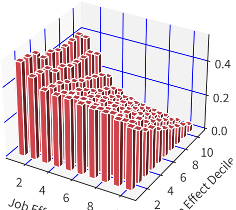
(e) Coworker Quality



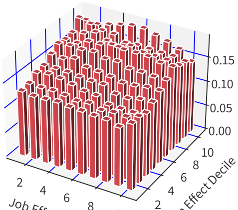
(f) Training



(g) Weekend, Holiday, Fixed Work-Time



(h) Work(-Time) Flexibility



Fact 4: Hedonic Regression Results are Mixed but in A Systematic Way

	(1)	(2)	(3)
Advanced Insurance	.117**	.087**	.014**
	(.001)	(.001)	(.001)
Backloading Wage	.054**	.030**	.010**
	(.001)	(.001)	(.001)
Stock Option	.114**	.058**	.087**
	(.001)	(.001)	(.001)
Coworker Quality	.140**	.059**	.024**
	(.001)	(.001)	(.001)
Work-Flexibility	.046**	.032**	.010**
	(.001)	(.001)	(.001)
Basic Insurance	-.062**	-.046**	-.025**
	(.000)	(.000)	(.000)
Training	-.057**	-.012**	-.003**
	(.001)	(.001)	(.001)
Work-Time	-.113**	-.081**	-.021**
	(.001)	(.000)	(.000)
Education FE	✓	✓	✓
Experience FE	✓	✓	✓
Year FE	✓	✓	✓
$\mathbf{C}_{\backslash comp}$		✓	✓
Firm FE			✓
Adj. R^2	.506	.633	.738
No. Obs	3998840	3998840	3998840

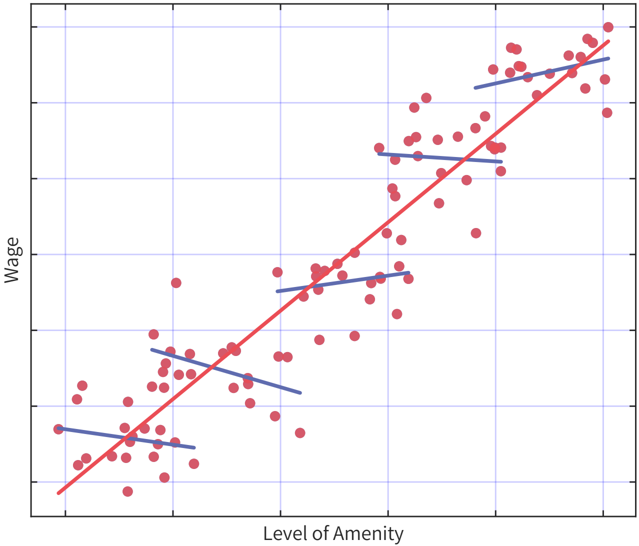
Summary of Empirical Findings & Implications on Theory

1. Firms use common non-wage compensations to attract job seekers: insurance, work-time, extra pay, workplace, ...
   $\rightarrow$ endogenous rather than exogenous variations in firm cost functions (& variations in worker preference?)
2. Non-wage compensations explain posted wage variance mainly via their correlations with job/firm effects
   $\rightarrow$ sorting is productivity-based; limited importance of compensating differential or co-determination with wage
3. Diff firms in diff jobs have distinct compensation-provision patterns
   $\rightarrow$ important mechanism of compensation provision linked with firm/worker quality
4. Hedonic regression shows systemically mixed results of compensating differential
   $\rightarrow$ reason of the empirical failures linked with the provision patterns

$\rightarrow$ These findings are inconsistent with the settings/views of compensating differential

Unobserved Worker Ability → Compensation Inequality?

◂ phantom of unobserved ability



Can Existing Theories Explain Positive Wage-Amenity Relationship?

- Hwang et al. (1992); Mortensen (2005): income effect
- Hwang et al. (1998): firms with low amenity-providing cost use both better amenity and higher wage to attract workers
- Problem 1: income effect cannot explain why it is low-pay firms provide leisure but not high-pay firms (e.g. notorious 996 working culture in Chinese IT industry)
- Problem 2: amenity-producing cost cannot explain why it is high-pay firms provide many superior amenities like insurance or backloading wages
- Problem 3: sorting is purely from exogenous heterogenous amenity-producing costs (and/or heterogenous worker preference) or wage-queue tradeoff

Model Overview

- We suggest a new theory that extends Compensating Differential with ”Efficiency Compensation” and productivity-based firm-worker Sorting

- Key idea: ”Efficiency” dimension
  1. Many compensations observed in data are (in)efficiency compensation
  2. The level of efficiency depends on firm & worker productivity

- Mechanism: A new channel works in addition to compensating differential
  1. When a compensation is efficient, it counteracts compensating differential effect
  2. When a compensation is inefficient, it magnifies compensating differential effect
  3. Extent of this (in)efficiency channel depends on firm-worker productivity sorting

$\rightarrow$ This simple modification reconciles all findings and generates many important general implications

Model Setting: Worker

- A continuum of worker with heterogenous productivity $q \in [0, 1]$ and additively separable (quasi-linear) utility function $U(C, a, h) = C + \phi_a a - \frac{h^{1+\phi_h}}{1+\phi_h}$
  - C is monetary consumption
  - $a \in \{0, 1\}$ is the indicator of a discrete amenity, e.g. insurance
  - h is a continuous disamenity, e.g. additional working hour

(Abstract from heterogenous preference)

Model Setting: Firm

- Firms are ex-ante homogenous with O-Ring production function:
  $Y_j = AN_j^{1+\alpha} \prod_{i=1}^{N_j} q_i e(a, h)$
  - N is assumed to be fixed exogenously ◂ can relax
  - Compensations are (in)efficient: $e(a, h) = 1 + \gamma_a a + \frac{h^{\gamma_h}}{\gamma_h}$
    (microfoundations: e.g. less exogenous or endogenous exit(Hwang et al., 1998; Dey and Flinn, 2005); convexity in hour productivity (Goldin, 2014))
  - Firm pay direct cost κ for a and compensate wage w for h

(Abstract from heterogenous (dis)amenity production function)

Competitive Equilibrium & Matching

- Competitive equilibrium in this economy is defined as an assignment of worker types to firms and a utility schedule, $u(q)$ such that
  - Firms maximize their profits
  - Labor market clears

- Complementary production function & additively separable utility function ensure positive assortative matching (PAM) even under imperfect transferable utility
  $\rightarrow$ each firm will employ workers with same q

  (Abstract from other-types of sorting)

Firms' Optimal Choices

- A firm chooses $\{q, a, h, w\}$ to maximize profit s.t. market utility schedule ◂ firm problem
- $a^* = \begin{cases} 1, & \text{if } q \geq q_a \\ 0, & \text{if } q < q_a \end{cases}$, and $\underbrace{AN^\alpha q_a^N \gamma_a + \phi_a}_{\text{mb}} = \underbrace{\kappa}_{\text{mc}}$
  - If a is not efficient, i.e. $\gamma_a = 0$, return back to the canonical compensating differential
  - If unit cost is $q\kappa$, higher q firms are still more likely to provide a
- $h^* = \left(AN^\alpha q^N\right)^{\frac{1}{1+\phi_h-\gamma_h}}$ increases in q
  - $h^*(q)$ will be fully compensated by $w(q)$, thus provision cost ex-post depends on q

Market Wage

market utility

- $$w(q) = \begin{cases} \bar{A}q^N + \underbrace{\gamma_a \bar{A}q^N - \kappa}_{\text{wage effect of } a} + \underbrace{\frac{\left(\bar{A}q^N\right)^{1+\omega}}{(1+\omega)(1+\gamma_h)} + \frac{\left(\bar{A}q^N\right)^{\omega}}{1+\gamma_h}}_{\text{wage effect of } h}, & \text{if } q \geq q_a \\ \bar{A}q^N + \underbrace{\frac{\left(\bar{A}q^N\right)^{1+\omega}}{(1+\omega)(1+\gamma_h)} + \frac{\left(\bar{A}q^N\right)^{\omega}}{1+\gamma_h}}_{\text{wage effect of } h}, & \text{if } q < q_a \end{cases}$$
  - Recall $\gamma_a \bar{A}q^N - \kappa = -\phi_a$ when $q = q_a$ and can be positive when $q \uparrow$
    $\rightarrow$ offsetting compensating differential
  - $\frac{\left(\bar{A}q^N\right)^{1+\omega}}{(1+\omega)(1+\gamma_h)}$ is the efficiency gain from h; $\frac{\left(\bar{A}q^N\right)^{\omega}}{1+\gamma_h}$ is the compensation for h
    $\rightarrow$ magnifying compensating differential

Model Implications

1. Testing compensating differential: Compensating effects can be confounded with productivity effects; Available variations for wage-amenity packages can be limited conditional on worker

   $\rightarrow$ *Field/choice experiments (WtP) or RCT-like experiments (exogenous variations) not necessarily capture the whole picture of how labor market works*

2. Labor market inequality: Efficiency compensations can enlarge both utility dispersion & wage dispersion

   $\rightarrow$ *Increased sorting or better use of efficiency compensations increases wage inequality*

3. Job mobility: The set of non-wage compensations that can justify job moves to low wage-premium firms is likely limited to inefficient amenities

   $\rightarrow$ *Potential implications for gender wage gap and etc.*

Outline

Chapter 1. Establishment Dynamics in Post-War Japan

Chapter 2. Post Wage Inequality

Chapter 3. Post Compensation Inequality

Chapter 4. Japanese Programmer and Technology Adoption

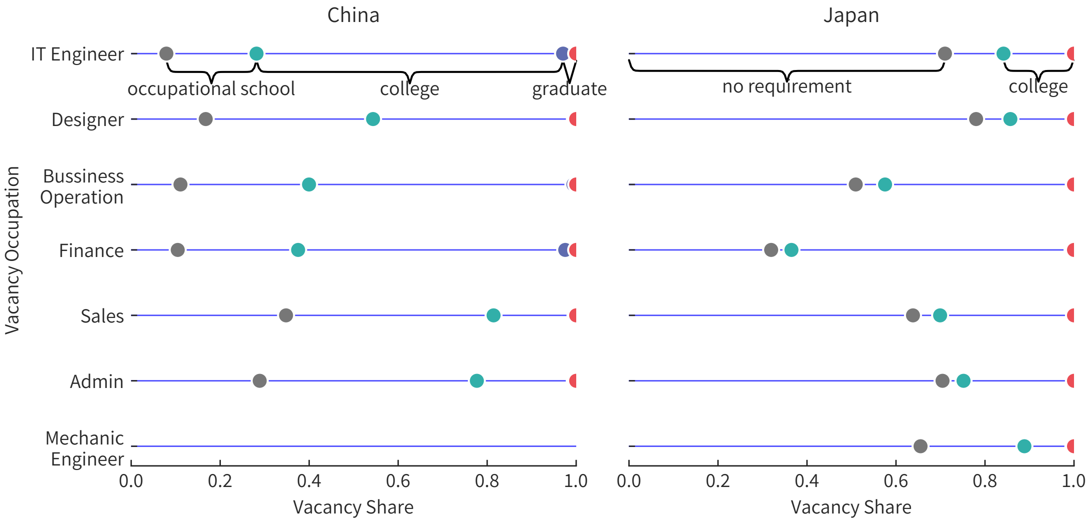
Research Questions

- Who (should) provide human capital investment for new skills under new tech? Firms (Training)? Workers (Learning)?
- Literature suggests both can, but their incentives diverges
  - Becker (1964): only workers have incentive in competitive labor market
  - Acemoglu and Pischke (1998, 1999a,b): firms also have incentive under labor market imperfections
- Less stress on efficiency differences:
  - Assume that can achieve optimal investment as long as one party is sufficiently incentivized
  - If efficiency differences exist, there can have mismatch with incentive structures
- Market structures and institutions that determine the incentive structures often exogenously given

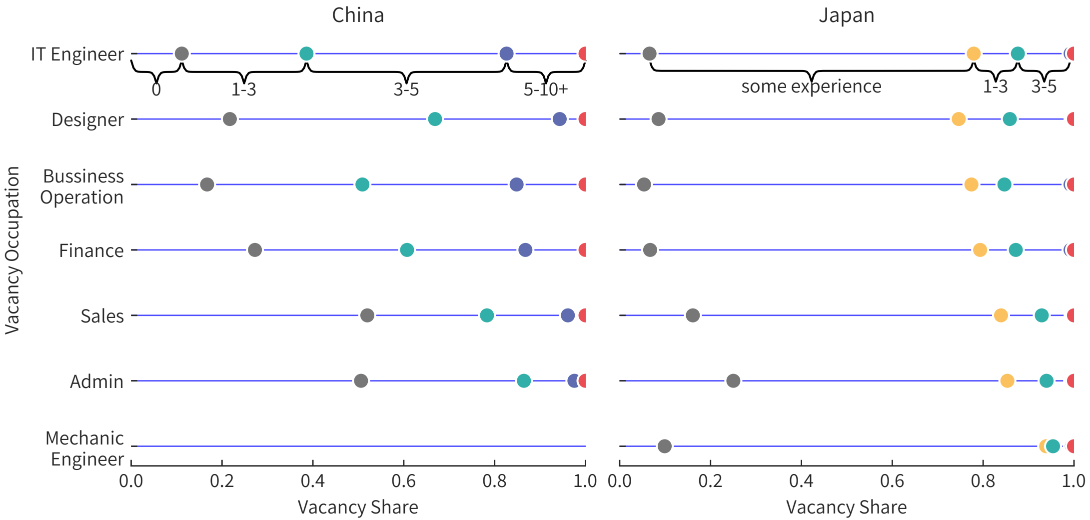
This Paper

- Study human capital investment and technological adoption behaviors/outcomes under different markets by comparing the IT industries in Japan and China
  - Utilize online vacancy data to identify otherwise hard-to-observe info ◂ data source
- Find distinct empirical features
  - Japanese IT firms: lower edu/exp requirements; more on-the-job training; less advanced technologies and skills; less wage premium ◂ other evidences from literature
- Build a simple model to explain why distinct qualitative results
  - Assume workers have higher investment efficiency than firms for IT technology
  - Illiquid labor market in Japan suppresses worker investment but encourages firm investment, despite its relative inefficiency
- Show that this model can also explain why endogenous labor market institutions emerge and/or resist to changes
  - Key is to allow relative efficiency contingent on the prevailing technological regime
  - Incumbent firms have limited incentives to change the institutions

Requirements on Education

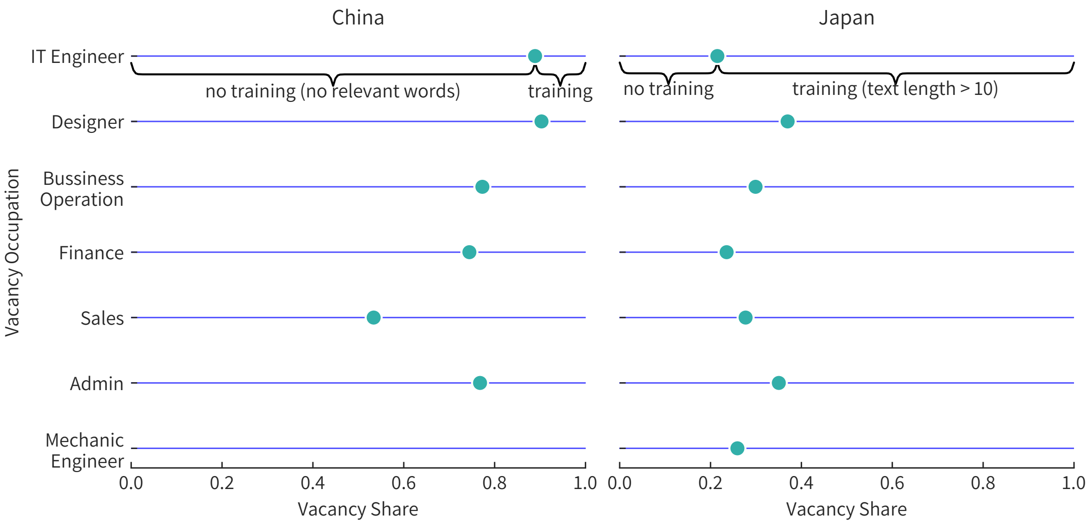


Requirements on Experience

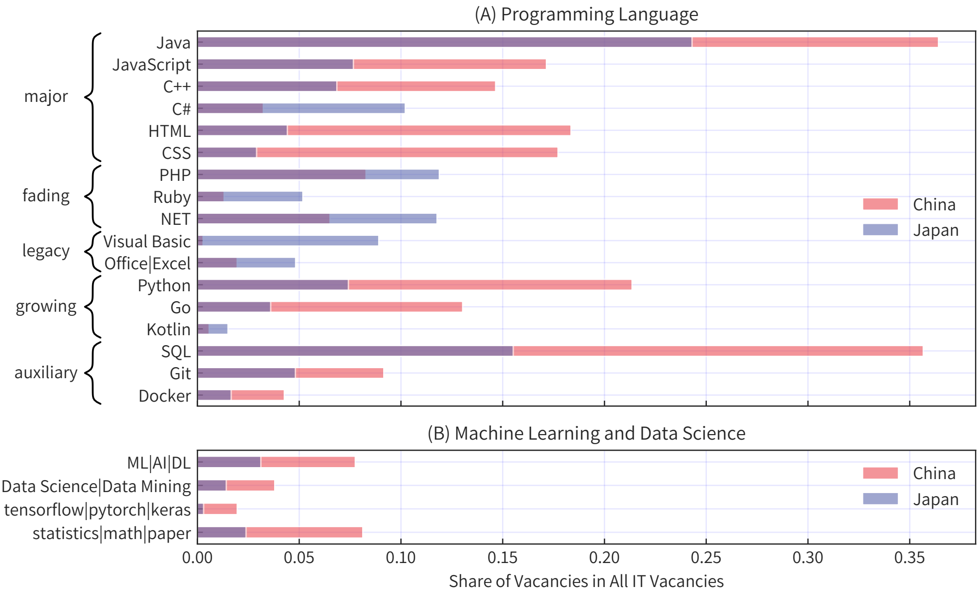


Provision of Training

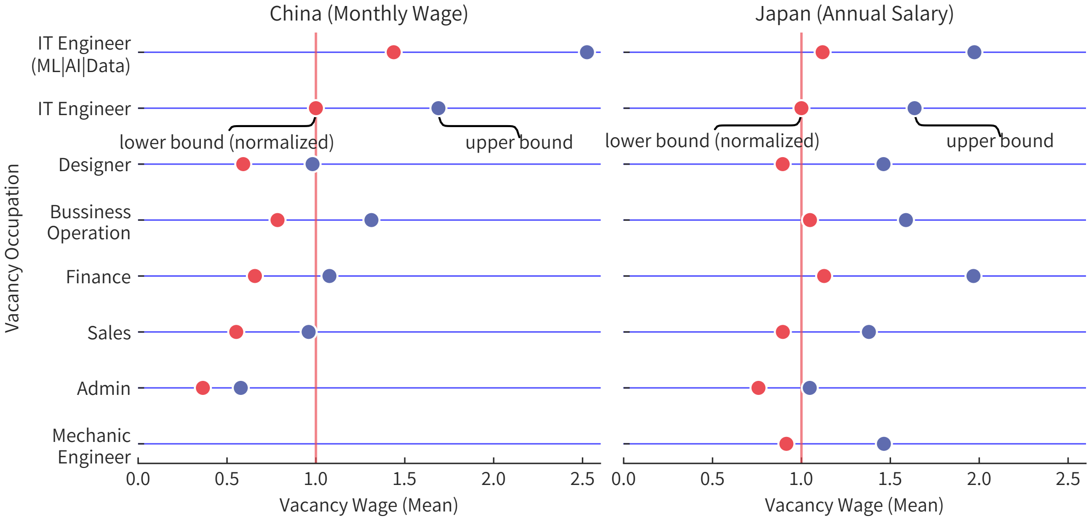
◂ distribution of training text length



IT Skills and Technologies Mentioned

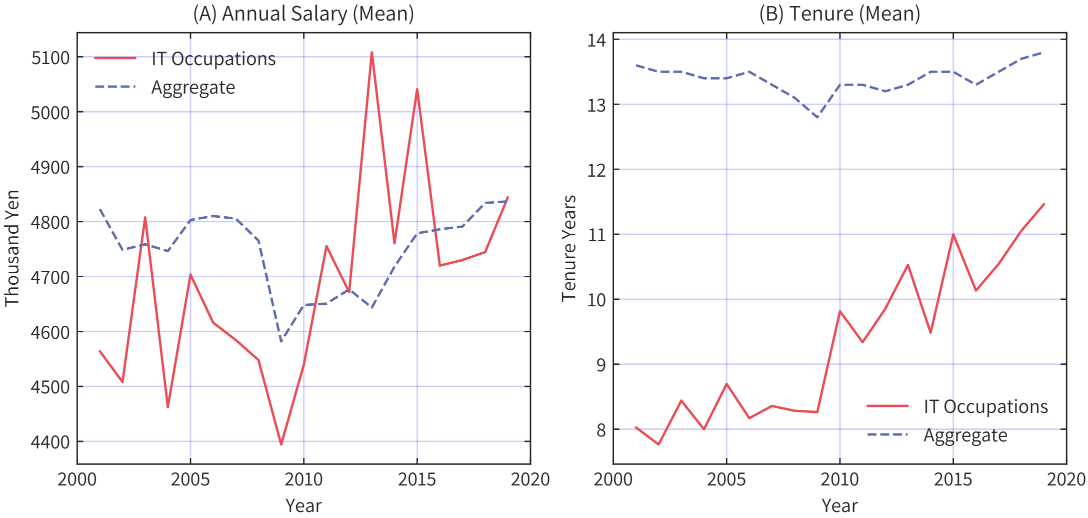


Posted Wage



▸ Monthly Wage ▸ Un-normalized

Japanese Census Data (BSWS)



US Census Data
China JJ Rate

Model Environment

A simple two-periods model of training/learning and production

- 1st period: a mass one of works and a mass one of firms match one-to-one randomly and then produce

- To simplify, assume homogenous endowments (and risk neutral agents; no discounting)
  - Workers have initial general human capital $h_0 = 1$
  - Firms have productivity $z = z_1$

- Production technology: $f(z, h) \equiv zh$

- Training technology (invest noncooperatively): $\Delta h(k, l) = Ak^{\alpha} l^{(1-\alpha)}$
  - Workers' input can be effort or leisure, with utility cost $\kappa \frac{l^{1+\gamma}}{(1+\gamma)}$ (thus no credit constraint)
  - Firms' input can be capital or any other training costs, with unit cost r

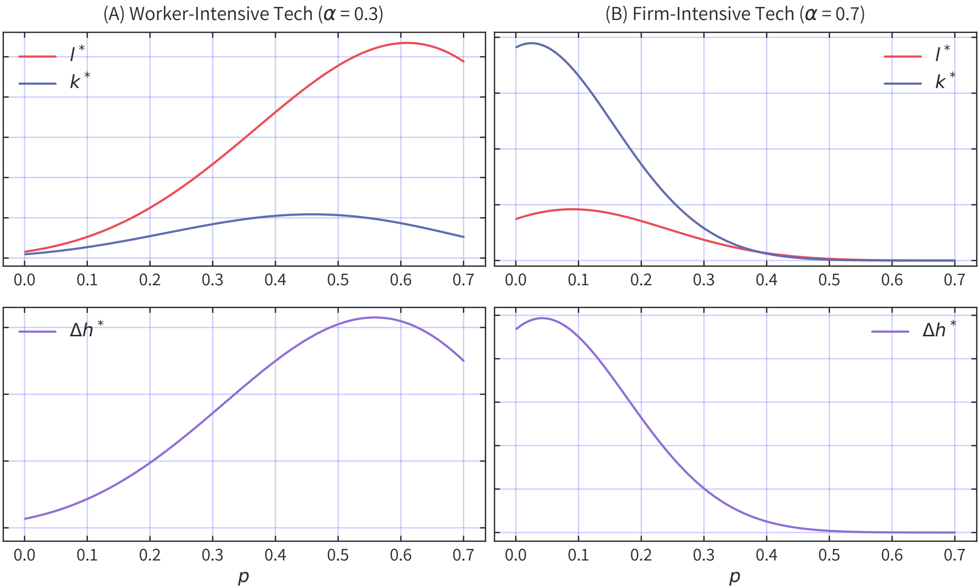
Model Environment (cont')

- 2nd period: there is a large mass of potential new entrants of firms
  - They can pay an entry cost c to open a vacancy
  - Assume they draw productivity $z = z_2 > z_1$
- Labor market is frictional; Employed workers do on-the-job search (abstract from unemployment)
  - Number of new matches: $m = M(v, s) = \xi v^{\phi} s^{1-\phi}$
  - Matching rate for workers: $p = \xi (v/s)^{\phi}$; for poaching firms: $q = \xi (v/s)^{\phi - 1}$
  - Normalize search effort: $s = 1$ (can endogenize search effort)
- Wages are determined by Nash bargaining, with worker bargaining power β

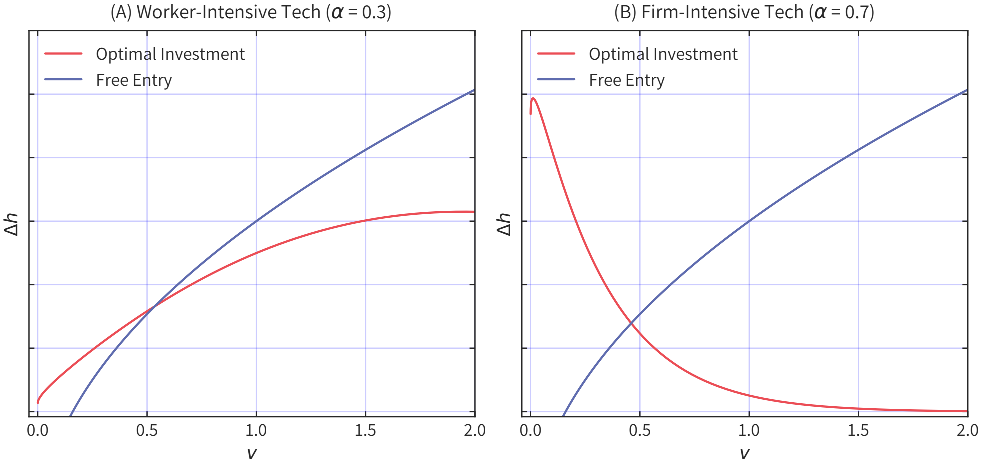
Value Functions

- Worker value: $W(l; k, v) =$
  $\beta z_1 - \kappa \frac{l^{1+\gamma}}{(1+\gamma)} + \left[(1 - p(v))\, \beta z_1 + p(v)\, (z_1 + \beta(z_2 - z_1))\right] (1 + \Delta h(l, k))$
  - Worker's outside option when bargaining with new firm is $\beta z_1 (1 + \Delta h(l, k))$
- Firm value: $F(k; l, v) = (1 - \beta) z_1 - rk + (1 - p(v))(1 - \beta) z_1 (1 + \Delta h(l, k))$
- Free entry condition: $q(v)(1 - \beta)(z_2 - z_1)(1 + \Delta h(l, k)) \geq c$

Investment/Training Under Different Technologies



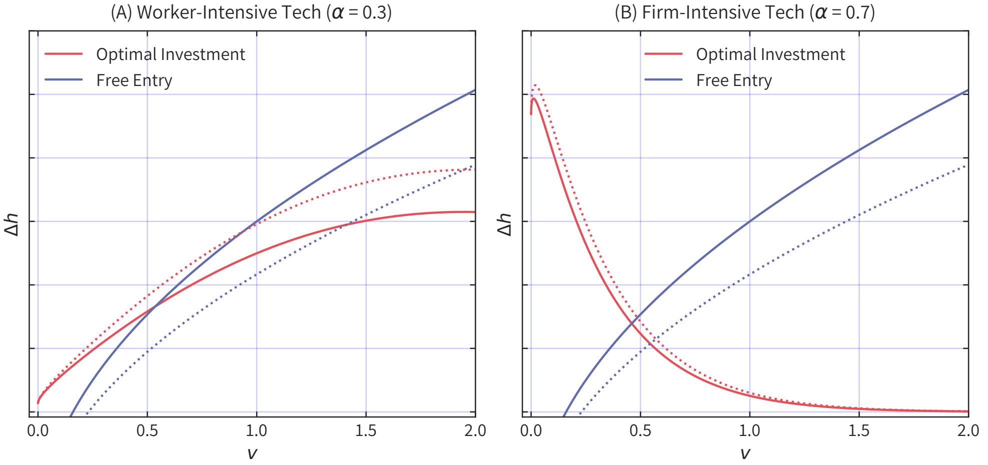
Equilibrium (and Inefficiency)



Model Implications

- The optimal investment requires different labor market structures under different technological regimes
  - If worker's efficiency in training is high in IT technology (small α), the illiquid labor market in Japan will generate large incentive mismatch and result low skill acquisition or tech adoption (Δh)
  - Wage premium will be low due to both low Δh and low p
- A similar logic can explain why the Japanese labor market institutions built at early post-war periods
  - Heavy manufacturing industries require firms to have incentive to invest
- Existence of a large amount of well-established incumbent firms will likely to generate resistance for regime changes
- We conjecture that China circumvent the Japanese path by utilizing separate labor markets (state-owned illiquid & private liquid markets) ◂ details

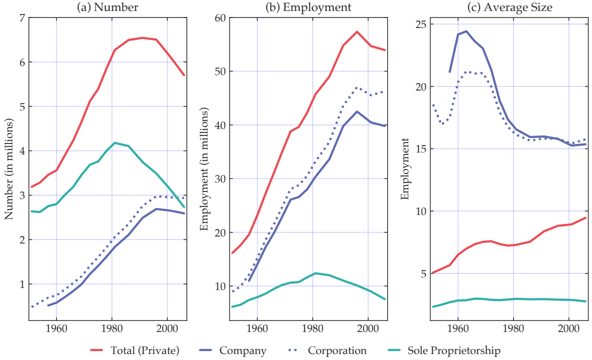
Increase in Technology Gap z_2



Thanks!

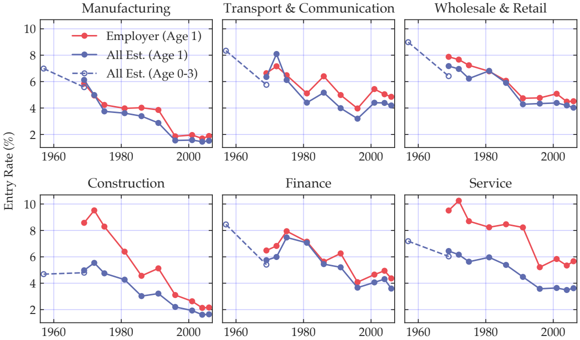
Appendix for Chapter 1.

”Employer” and ”Nonemployer”



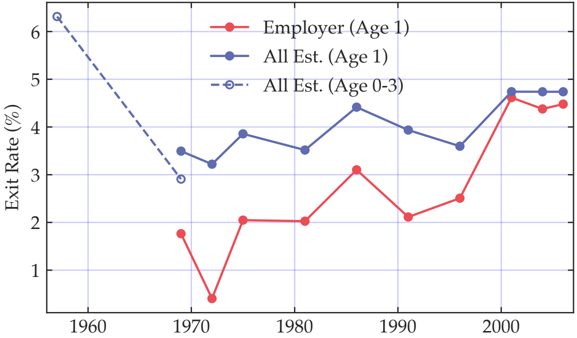
- Could it be more ”nonemployer” turning to ”employer” that drug down average size?
- Not likely. Because i) These two groups have similar trend on entry rate; ii) Nonemployers have more shares in Wholesale&Retail and Service sectors, where we see the least decline in average size; iii) There are larger initial and on-going costs for ”employer”, thus a change in organization type can be regarded as a de-facto ”entry”

Declining Entry Rate by Industry ◂ Back



- This decline is pervasive across all sectors and industries
- This decline is also shown in firm statistics ▸ firm entry rate

Low and Stagnated Exit Rate ◂ Back



- Calculated based on entry rate
- Declined before 1970 but then stagnated at very low level (2% per year) thereafter until the end of 1990s
- Decreasing entry rate could contribute to this low exit level since young establishments are more likely to exit

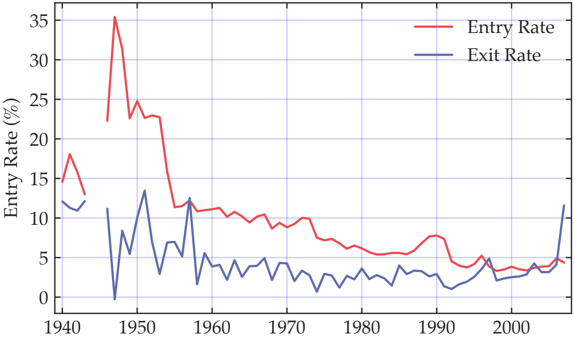
One Natural Result is The Aging of Businesses in Japan

Back

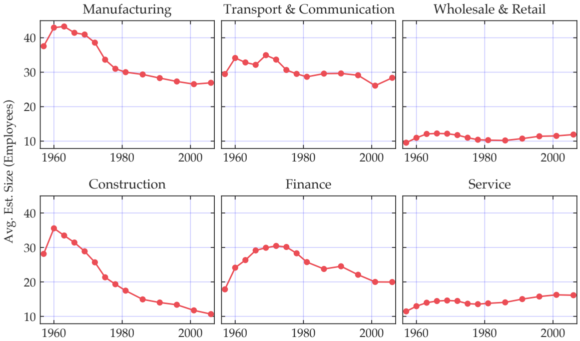
- A nature result of declining entry and low exit rate is the decreasing share of the young business units in the economy and aging of the establishment population in Japan
- In 2001, nearly 35% of the employees in Japan work at an establishment of 27+ years old

Firm Entry/Exit Rate

Back



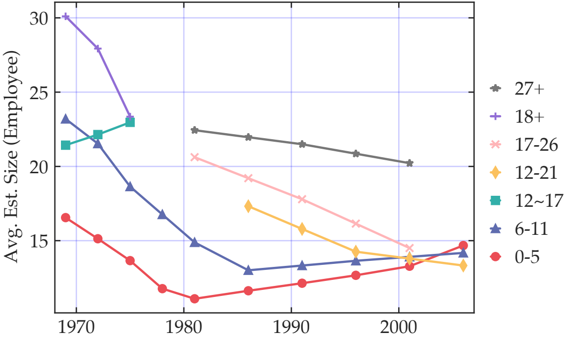
Average Size Declines Diverge by Industry ◂ Back



- Manufacturing and Construction industries decline the first (since early 1960s) and the most
- Wholesale&Retail and Service sector seems to be more resilient to this decline, and recovered since 1980s.

Average Size Decline Diverges by Age

Back

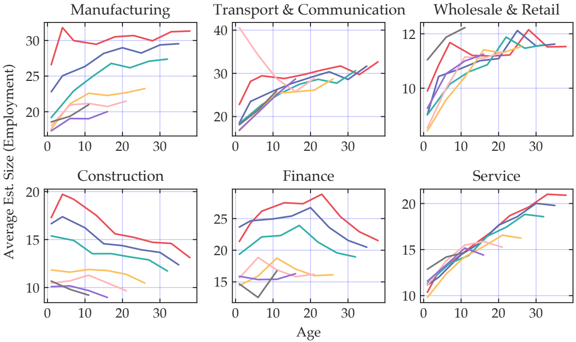


By Industry

- Before 1980, the average size declines in most age groups
- However since 1980, the average size of the young establishments began to recover, while elder ones kept declining in census
- Note that the change of the average size of an age group over time depends on two dimensions: initial level and life cycle growth

Birth Cohort Effect Also Diverges by Industry

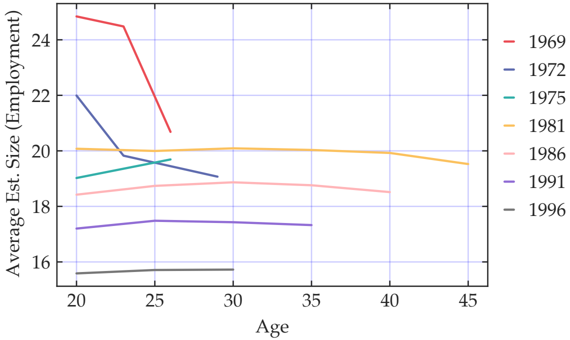
Back



- With large difference in average entrant size over time in Manufacturing and Construction, we see clear birth cohort effect
- In Wholesale&Retail and Service sector, it seems that the life cycle growth paths are more likely to converge despite the time-variant average entry size

Life Cycle Growth of Early Cohorts

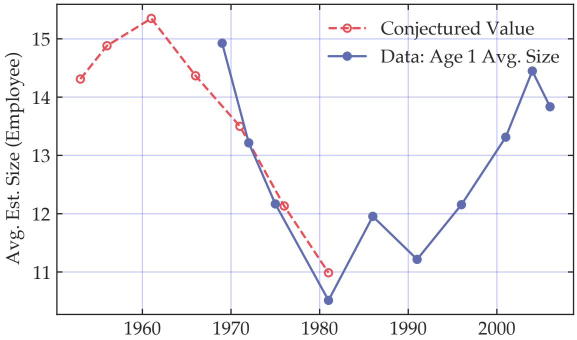
◂ Back



▸ By Industry

- Using the same imputation method, we can confirm the non-converged life cycle growth even for cohorts born in 1960s and even before
- Moreover, we can confirm that the forces that led to the decline in avg. size in 1960s and 1970s also affected the elder groups, generating average size decline for even aging establishments

Conjectured Average Entrant Size in Early Periods



- Apply the average life-cycle growth of the birth cohorts in 1969-1981 to the average size of old groups in census after 1981, we back out the average entry size in early periods when no age data exist
- Just like the trend of average size, the average entry size saw a turning point in around 1960

Life Cyle Growth in Manufacturing (Hsieh & Klenow 2014)

Back

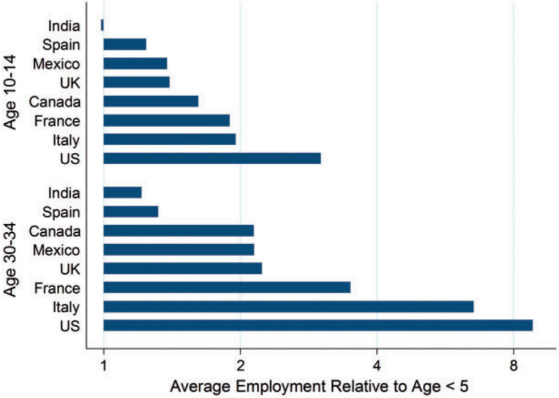


FIGURE IX

Employment Growth over the Life Cycle

Employment growth by age 10–14 and age 30–34 relative to age < 5. Indian data are from plants in the 2009–2010 ASI/NSS. Data for France, Italy, and Spain are for firms in the 2006–2007 Amadeus Database. U.K. data are for plants from 1997–2001 to 2002–2006 in the ARD. Canadian data are for plants from 1999–2001 to 2004–2006 in the Canadian ASM. See Appendix I for additional details.

Summarizing Facts

1. Persistent decline in market dynamism in Japan since 1950s
   - Potential fundamental long-term deriver since early post-war period
   - Less likely for drivers stressed during lost decades: e.g. ”zombie” firms or financial policy
2. Establishment size decline in 1960s and 1970s
   - A strong force reduce average est. sizes for all ages esp. in manufacturing and construction sectors
   - Puzzle as literature documents a positive relationship between development and firm size (except Portugal 1980-2010)
3. Lifecycle growth downward shifted over time
   - Entrants size decline thus has a feedback effect over time through the cohort effect of life cycle growth (esp. strong given the low levels of entry/exit)
   - Thus history matters for recent est. dynamics and demographics

Benchmark Model: Calibrating to Period Average

Back

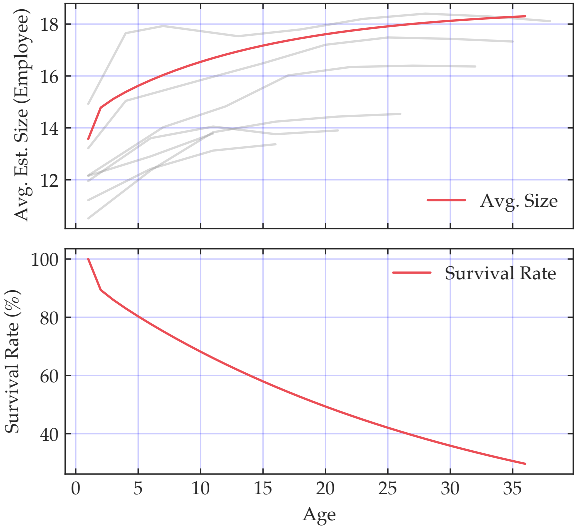
- Value of β, θ follows the literature; η is the peorid average value from data; and the others parameters are calibrated jointed

Parameters	Values	Definition	Calibration
β	0.96	Discounter factor	Assigned
θ	0.64	Labor share (”span of control”)	Assigned
η	0.02	Average labor force growth rate	Assigned
c_e	76.050	Entry cost (in unit of product)	Jointly Calibrated
c_f	2.123	Operation cost (in unit of labor)	Jointly Calibrated
a	0.008	Drift in AR(1)	Jointly Calibrated
ρ	0.966	Persistence in AR(1)	Jointly Calibrated
σ_ε	0.181	Std. of AR(1) shocks	Jointly Calibrated
μ_G	1.200	Mean of entrant productivity (log normal)	Jointly Calibrated
σ_G	0.527	Std. of entrant productivity (log normal)	Jointly Calibrated

- Entry cost c_e is large in order to pin down the low entry and exit rate in Japan

Benchmark Model: Life Cycle Growth and Survival Rate

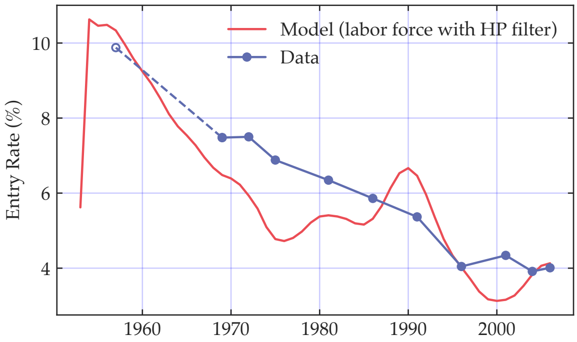
Back



- The benchmark model simulates a life cycle growth similar to the early period of the data. The model growth would be higher in elder period because the evolution of productivity (AR1) in model is non-decreasing in expectation
- The survival curve shows that in our model around 50% of the entrants can survive for 20 years

Declining Labor Force Can Drive Long-Run Entry Decline

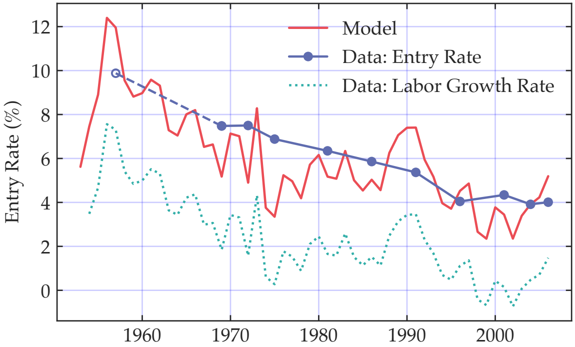
Back



- Quantitatively, the labor force growth decline can account for at least 2.4 percent points in the 3.5 percent points entry rate decline btw. 1969-2006

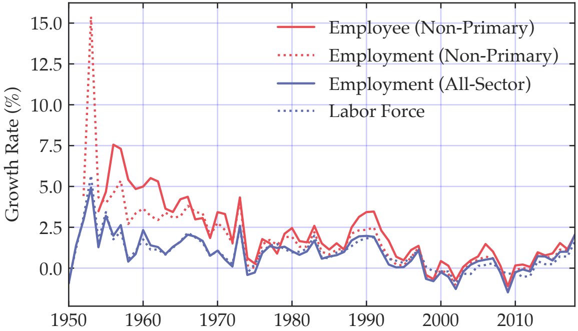
Declining Labor Force Can Drive Long-Run Entry Decline But

◂ Back



- The simulated entry rate is completely driven by the changes in labor force growth rate
- In theory, a decline in entry rate would lower exit rate and enlarge average size due to changes in age composition
- These changes should generate feedback effects through incumbent est. labor demand, further reducing entry rate over time. But we don't see these effects here

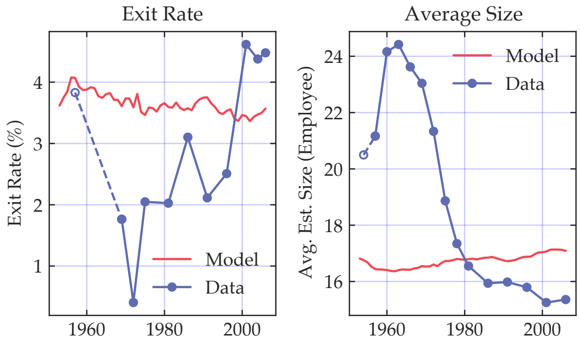
Labor Force Growth Rate ◂ Back



(Source: Labor Force Survey)

Feedback Effect Is Weak And At Odds With Data

Back



- In our empirical case, due to the fairly low exit rate and life cycle growth in Japan, these feedback effects are very week
- Also these potential effects are qualitatively at odds with the changes of exit rate and average size in the data
- There is also no effect on entry size and life cycle growth.

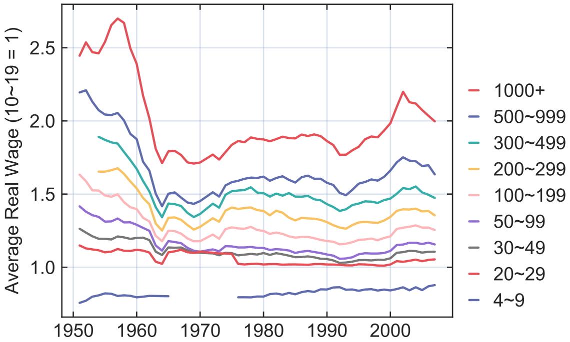
Combined Labor Force Decline With Traditional Explanations

A 2pp decline in labor force growth rate + A further 1pp entry rate decline led by other derivers

	Benchmark	Labor Growth	Entry Cost	Exit Value	Fixed Cost
η, %	2.00	0.00	0.00	0.00	0.00
c_e	76.05	-	99.88	-	-
V^x	0.00	-	-	-10.35	-
c_f	2.12	-	-	-	1.39
w^*	0.98	0.98	0.89	0.96	1.03
$\bar{x}^*$	1.32	1.32	1.09	1.09	1.09
Entry Rate, %	5.62	3.43	2.46	2.46	2.46
Exit Rate, %	3.62	3.43	2.46	2.46	2.46
Avg. Entry Size	13.57	13.57	17.29	14.22	11.30
Avg. Entry Size (after exit)	14.89	14.89	18.15	14.90	11.86
Avg. Est. Size	16.82	17.31	18.98	15.57	12.40
LifeCycle Growth Rate 10y, %	21.88	21.88	9.01	8.74	9.01
LifeCycle Growth Rate 20y, %	29.72	29.72	10.68	10.36	10.68

- Now all 3 cases generate moderate decline in lifecycle growth
- The case of fixed cost decline also well fits a decline in both entrant and overall average size
- However robust? And the nature of the fixed cost is quite abstract, which mainly implies a cost decline in the operation of the young establishments.

Empirical Problems with Distortion Explanation



(Source: Manufacturing Census)

- The distortion should be generated in 1960s and 1970s
- Wage inequality across establishment size groups in manufacturing declines in early 1960s and doesn't increase too much thereafter
- Other implicit labor cost distortions?

Appendix for Chapter 2.

Data Concerns & Reliefs

◂ Back Intro ◂ Back Data

- Vacancy data may be selective or less representative
  - Vacancy data is incline to young and more educated workers, esp. here
  - Not all jobs on the internet or different post frequency than job composition
  - Ideal match but not real match results
  - Only entry wage thus missing (re-)bargaining, discrimination, promotion, rent-sharing, revealing of worker ability or matching productivity, ...

(Valid issue for all vacancy data; Partially justified in the literature; Extent is an empirical question; Can improve with better data and adjust composition; Better fit liquid labor market; Not all bad for estimation)

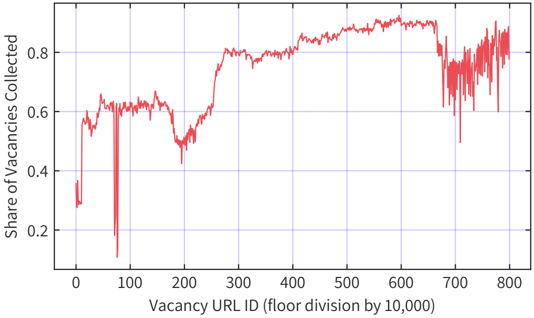
- Our wage measure incorporates variation in hours
  - One might worry that wage variation could be thus over-estimated
  - One might worry that those efficient compensations are solely compensating more working hours

(Often additional pay for overtime hours; Variation is limited comparing to wage; Inequality is often considered on overall compensation level; Need to think hour and wage as a package)

Trends on Collected Vacancies

◂ Back

# of Vacancies by Posted Month



% of Vacancies by Post ID Chunks

A Sample Vacancy

◂ Back Intro ◂ Back Data


Job Title
iOS开发工程师
Wage
18k–22k
(该职位已下线)

深圳 / 经验1年以下 / 本科及以上 / web前端 / 全职
内容资讯 短视频
Basic Job Info

收藏
已下线

完善在线简历
上传附件简历

字节跳动 2018-09-10 发布于拉勾网
Post Info

查看原职位详情

职位诱惑:
Job Benefits
六险一金，弹性工作，免费三餐，餐补，租房补贴，带薪休假，扁平管理，晋升空间，团队氛围好

职位描述:
Job Description and Reqirement

职位职责:
1、负责产品迭代改进及移动新产品的开发;
2、参与 APP 性能、体验优化及质量监控评估体系建设;
3、参与客户端基础组件及架构设计, 推进研发效率;
4、参与 hybrid 容器搭建，插件、React Native 等动态技术调研。

职位要求:
1、本科及以上学历，计算机相关专业;
2、热爱计算机科学和互联网技术，对移动产品有浓厚兴趣;
3、扎实的数据结构和算法基础；精通至少一门编程语言，包括但不限于：Objective-C、Swift、C、C++、Java;
4、熟悉 iOS平台原理，具备将产品逻辑抽象为技术方案的能力;
5、关注用户体验，能够积极把技术转化到用户体验改进上;
6、对新技术保持热情，具备良好的分析、解决问题的能力。

工作地址
深圳 - 南山区 - 广东省深圳市南山区南海大道2163号来福士广场15层
Work Address 查看地图

ByteDance 字节跳动
Firm Info
字节跳动
内容资讯,短视频
D轮及以上
2000人以上
http://www.bytedance.com

Sample Cleaning

Back

- Drop vacancies with not full-time jobs, outlier wages, job descriptions less than 20 words, nonChinese content
- Drop vacancies in 2013
- Drop vacancies from firms with less than 10 posts and from all the locations that have less than 1000 vacancies
- Drop duplicated vacancies based on job descriptions and education and experience requirements
- Drop vacancies with occupations not in selected major occupations

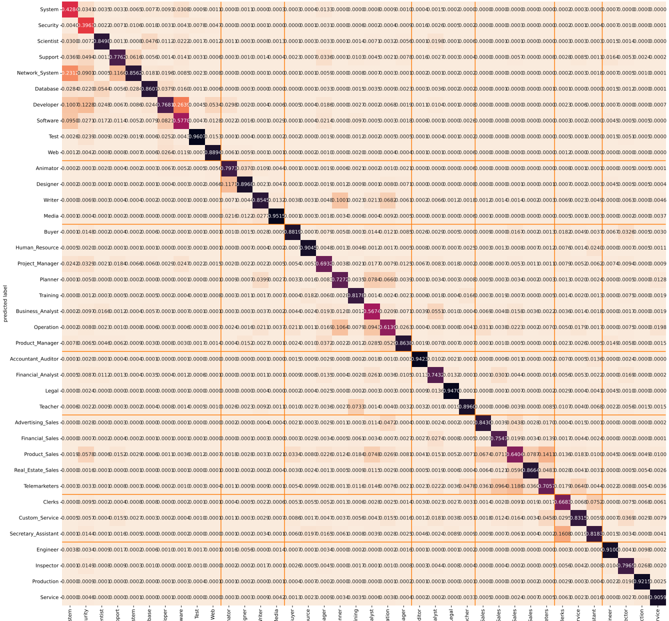
Data: Occupation Classification

◂ Back Data ◂ Back Estimation

- No ready-for-use occupation classification
- Match to a set of selected 6-digit occupations (”minor”) in six 2-digit occupations (”major”) in U.S. SOC 2018
- Key idea: an occupation is defined by a bundle of skills and tasks
- 1st step: for each occupation choose several exclusive keywords, and find the set of just-match vacancies as the ”learning” sample
- 2nd step: use the ”learning” group to train a Naive Bayes classifier based on the job titles and job descriptions
- 3rd step: apply the trained classifier to both the ”unknown” sample and the ”learning” sample ▸ confusion matrix

Confusion Matrix of Occupation Assignment

Back



Data: Summary Statistics

▸ back

	Pooled			Major Occupation			
	-	Computer	Design_ Media	Business_ Operations	Financial_ Legal	Sales	Admin
Vacancy #	3,999,005	1,330,001	561,236	1,162,404	214,661	452,771	277,932
- share	1.00	.33	.14	.29	.05	.11	.07
Avg # Words	108.91	104.26	103.05	115.60	110.69	120.31	95.09
Wage (1k CNY):							
- Mean	13.64	17.38	10.68	14.19	11.95	10.21	6.32
- SD	9.24	9.79	6.31	9.52	9.19	6.53	3.90
Firm:							
- #	86,330	67,369	68,092	78,244	41,285	58,847	59,016
- Avg Posts	46.32	19.74	8.24	14.86	5.20	7.69	4.71
- Median Posts	20.0	9.0	4.0	6.0	2.0	3.0	2.0
Firm Size (share):							
- -15	.03	.03	.05	.02	.02	.03	.03
- 15-50	.18	.17	.25	.16	.15	.19	.20
- 50-150	.23	.21	.26	.22	.22	.23	.26
- 150-500	.21	.21	.21	.22	.23	.20	.23
- 500-2000	.15	.16	.12	.16	.18	.15	.14
- 2000+	.20	.23	.11	.22	.21	.19	.13
Education (share):							
- Vocational College	.33	.24	.38	.29	.27	.51	.52
- Bachelor	.54	.66	.47	.61	.63	.22	.24
- Master/Doctor	.01	.02	.00	.01	.03	.00	.00
- Not Specified	.12	.08	.15	.09	.07	.27	.23
Experience (share):							
- 0	.22	.12	.21	.16	.25	.48	.50
- 1-3	.37	.33	.48	.37	.36	.31	.38
- 3-5	.31	.41	.25	.33	.26	.16	.10
- 5-10	.11	.14	.05	.14	.13	.05	.03

Education, Experience, Occupation $\subset$ {Skills, Tasks}

- One way: $X = \{\text{EDU, EXP, OCC}\}$ ▸ results ▸ compare with $X = \{\text{EDU, EXP}\}$ ▸ bias correction
- All are different subspaces of the full skill/task space
- In theory, an occupation is a subset in the skill/task space
  - A pre-defined bundle of different skills/tasks
  - Lack of within-occupation skill/task variations
- In practice, occupation info of vacancy data is generated by mapping job title or content to the official categories ▸ occupation classifiction
- Below, we directly exploit all information in vacancy texts to create proxy variables for various skills/tasks
  - By doing this, we also show a data-driven skill/task structure

Variance Decomposition

Back

	Pooled		Computer		Design_Media		Admin	
	Comp.	Share	Comp.	Share	Comp.	Share	Comp.	Share
Var(ln w)	.360	-	.279	-	.251	-	.164	-
Panel A: X=$\{$EDU, EXP$\}$								
Var(θ_i)	.102	.283	.052	.188	.053	.212	.050	.307
Within-Firm:								
Var$(\theta_i - \bar{\theta}_j)$	.072	.199	.037	.133	.036	.144	.033	.204
Var(ϵ_i)	.132	.367	.089	.318	.078	.310	.061	.371
Between-Firm:								
Var$(\bar{\theta}_j)$	.030	.084	.015	.055	.017	.068	.017	.102
Var(ψ_j)	.076	.212	.102	.365	.086	.342	.041	.253
2 Cov$(\bar{\theta}_j, \psi_j)$	.049	.137	.036	.130	.034	.136	.011	.069
Panel B: X=$\{$EDU, EXP, OCC$\}$ (Change from Panel A)								
Var(θ_i)	$+$.045	$+$.124	$+$.012	$+$.044	$+$.008	$+$.031	$+$.002	$+$.013
Within-Firm:								
Var$(\theta_i - \bar{\theta}_j)$	$+$.031	$+$.087	$+$.012	$+$.043	$+$.004	$+$.015	$+$.002	$+$.010
Var(ϵ_i)	$-$.031	$-$.087	$-$.012	$-$.043	$-$.004	$-$.015	$-$.002	$-$.010
Between-Firm:								
Var$(\bar{\theta}_j)$	$+$.013	$+$.037	$+$.000	$+$.002	$+$.004	$+$.017	$+$.001	$+$.005
Var(ψ_j)	$-$.012	$-$.033	$-$.006	$-$.021	$-$.007	$-$.028	$-$.001	$-$.008
2 Cov$(\bar{\theta}_j, \psi_j)$	$-$.001	$-$.003	$+$.005	$+$.018	$+$.003	$+$.012	$+$.001	$+$.005
Obs	3998840		1325260		548808		260364	
Firm	86165		62628		55664		41448	

Variance Bias Correction

Back

	Pooled		Computer		Design_Media		Admin	
	Comp.	Share	Comp.	Share	Comp.	Share	Comp.	Share
Var(ln w)	.360	-	.279	-	.251	-	.164	-
Panel A: Plug-In								
Var(θ_i)	.102	.283	.052	.188	.053	.212	.050	.307
Var(ϵ_i)	.132	.367	.089	.318	.078	.310	.061	.371
Var(ψ_j)	.076	.212	.102	.365	.086	.342	.041	.253
2 Cov(θ_j, ψ_j)	.049	.137	.036	.130	.034	.136	.011	.069
Panel B: Homoscedasticity Correction (Change from Panel A)								
Var(θ_i)	−.000	+.000	+.000	+.000	+.000	+.000	−.000	+.000
Var(ϵ_i)	+.003	+.009	+.004	+.016	+.009	+.035	+.011	+.070
Var(ψ_j)	−.003	−.008	−.004	−.016	−.009	−.035	−.011	−.070
2 Cov(θ_j, ψ_j)	+.000	+.000	−.000	+.000	−.000	+.000	+.000	+.000
Panel C: KSS (Leave-Out) Correction (Change from Panel A)								
Var(θ_i)	−.000	+.000	+.000	+.000	−.000	+.000	−.000	+.000
Var(ϵ_i)	+.003	+.007	+.004	+.014	+.007	+.029	+.010	+.060
Var(ψ_j)	−.003	−.007	−.004	−.015	−.007	−.028	−.010	−.060
2 Cov(θ_j, ψ_j)	+.000	+.001	−.000	+.000	+.000	+.000	−.000	+.000
Obs	3998840		1325260		548808		260364	
Firm	86165		62628		55664		41448	

Feature Selection: Lasso Regression ▸ Overview

1st step: extract the useful information in vacancy text

- First we transform the vacancy text into an indicator matrix $\mathbf{C}$ with dimension $N \times K$ where each entry c_{ik} is an indicator of a token (word/phrase) k in vacancy i and the total vocabulary set is V

- Then we use (regularized linear) Lasso regression (L1 penalization):
  $\hat{\zeta} = \arg\min_{\zeta} \sum_{i=1}^{N} \left(\ln w_i - \sum_{k=1}^{K} c_{ik}\zeta_k \right)^2 + \lambda \sum_{k=1}^{K} |\zeta_k|$

Feature Selection: Tune Lasso ▸ Overview

- Following the suggestion in the literature, we use BIC as the criterion to gauge the hyperparameter λ: $\min \mathrm{BIC}(\lambda) = \frac{\| \ln \mathbf{w} - \mathbf{C}\hat{\zeta}_\lambda \|^2}{\sigma^2} + \widehat{df}_\lambda \log N$
- The estimation results 700-3100 features (V') with nonzero coefficients

	Pooled	Computer	Design_ Media	Admin
λ^*	332.0	190.3	238.5	155.0
MSE	.162	.149	.142	.100
R^2	.566	.494	.461	.418
BIC/N	.446	.527	.561	.613
df	3,144	1,922	929	691
K	109,123	51,602	39,306	24,896
N	3,999,005	1,330,001	561,236	277,932

Feature Selection: Inference and Interpretation on Lasso Results

▸ Overview

- In general, features selected and their coefficients in high-dimensional penalized model are not interpretable due to multicollinearity and flexibility
- Inference via subsampling (10x10) shows that our selected features/tokens are rather robust (small confidence interval) ▸ subsampling results
- Interpretation on coefficients are still forbidden, but now we can inspect important features to see if they make some intuitive sense ▸ top positive tokens ▸ top negative tokens

Feature Clustering: Word Embedding ▸ Overview

2nd step: examine what are these selected features (beyond eyeballing)

- Indicator matrix $\mathbf{C}$ tells nothing about the meaning of the words
- We train a word embedding model, Word2Vec (CBOW), to learn the relationship between tokens
  - it maps each word to a latent vector space (with dimension $H = 100$), which best predicts the probability of a word given the context (adjacent words)
- The result is a $K \times H$ embedding weight matrix $\mathbf{U}$, where each row of the matrix, $\mathbf{u}_k$, is the representation vector of the word k in the latent embedding space
- We only use the part of the selected features: $\mathbf{U}' \equiv \{\mathbf{u}_k\}$ where $k \in V'$

Feature Clustering: K-Means Clustering ▸ Overview

- We now can use unsupervised clustering algorithms to cluster our selected features
- We use K-Means classifier, which finds the centroids for the clusters $\{V'_p\}$ in the embedding space to minimize the sum of within-cluster Euclidean distances:
  $$\underset{\{V'_1, V'_2, \ldots, V'_P\}}{\arg\min} \sum_{p=1}^{P} \sum_{k \in V'_p} \left\| \mathbf{u}_k - \frac{1}{|V'_p|} \sum_{j \in V'_p} \mathbf{u}_j \right\|^2$$
- P is the predetermined cluster numbers, and we set $P = 8$ *(arbitrary)*
- Visualization of clustering results in 2D (through t-SNE only for demonstration):
  ▸ Pooled ▸ Computer ▸ Design & Media ▸ Admin

Dimension Reduction ‣ Overview

3rd step: further reduce the dimension of these features

- Instead of PCA (unsupervised), we use partial least squares (PLS) (supervised) regression which uses the covariance of the predictive and target variables

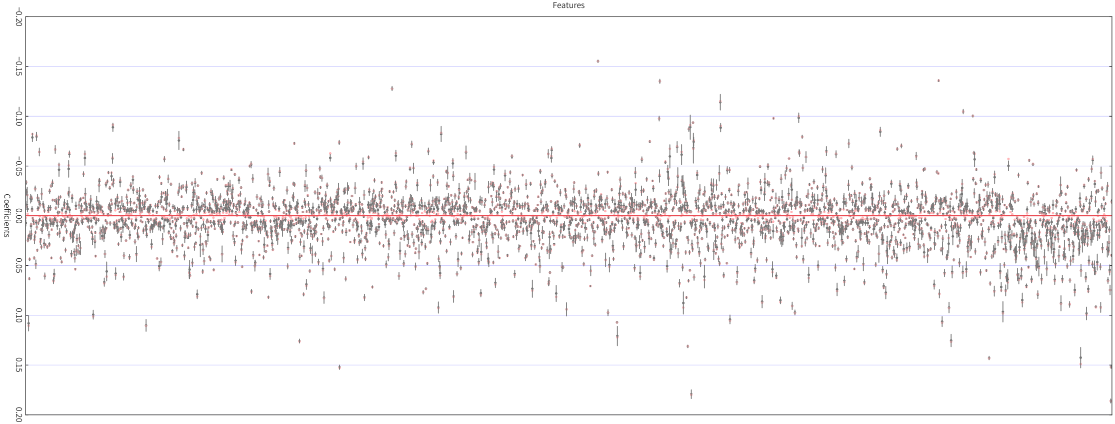
- Transform the indicator matrix $\mathbf{C'}_p \equiv \{\mathbf{c}_k\}, k \in V'_p$ of each cluster p into a low dimensional representation Ξ_p; Set reduced dimension $Q = 3$ *(arbitrary)*

- Thus for each occupation, we now have 8 proxy matrices (linear combination) $\Xi_1, \Xi_2, \ldots, \Xi_8$ corresponding to 8 clusters $V'_1, V'_2, \ldots, V'_8$

- OLS regressions show that they preserve over 95% predictive power (R^2) of the Lasso regression

Confidence Intervals on Lasso Coefficients via Subsampling

◂ Back



Feature Selection: Top Features (Positive)

Back

	Pooled			Computer			Design_Media			Admin		
	token	coef	feq	token	coef	feq	token	coef	feq	token	coef	feq
1	14th month pay	.152	.014	15th month pay	.181	.010	14th month pay	.193	.011	undergraduate	.161	.014
2	three meals	.143	.014	three meals	.148	.014	lead	.155	.025	undergraduate	.157	.156
3	large platform	.131	.019	14th month pay	.140	.017	three meals	.129	.015	president	.120	.014
4	master degree	.126	.015	master degree	.109	.027	c++	.121	.017	ceo	.117	.010
5	lead	.107	.041	lead	.089	.038	crisis	.113	.011	build	.117	.016
6	c++	.092	.051	golang	.080	.017	games	.098	.180	lead	.105	.017
7	algorithm	.082	.061	guru	.079	.047	europe & america	.090	.011	government	.103	.030
8	guru	.082	.028	deep learning	.078	.022	engine	.090	.046	high salary	.089	.018
9	famous	.079	.019	famous	.070	.014	4a	.090	.014	translation	.083	.012
10	machine learning	.077	.016	high salary	.070	.018	six insurance & one fund	.086	.046	bachelor degree	.082	.018
11	formation	.076	.013	maestro	.068	.012	finance	.084	.016	strategy	.077	.015
12	undergraduate	.074	.319	overseas	.067	.010	undergraduate	.078	.238	large scale	.076	.030
13	overseas	.072	.026	go	.065	.027	listed company	.076	.021	landing	.070	.018
14	react	.072	.020	c++	.064	.144	finance	.076	.031	project management	.067	.011
15	development	.071	.374	algorithm	.064	.164	outsourcing	.074	.012	overseas	.066	.021
16	undergraduate	.066	.029	react	.064	.061	guru	.070	.022	background	.064	.032
17	high salary	.063	.028	machine learning	.061	.045	overseas	.068	.024	develop	.063	.097
18	landing	.060	.067	landing	.061	.037	journalists	.068	.011	13th month pay	.063	.019
19	strategy	.057	.047	development	.059	.776	13th month pay	.068	.023	unified recruitment	.058	.031
20	live streaming	.056	.014	audio & video	.058	.012	c4d	.066	.021	budget	.057	.021
21	listed company	.055	.027	unified recruitment	.054	.044	famous	.065	.023	major	.055	.019
22	large scale	.055	.072	beijing	.053	.012	unity	.065	.043	decoration	.055	.016
23	responsibilities	.055	.048	live streaming	.052	.011	high salary	.064	.016	resources	.053	.043
24	shuttle	.054	.018	recommend	.052	.023	management	.063	.010	promote	.051	.029
25	finance	.054	.070	management	.051	.016	3d	.063	.106	finance	.051	.036
26	six insurance & one fund	.053	.055	ai	.051	.015	large scale	.063	.043	english	.050	.054
27	python	.052	.066	stock	.049	.025	performance	.063	.016	business negotiations	.048	.010
28	director	.052	.022	undergraduate	.048	.365	unified recruitment	.059	.019	optimization	.046	.079
29	unified recruitment	.051	.042	salary	.048	.049	undergraduate	.059	.023	responsibilities	.046	.035
30	hive	.051	.013	supplementary	.045	.019	ip	.057	.017	integrated planning	.046	.028

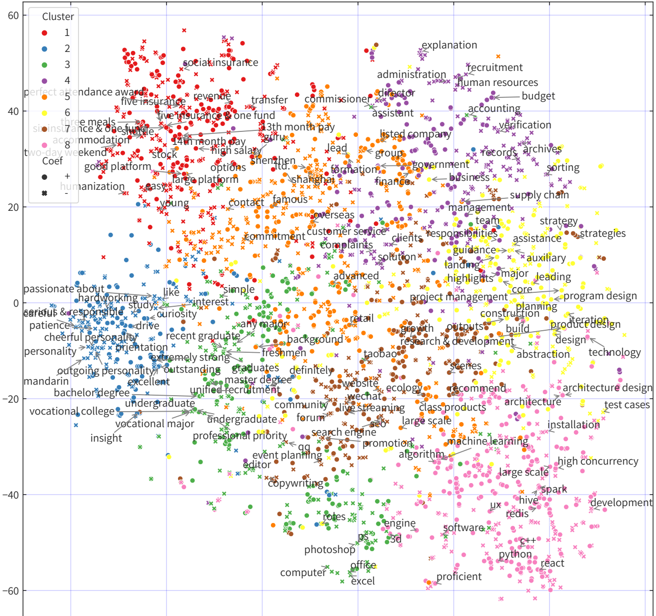
Feature Selection: Top Features (Negative)

Back

	Pooled			Computer			Design_Media			Admin		
	token	coeff	feq	token	coeff	feq	token	coeff	feq	token	coeff	feq
1	freshmen	-.155	.018	graduates	-.205	.013	freshmen	-.188	.017	five insurance	-.070	.052
2	five insurance	-.136	.030	five insurance	-.197	.016	internship	-.133	.011	graduates	-.061	.082
3	graduates	-.128	.033	vocational college	-.134	.072	five insurance	-.132	.033	vocational school	-.059	.038
4	vocational major	-.100	.036	social insurance	-.121	.012	graduates	-.132	.030	freshmen	-.057	.048
5	two-day weekend	-.098	.166	vocational major	-.119	.030	two-day weekend	-.090	.176	internship	-.056	.012
6	vocational college	-.094	.148	two-day weekend	-.115	.147	recent graduate	-.072	.026	interns	-.053	.017
7	assistant	-.079	.011	recent graduate	-.106	.011	vocational college	-.070	.144	two-day weekend	-.051	.214
8	customer service	-.075	.030	test cases	-.067	.068	social insurance	-.068	.023	player	-.046	.024
9	social insurance	-.073	.028	installation	-.067	.048	vocational major	-.066	.041	mandarin	-.046	.172
10	accounting	-.071	.019	th	-.066	.014	ltd.	-.059	.012	women	-.038	.015
11	accommodation	-.067	.016	computer	-.065	.011	any major	-.055	.011	social insurance	-.037	.060
12	administration	-.067	.027	after sales	-.061	.011	humanization	-.055	.019	qq	-.037	.036
13	commissioner	-.063	.011	young	-.060	.013	comics	-.053	.014	easy	-.035	.043
14	taobao	-.059	.015	five insurance & one fund	-.059	.273	cad	-.052	.010	website	-.033	.032
15	assistance	-.058	.164	business trip	-.051	.030	photoshop	-.049	.235	cleaning	-.030	.015
16	ps	-.056	.029	records	-.048	.015	cdr	-.047	.012	health	-.029	.024
17	ltd.	-.056	.012	hardworking	-.048	.015	website	-.047	.180	clerks	-.029	.014
18	installation	-.055	.020	holidays	-.046	.059	assistance	-.046	.131	attendance	-.029	.104
19	photoshop	-.052	.039	clients	-.046	.078	ps	-.045	.142	e-commerce	-.029	.031
20	careful	-.050	.032	easy	-.043	.017	hardworking	-.044	.023	input	-.028	.044
21	hardworking	-.050	.032	software testing	-.043	.047	anime	-.044	.019	shift	-.028	.013
22	verification	-.048	.011	wechat	-.041	.042	easy	-.044	.033	answer the phone	-.027	.101
23	human resources	-.047	.032	.net	-.041	.034	contact	-.042	.011	administration	-.027	.256
24	website	-.047	.090	patience	-.040	.023	editor	-.039	.204	perfect attendance award	-.026	.032
25	any major	-.047	.020	website	-.039	.101	artwork	-.038	.032	apply for the job	-.025	.018
26	humanization	-.046	.012	focused	-.038	.011	forum	-.038	.034	mobile	-.025	.013
27	excel	-.046	.047	network equipment	-.037	.016	taobao	-.038	.024	hardworking	-.025	.055
28	mandarin	-.045	.027	bug	-.036	.053	young	-.038	.034	join	-.024	.041
29	explanation	-.044	.013	works	-.035	.023	commission	-.037	.017	games	-.024	.039
30	young	-.044	.025	holiday	-.034	.037	clients	-.037	.096	front desk	-.023	.088
31	contact	-.044	.010	dividend	-.034	.012	wechat	-.037	.172	department manager	-.023	.014

Feature Clustering: Visualization (Pooled)

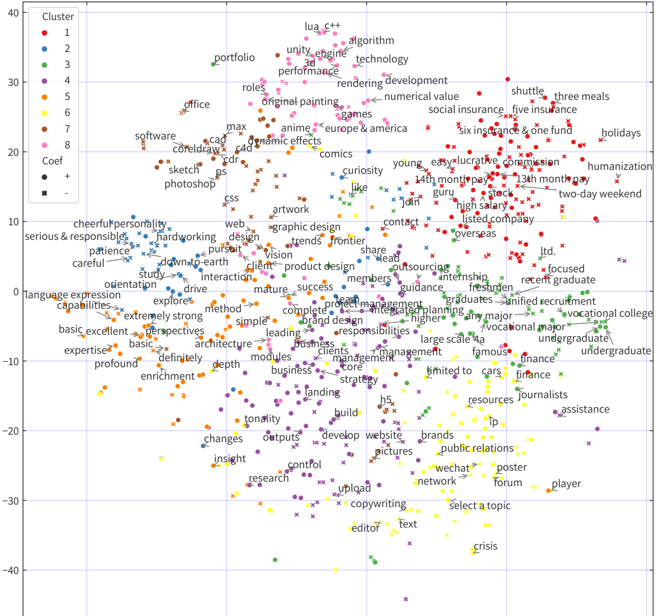
Back



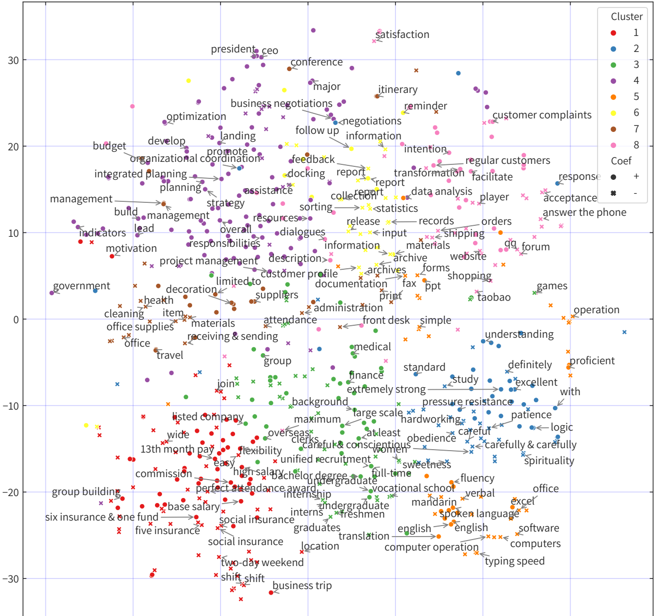
Feature Clustering: Visualization (Computer) ◂ Back

Feature Clustering: Visualization (Design_Media)

Back

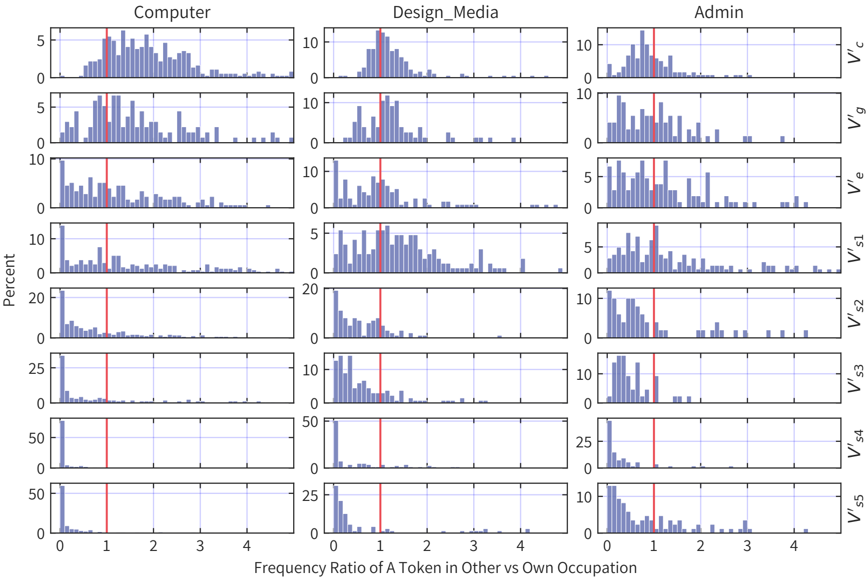


Feature Clustering: Visualization (Admin) ◂ Back

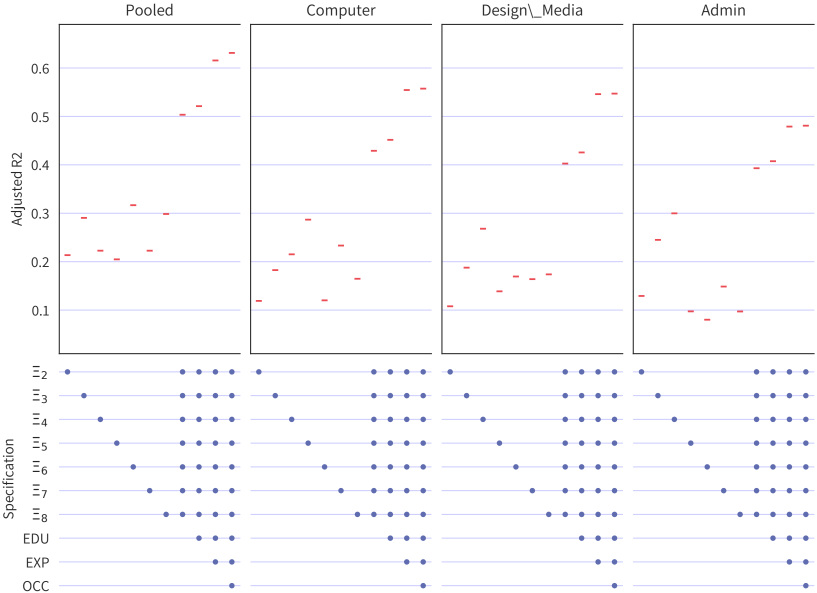


Feature Clustering: General vs Specific

◂ Back



R2 Under Different Specifications ◂ Back



Variance Bias Correction

Back

	Pooled		Computer		Design_Media		Admin	
	Comp.	Share	Comp.	Share	Comp.	Share	Comp.	Share
Var(ln w)	.362	-	.281	-	.253	-	.164	-
Panel A: Plug-In								
Var(θ_i)	.163	.450	.082	.291	.084	.331	.067	.408
Var(ϵ_i)	.096	.267	.071	.252	.065	.255	.050	.304
Var(ψ_j)	.051	.141	.074	.263	.062	.243	.035	.216
2 Cov(θ_i, ψ_j)	.051	.142	.054	.193	.043	.171	.012	.072
Panel B: Homoscedasticity Correction (Change from Panel A)								
Var(θ_i)	$+$.000	$+$.000	$-$.000	$+$.000	$-$.000	$+$.000	$+$.000	$+$.001
Var(ϵ_i)	$+$.002	$+$.006	$+$.004	$+$.012	$+$.007	$+$.029	$+$.009	$+$.057
Var(ψ_j)	$-$.002	$-$.006	$-$.004	$-$.012	$-$.007	$-$.029	$-$.009	$-$.057
2 Cov(θ_i, ψ_j)	$-$.000	$+$.000	$+$.000	$+$.001	$-$.000	$+$.000	$-$.000	$-$.002
Panel C: KSS (Leave-Out) Correction (Change from Panel A)								
Var(θ_i)	$-$.000	$+$.000	$+$.000	$+$.000	$+$.000	$+$.000	$-$.000	$-$.001
Var(ϵ_i)	$+$.002	$+$.005	$+$.003	$+$.012	$+$.006	$+$.024	$+$.008	$+$.048
Var(ψ_j)	$-$.002	$-$.005	$-$.003	$-$.012	$-$.006	$-$.024	$-$.008	$-$.048
2 Cov(θ_i, ψ_j)	$+$.000	$+$.000	$+$.000	$+$.001	$+$.000	$+$.002	$+$.000	$+$.001
Obs	3998840		1325260		548808		260364	
Firm	86165		62628		55664		41448	

Conditional On EXP=0

Back

	Pooled		Computer		Design_Media		Admin	
	Comp.	Share	Comp.	Share	Comp.	Share	Comp.	Share
$\text{Var}(\ln w)$	.305	-	.407	-	.226	-	.097	-
Panel A: $X = \{$**EDU**, **EXP**, $\Xi_2, \ldots, \Xi_8\}$								
$\text{Var}(\theta_i)$	.079	.258	.069	.169	.036	.159	.014	.146
$\text{Var}(\epsilon_i)$	.115	.377	.111	.273	.084	.372	.049	.512
$\text{Var}(\psi_j)$	.068	.222	.138	.339	.075	.333	.029	.298
$2\,\text{Cov}(\theta_i, \psi_j)$	.044	.143	.089	.219	.033	.145	.005	.047
Panel B: Decompose θ **Terms**								
$\text{Var}(X_{int})$	.000	.000	.000	.000	.000	.000	.000	.000
$\text{Var}(X_{ext})$	.079	.258	.069	.169	.036	.159	.014	.146
$2\,\text{Cov}(X_{int}, X_{ext})$	.000	.000	.000	.000	.000	.000	.000	.000
$2\,\text{Cov}(X_{int}, \psi_j)$	.000	.000	.000	.000	.000	.000	.000	.000
$2\,\text{Cov}(X_{ext}, \psi_j)$	.044	.143	.089	.219	.033	.145	.005	.047
Panel C: Further Decompose X_{ext} **Terms**								
$\text{Var}(\Xi_g)$	.001	.004	.001	.003	.001	.005	.000	.002
$\text{Var}(\Xi_m)$	.005	.018	.010	.024	.004	.016	.003	.031
$\text{Var}(\Xi_s)$	.047	.153	.036	.087	.021	.094	.007	.068
$2\,\text{Cov}(\Xi_g, \Xi_m)$	.001	.004	.001	.004	.001	.002	.000	.004
$2\,\text{Cov}(\Xi_g, \Xi_s)$	.006	.021	.003	.008	.003	.012	.001	.009
$2\,\text{Cov}(\Xi_m, \Xi_s)$	.018	.058	.017	.043	.007	.032	.003	.032
$2\,\text{Cov}(\Xi_g, X_{int})$	.000	.000	.000	.000	.000	.000	.000	.000
$2\,\text{Cov}(\Xi_m, X_{int})$	.000	.000	.000	.000	.000	.000	.000	.000
$2\,\text{Cov}(\Xi_s, X_{int})$	.000	.000	.000	.000	.000	.000	.000	.000
$2\,\text{Cov}(\Xi_g, \psi_j)$	.003	.010	.005	.013	.002	.008	.000	.002
$2\,\text{Cov}(\Xi_m, \psi_j)$	.008	.027	.024	.060	.006	.029	.002	.022
$2\,\text{Cov}(\Xi_s, \psi_j)$	.032	.106	.059	.146	.024	.108	.002	.023
Obs	858147		144122		104960		120241	

Conditional On EXP=1-3

Back

	Pooled		Computer		Design_Media		Admin	
	Comp.	Share	Comp.	Share	Comp.	Share	Comp.	Share
$\text{Var}(\ln w)$	.204	-	.195	-	.140	-	.104	-
Panel A: $X = \{\textbf{EDU}, \textbf{EXP}, \Xi_2, \ldots, \Xi_8\}$								
$\text{Var}(\theta_i)$	.062	.302	.034	.174	.022	.158	.027	.259
$\text{Var}(\epsilon_i)$	.081	.396	.064	.331	.057	.407	.049	.468
$\text{Var}(\psi_j)$	.043	.213	.068	.348	.048	.343	.024	.235
$2\,\text{Cov}(\theta_i, \psi_j)$	.018	.088	.029	.147	.013	.095	.004	.036
Panel B: Decompose θ Terms								
$\text{Var}(X_{int})$	.000	.000	.000	.000	.000	.000	.000	.000
$\text{Var}(X_{ext})$	.062	.302	.034	.174	.022	.158	.027	.259
$2\,\text{Cov}(X_{int}, X_{ext})$	.000	.000	.000	.000	.000	.000	.000	.000
$2\,\text{Cov}(X_{int}, \psi_j)$	.000	.000	.000	.000	.000	.000	.000	.000
$2\,\text{Cov}(X_{ext}, \psi_j)$	.018	.088	.029	.147	.013	.095	.004	.036
Panel C: Further Decompose X_{ext} Terms								
$\text{Var}(\Xi_g)$	.001	.003	.000	.002	.000	.002	.000	.001
$\text{Var}(\Xi_m)$	.005	.024	.004	.020	.002	.013	.005	.051
$\text{Var}(\Xi_s)$	.036	.177	.021	.106	.016	.116	.013	.126
$2\,\text{Cov}(\Xi_g, \Xi_m)$	.001	.006	.000	.002	.000	.001	.000	.005
$2\,\text{Cov}(\Xi_g, \Xi_s)$	.005	.023	.002	.009	.001	.006	.001	.012
$2\,\text{Cov}(\Xi_m, \Xi_s)$	.014	.068	.007	.036	.003	.020	.007	.066
$2\,\text{Cov}(\Xi_g, X_{int})$	.000	.000	.000	.000	.000	.000	.000	.000
$2\,\text{Cov}(\Xi_m, X_{int})$	.000	.000	.000	.000	.000	.000	.000	.000
$2\,\text{Cov}(\Xi_s, X_{int})$	.000	.000	.000	.000	.000	.000	.000	.000
$2\,\text{Cov}(\Xi_g, \psi_j)$	.001	.005	.001	.007	.000	.003	.000	.000
$2\,\text{Cov}(\Xi_m, \psi_j)$	.006	.031	.009	.046	.005	.034	.003	.031
$2\,\text{Cov}(\Xi_s, \psi_j)$	.011	.052	.018	.094	.008	.058	.001	.005
Obs	1457630		432077		254456		88030	

Conditional On EXP=3-5

◂ Back

	Pooled		Computer		Design_Media		Admin	
	Comp.	Share	Comp.	Share	Comp.	Share	Comp.	Share
Var(ln w)	.202	-	.167	-	.162	-	.192	-
Panel A: $X = \{$**EDU**, **EXP**, $\Xi_2, \ldots, \Xi_8\}$								
Var(θ_i)	.043	.212	.020	.121	.021	.129	.047	.246
Var(ϵ_i)	.079	.390	.055	.332	.060	.368	.085	.442
Var(ψ_j)	.054	.266	.065	.392	.061	.374	.049	.254
2 Cov(θ_i, ψ_j)	.027	.132	.026	.156	.021	.129	.013	.067
Panel B: Decompose θ Terms								
Var(X_{int})	.000	.000	.000	.000	.000	.000	.000	.000
Var(X_{ext})	.043	.212	.020	.121	.021	.129	.047	.246
2 Cov(X_{int}, X_{ext})	.000	.000	.000	.000	.000	.000	.000	.000
2 Cov(X_{int}, ψ_j)	.000	.000	.000	.000	.000	.000	.000	.000
2 Cov(X_{ext}, ψ_j)	.027	.132	.026	.156	.021	.129	.013	.067
Panel C: Further Decompose X_{ext} Terms								
Var(Ξ_g)	.000	.002	.000	.000	.000	.000	.001	.004
Var(Ξ_m)	.004	.019	.002	.013	.001	.008	.010	.054
Var(Ξ_s)	.026	.129	.013	.080	.016	.096	.024	.125
2 Cov(Ξ_g, Ξ_m)	.001	.004	.000	.001	.000	.001	.001	.005
2 Cov(Ξ_g, Ξ_s)	.003	.015	.001	.005	.001	.009	.002	.009
2 Cov(Ξ_m, Ξ_s)	.009	.044	.004	.023	.002	.014	.011	.056
2 Cov(Ξ_g, X_{int})	.000	.000	.000	.000	.000	.000	.000	.000
2 Cov(Ξ_m, X_{int})	.000	.000	.000	.000	.000	.000	.000	.000
2 Cov(Ξ_s, X_{int})	.000	.000	.000	.000	.000	.000	.000	.000
2 Cov(Ξ_g, ψ_j)	.001	.007	.001	.006	.001	.007	.000	.000
2 Cov(Ξ_m, ψ_j)	.007	.035	.007	.041	.005	.030	.007	.038
2 Cov(Ξ_s, ψ_j)	.018	.090	.018	.109	.015	.092	.006	.029
Obs	1222973		533940		127417		17247	

Conditional On EDU=C

Back

	Pooled		Computer		Design_Media		Admin	
	Comp.	Share	Comp.	Share	Comp.	Share	Comp.	Share
Var(ln w)	.244	-	.211	-	.200	-	.106	-
Panel A: $X = \{$**EDU**, **EXP**, $\Xi_2, \ldots, \Xi_8\}$								
Var(θ_i)	.111	.454	.072	.342	.066	.330	.033	.307
Var(ϵ_i)	.085	.349	.064	.303	.059	.293	.046	.428
Var(ψ_j)	.038	.154	.052	.245	.047	.234	.024	.229
2 Cov(θ_i, ψ_j)	.011	.044	.023	.109	.028	.142	.003	.028
Panel B: Decompose θ Terms								
Var(X_{int})	.033	.135	.028	.134	.024	.119	.010	.095
Var(X_{ext})	.046	.188	.026	.122	.024	.121	.013	.122
2 Cov(X_{int}, X_{ext})	.032	.130	.018	.085	.018	.090	.010	.091
2 Cov(X_{int}, ψ_j)	.005	.021	.014	.065	.012	.062	.002	.015
2 Cov(X_{ext}, ψ_j)	.005	.022	.009	.044	.016	.080	.001	.013
Panel C: Further Decompose X_{ext} Terms								
Var(Ξ_g)	.001	.004	.000	.002	.000	.001	.000	.003
Var(Ξ_m)	.002	.010	.001	.005	.001	.005	.001	.008
Var(Ξ_s)	.028	.114	.019	.092	.018	.090	.009	.084
2 Cov(Ξ_g, Ξ_m)	.001	.004	.000	.001	.000	.001	.000	.001
2 Cov(Ξ_g, Ξ_s)	.005	.019	.002	.009	.002	.008	.001	.007
2 Cov(Ξ_m, Ξ_s)	.009	.037	.003	.013	.003	.017	.002	.020
2 Cov(Ξ_g, X_{int})	.003	.012	.001	.006	.001	.005	.001	.005
2 Cov(Ξ_m, X_{int})	.005	.022	.002	.011	.003	.013	.002	.014
2 Cov(Ξ_s, X_{int})	.023	.096	.014	.068	.014	.072	.008	.072
2 Cov(Ξ_g, ψ_j)	.001	.003	.001	.004	.001	.003	-.000	.003
2 Cov(Ξ_m, ψ_j)	.001	.005	.002	.010	.002	.011	.001	.008
2 Cov(Ξ_s, ψ_j)	.004	.015	.007	.031	.013	.066	.001	.008
Obs	1302141		308332		198391		127547	

Conditional On EDU=B

Back

	Pooled		Computer		Design_Media		Admin	
	Comp.	Share	Comp.	Share	Comp.	Share	Comp.	Share
Var(ln w)	.313	-	.244	-	.244	-	.223	-
Panel A: $X = \{$**EDU**, **EXP**, $\Xi_2, \ldots, \Xi_8\}$								
Var(θ_i)	.129	.411	.063	.259	.085	.349	.101	.455
Var(ϵ_i)	.094	.299	.070	.287	.071	.291	.073	.326
Var(ψ_j)	.052	.166	.070	.286	.054	.220	.037	.166
2 Cov(θ_i, ψ_j)	.039	.124	.041	.167	.035	.142	.010	.045
Panel B: Decompose θ Terms								
Var(X_{int})	.043	.138	.027	.113	.036	.145	.036	.160
Var(X_{ext})	.052	.165	.022	.091	.026	.108	.036	.163
2 Cov(X_{int}, X_{ext})	.034	.108	.014	.056	.023	.095	.030	.133
2 Cov(X_{int}, ψ_j)	.014	.044	.013	.054	.016	.067	.008	.036
2 Cov(X_{ext}, ψ_j)	.025	.081	.028	.113	.018	.075	.002	.009
Panel C: Further Decompose X_{ext} Terms								
Var(Ξ_g)	.001	.003	.000	.001	.000	.001	.001	.004
Var(Ξ_m)	.002	.006	.001	.004	.001	.004	.002	.009
Var(Ξ_s)	.034	.110	.017	.069	.020	.080	.025	.112
2 Cov(Ξ_g, Ξ_m)	.001	.003	.000	.001	.000	.001	.000	.001
2 Cov(Ξ_g, Ξ_s)	.005	.016	.001	.005	.002	.007	.003	.012
2 Cov(Ξ_m, Ξ_s)	.009	.027	.003	.011	.003	.014	.005	.023
2 Cov(Ξ_g, X_{int})	.003	.009	.001	.003	.001	.006	.002	.008
2 Cov(Ξ_m, X_{int})	.005	.015	.002	.007	.003	.013	.005	.022
2 Cov(Ξ_s, X_{int})	.026	.084	.011	.045	.019	.077	.023	.103
2 Cov(Ξ_g, ψ_j)	.002	.006	.001	.005	.001	.005	-.001	.005
2 Cov(Ξ_m, ψ_j)	.003	.010	.004	.015	.003	.011	.003	.013
2 Cov(Ξ_s, ψ_j)	.020	.064	.023	.093	.014	.058	.000	.002
Obs	2142593		863523		248143		55786	

If $\Xi_m \equiv \{\text{EDU}, \Xi_3, \Xi_4\}$

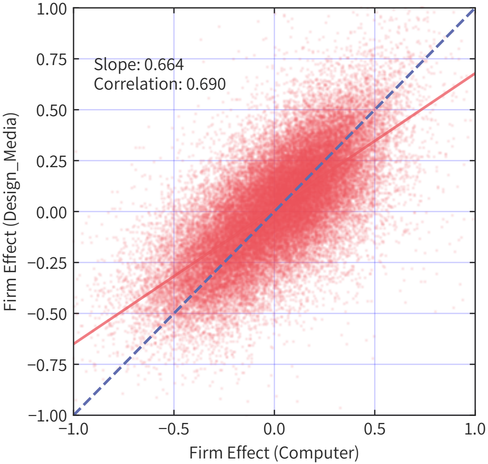
Back

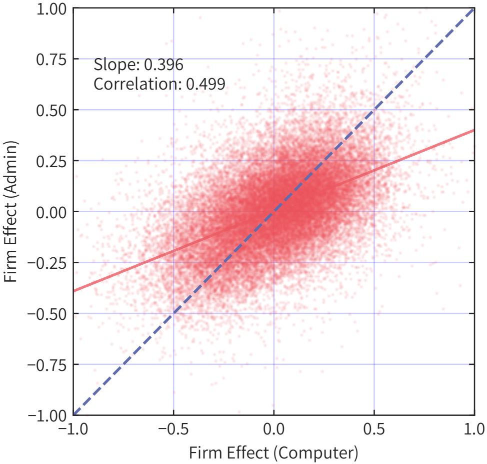
	Pooled		Computer		Design_Media		Admin	
	Comp.	Share	Comp.	Share	Comp.	Share	Comp.	Share
$\text{Var}(\ln w)$	.362	-	.281	-	.253	-	.164	-
Panel A: $X = \{$**EDU**, **EXP**$, \Xi_2, \ldots, \Xi_8\}$								
$\text{Var}(\theta_i)$	.163	.450	.082	.291	.084	.330	.067	.409
$\text{Var}(\epsilon_i)$	.098	.272	.074	.264	.071	.279	.058	.353
$\text{Var}(\psi_j)$	.049	.136	.071	.251	.056	.219	.027	.168
$2\,\text{Cov}(\theta_i, \psi_j)$	.052	.142	.054	.193	.043	.170	.012	.072
Panel B: Decompose θ Terms								
$\text{Var}(X_{int})$	.042	.115	.028	.099	.030	.119	.016	.096
$\text{Var}(X_{ext})$	.072	.199	.035	.126	.030	.117	.030	.184
$2\,\text{Cov}(X_{int}, X_{ext})$	.049	.136	.019	.067	.024	.094	.021	.129
$2\,\text{Cov}(X_{int}, \psi_j)$	.017	.048	.017	.060	.018	.072	.004	.025
$2\,\text{Cov}(X_{ext}, \psi_j)$	.034	.094	.037	.133	.025	.099	.008	.047
Panel C: Further Decompose X_{ext} Terms								
$\text{Var}(\Xi_g)$	.001	.003	.000	.001	.000	.001	.000	.002
$\text{Var}(\Xi_m)$	.017	.048	.007	.026	.006	.025	.018	.109
$\text{Var}(\Xi_s)$	.022	.062	.014	.051	.011	.045	.003	.019
$2\,\text{Cov}(\Xi_g, \Xi_m)$	.004	.010	.001	.003	.001	.004	.002	.011
$2\,\text{Cov}(\Xi_g, \Xi_s)$	.005	.012	.001	.005	.001	.004	.001	.003
$2\,\text{Cov}(\Xi_m, \Xi_s)$	.023	.064	.011	.039	.009	.037	.007	.041
$2\,\text{Cov}(\Xi_g, X_{int})$	.004	.011	.001	.004	.001	.005	.001	.006
$2\,\text{Cov}(\Xi_m, X_{int})$	.020	.054	.006	.022	.011	.042	.017	.102
$2\,\text{Cov}(\Xi_s, X_{int})$	.026	.071	.011	.041	.012	.047	.003	.020
$2\,\text{Cov}(\Xi_g, \psi_j)$	.002	.007	.002	.007	.001	.005	.000	.001
$2\,\text{Cov}(\Xi_m, \psi_j)$	.014	.040	.015	.052	.012	.048	.007	.040
$2\,\text{Cov}(\Xi_s, \psi_j)$	.017	.048	.021	.075	.012	.046	.001	.007
Obs	3998840		1325260		548808		260364	

If $\Xi_m \equiv \{\text{EDU}, \Xi_3, \Xi_4, \Xi_5\}$ ◂ Back

	Pooled		Computer		Design_Media		Admin	
	Comp.	Share	Comp.	Share	Comp.	Share	Comp.	Share
Var(ln w)	.362	-	.281	-	.253	-	.164	-
Panel A: $X = \{$**EDU**, **EXP**, $\Xi_2, \ldots, \Xi_8\}$								
Var(θ_i)	.163	.450	.082	.291	.084	.331	.066	.405
Var(ϵ_i)	.098	.272	.074	.264	.071	.279	.058	.352
Var(ψ_j)	.049	.136	.071	.251	.056	.219	.027	.168
2 Cov(θ_i, ψ_j)	.051	.142	.054	.194	.043	.171	.012	.070
Panel B: Decompose θ Terms								
Var(X_{int})	.042	.115	.028	.099	.030	.119	.016	.096
Var(X_{ext})	.072	.199	.035	.125	.030	.118	.029	.180
2 Cov(X_{int}, X_{ext})	.049	.136	.019	.067	.024	.094	.021	.129
2 Cov(X_{int}, ψ_j)	.017	.048	.017	.060	.018	.072	.004	.025
2 Cov(X_{ext}, ψ_j)	.034	.094	.038	.134	.025	.099	.007	.046
Panel C: Further Decompose X_{ext} Terms								
Var(Ξ_g)	.001	.002	.000	.001	.000	.001	.000	.001
Var(Ξ_m)	.021	.057	.015	.055	.008	.033	.020	.122
Var(Ξ_s)	.018	.051	.007	.024	.010	.038	.002	.011
2 Cov(Ξ_g, Ξ_m)	.004	.011	.002	.005	.001	.005	.002	.012
2 Cov(Ξ_g, Ξ_s)	.004	.011	.001	.003	.001	.004	.000	.002
2 Cov(Ξ_m, Ξ_s)	.024	.066	.010	.037	.010	.038	.005	.032
2 Cov(Ξ_g, X_{int})	.004	.011	.001	.004	.001	.005	.001	.006
2 Cov(Ξ_m, X_{int})	.022	.062	.012	.041	.013	.050	.018	.109
2 Cov(Ξ_s, X_{int})	.023	.063	.006	.022	.010	.039	.002	.014
2 Cov(Ξ_g, ψ_j)	.002	.007	.002	.007	.001	.005	.000	.001
2 Cov(Ξ_m, ψ_j)	.017	.047	.025	.089	.014	.053	.007	.041
2 Cov(Ξ_s, ψ_j)	.015	.041	.011	.038	.010	.041	.001	.003
Obs	3998840		1325260		548808		260364	

Firm Wage Premium: Difference Between Occupations ▸ robustness ◂ Back





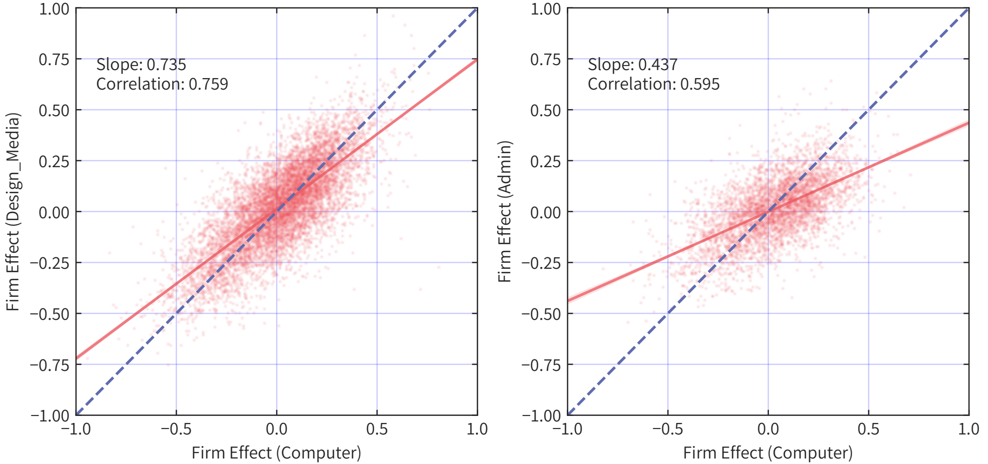
Firm Wage Premium: Firm Size and Firm Location

robustness Back

	Pooled			Computer			Design_Media			Admin		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
fsize.15-50	.019**	.018**	.023**	.011+	.013*	.019**	.022**	.013**	.020**	.006	.005	.005
	(.004)	(.003)	(.003)	(.006)	(.005)	(.004)	(.005)	(.005)	(.004)	(.006)	(.006)	(.006)
fsize.50-150	.042**	.037**	.050**	.037**	.032**	.038**	.050**	.033**	.045**	.020**	.018**	.021**
	(.004)	(.003)	(.003)	(.006)	(.005)	(.004)	(.005)	(.005)	(.004)	(.006)	(.006)	(.005)
fsize.150-500	.067**	.057**	.067**	.072**	.054**	.051**	.086**	.058**	.063**	.035**	.031**	.030**
	(.004)	(.004)	(.003)	(.006)	(.005)	(.005)	(.005)	(.005)	(.004)	(.006)	(.006)	(.006)
fsize.500-2000	.095**	.078**	.085**	.108**	.074**	.066**	.127**	.087**	.086**	.050**	.043**	.040**
	(.005)	(.004)	(.004)	(.007)	(.006)	(.005)	(.006)	(.006)	(.005)	(.007)	(.007)	(.006)
fsize.2000+	.121**	.102**	.120**	.140**	.084**	.082**	.161**	.107**	.108**	.064**	.055**	.058**
	(.005)	(.005)	(.004)	(.008)	(.007)	(.006)	(.007)	(.007)	(.006)	(.008)	(.008)	(.007)
Job Effect ($\bar{\theta}$)		.287**	.201**		.643**	.498**		.391**	.292**		.118**	.063**
		(.004)	(.003)		(.007)	(.006)		(.006)	(.005)		(.008)	(.008)
const	.146**	-1.115**	-.633**	.222**	-2.684**	-1.905**	-.030**	-1.759**	-1.208**	.024**	-.478**	-.166**
	(.003)	(.016)	(.015)	(.005)	(.030)	(.027)	(.004)	(.028)	(.024)	(.006)	(.036)	(.033)
Location FE			✓			✓			✓			✓
Adj. R^2	.016	.096	.377	.016	.168	.436	.022	.100	.390	.006	.014	.229
No. Obs	86165	86165	86165	62628	62628	62628	55664	55664	55664	41448	41448	41448

Firm Wage Premium: Difference Between Occupations

Back

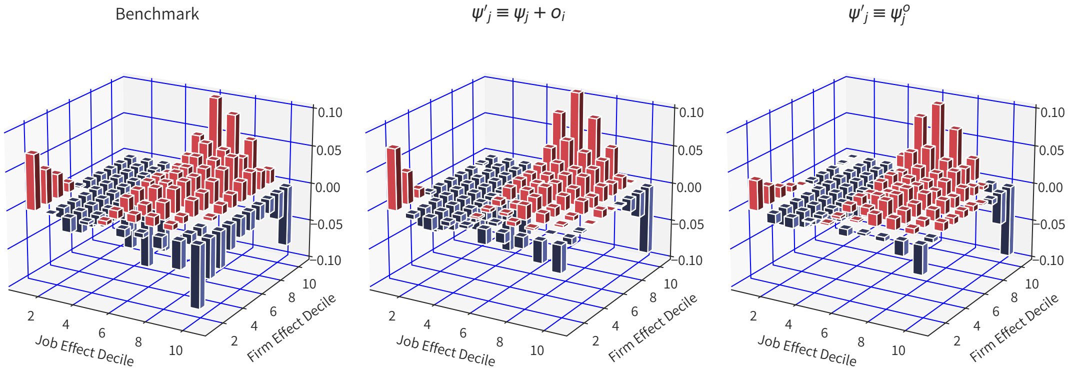


Firm Wage Premium: Firm Size and Firm Location ▸ Back

	Pooled			Computer			Design_Media			Admin		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
fsize.15-50	.019**	.018**	.023**	.012	.011	.014+	.049**	.035**	.045**	-.032	-.039	-.034
	(.004)	(.004)	(.003)	(.010)	(.009)	(.008)	(.011)	(.010)	(.008)	(.038)	(.034)	(.033)
fsize.50-150	.044**	.038**	.050**	.043**	.034**	.032**	.083**	.058**	.073**	-.023	-.038	-.035
	(.004)	(.004)	(.003)	(.010)	(.009)	(.007)	(.010)	(.010)	(.008)	(.038)	(.034)	(.033)
fsize.150-500	.069**	.059**	.068**	.079**	.053**	.043**	.127**	.087**	.094**	-.009	-.032	-.032
	(.004)	(.004)	(.003)	(.010)	(.009)	(.008)	(.011)	(.010)	(.009)	(.038)	(.034)	(.033)
fsize.500-2000	.099**	.081**	.086**	.119**	.070**	.053**	.176**	.121**	.120**	.015	-.014	-.019
	(.005)	(.004)	(.004)	(.011)	(.009)	(.008)	(.012)	(.011)	(.009)	(.038)	(.035)	(.033)
fsize.2000+	.125**	.105**	.121**	.154**	.077**	.065**	.213**	.140**	.134**	.028	-.005	-.006
	(.005)	(.005)	(.004)	(.011)	(.010)	(.008)	(.013)	(.012)	(.010)	(.038)	(.035)	(.034)
Job Effect ($\bar{\theta}$)		.284**	.200**		.793**	.622**		.479**	.395**		.262**	.171**
		(.004)	(.003)		(.009)	(.008)		(.010)	(.009)		(.020)	(.018)
const	.148**	-1.101**	-.630**	-.176**	-3.946**	-3.018**	.157**	-1.931**	-1.488**	.175**	-.919**	-.468**
	(.003)	(.016)	(.015)	(.010)	(.042)	(.037)	(.010)	(.046)	(.040)	(.038)	(.079)	(.073)
Location FE			✓			✓			✓			✓
Adj. R^2	.017	.096	.381	.025	.243	.515	.053	.190	.473	.014	.062	.292
No. Obs	84023	84023	84023	30658	30658	30658	13871	13871	13871	5592	5592	5592

Mean Residual for Work-Firm cells

◂ Back



Deming & Kahn (2018)

◂ Back

Job Skills	Keywords and Phrases: Deming & Kahn (2018)	Keywords and Phrases: Chinese Correspondents
Cognitive	Problem solving, research, analytical, critical thinking, math, statistics	解决, 问题, 研究, 分析, 批判, 思考, 数学, 统计
Social	Communication, teamwork, collaboration, negotiation, presentation	交流, 沟通, 讨论, 演示, 展示, 合作, 团队, 协作
	Matched Keywords and Phrases in V': V_g, V_e	**Matched Keywords and Phrases in V':** $V_{s1}, \ldots, V_{s5}$
Cognitive	分析判断(analysis & judgment); 思考(reflections); 独立思考(independent thinking); 解决问题(problem solving); 数学(mathematics); 研究生(graduate students); 研究者(researchers); 统计学(statistics); 认真思考(think carefully)	统计(statistics); 统计分析(statistical analysis); 问题解答(question answers); 商业分析(business analysis); 行业研究(industry research); 业务分析(business analysis); 关键问题(key issues); 分析(analysis); 分析报告(analysis report); 功能分析(functional analysis); 可行性研究(feasibility study); 解决(solutions); 解决方案(solutions); 问题(question); 市场分析(market analysis); 数据分析(data analysis); 深入分析(in-depth analysis); 深入研究(in-depth research); 研究(research); 兼容性问题(compatibility issues); 定位问题(positioning issues); 疑难问题(difficult questions); 系统分析(system analysis); 面向对象分析(object-oriented analysis)
Social	交流(communication); 人际沟通(interpersonal communication); 协作(collaboration); 合作(cooperation); 团队(team); 团队精神(team spirit); 沟通(communication); 沟通交流(communication);学术交流(academic exchange)	合作项目(cooperation projects); 沟通了解(communication & understanding); 合作方(partners)

Deming & Kahn (2018) Back

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cognitive	.045	.054	.027	.047	.013	.032	.011	.033
	(.000)	(.001)	(.000)	(.001)	(.000)	(.001)	(.000)	(.001)
Social	.035	.041	.030	.045	.020	.033	.025	.041
	(.001)	(.001)	(.001)	(.001)	(.000)	(.001)	(.001)	(.001)
Both required		-.012		-.026		-.024		-.029
		(.001)		(.001)		(.001)		(.001)
Ξ_g, Ξ_m			✓	✓			✓	✓
Ξ_s					✓	✓	✓	✓
Education FE	✓	✓	✓	✓	✓	✓	✓	✓
Experience FE	✓	✓	✓	✓	✓	✓	✓	✓
Occupation FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Adj. R^2	.582	.582	.604	.604	.636	.636	.641	.641

Firm Wage Premium Varies Across Occupations

- Shares of firm effect and sorting (job effect) are larger (smaller) in high-skill occupation than low skill occupation, despite of more features ▸ compare shares
- We also find for low-skilled occupations have estimated firm effects less consistent with the firm effects estimated in high-skilled occupation ▸ compare firm FE

Occupational Specific Specification

- Allow for firm wage premiums varying across major occupations
  $\ln w_i = X_i\beta + \psi_j^o + \iota_t + \epsilon_i$
  - Also compare with $\ln w_i = X_i\beta + \psi_j + o_i + \iota_t + \epsilon_i$

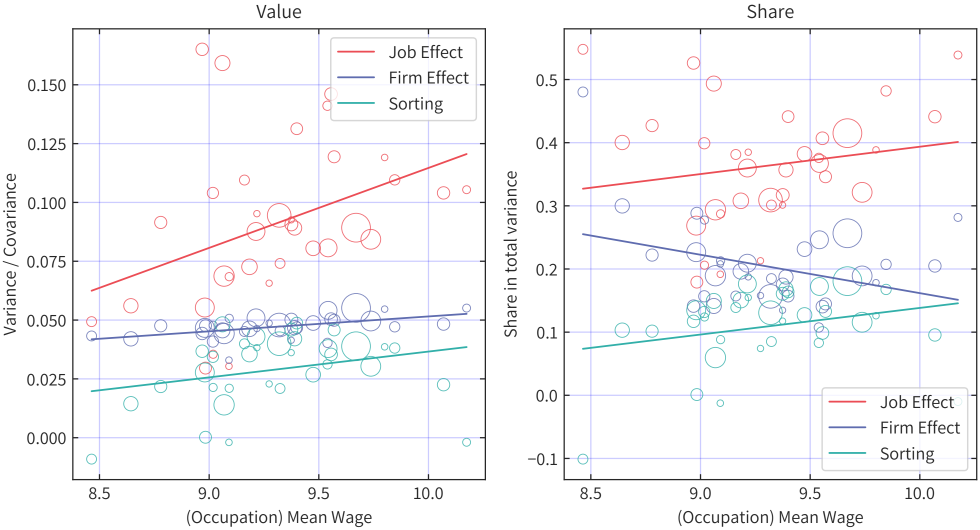
- Allow for skill prices varying across major occupations
  $\ln w_i = \sum_o \mathbb{1}_{[i \in o]} X_i \beta_o + \psi_j + \iota_t + \epsilon_i$

	Benchmark		$\psi_j \equiv \hat{\psi}_j + \hat{o}_i$		$\psi_j \equiv \hat{\psi}_j^o$	
	Comp.	Share	Comp.	Share	Comp.	Share
Var($\ln w$)	.362	-	.362	-	.360	-
Var(θ_i)	.163	.450	.141	.391	.136	.378
Var(ϵ_i)	.098	.272	.096	.265	.088	.245
Var(ψ_j)	.049	.136	.056	.156	.065	.182
2 Cov(θ_i, ψ_j)	.051	.142	.068	.188	.070	.196
Obs	3998840		3998840		3926231	
Firm	86165		86165		300079	

▸ mean residual distribution

Shares Across Occupations

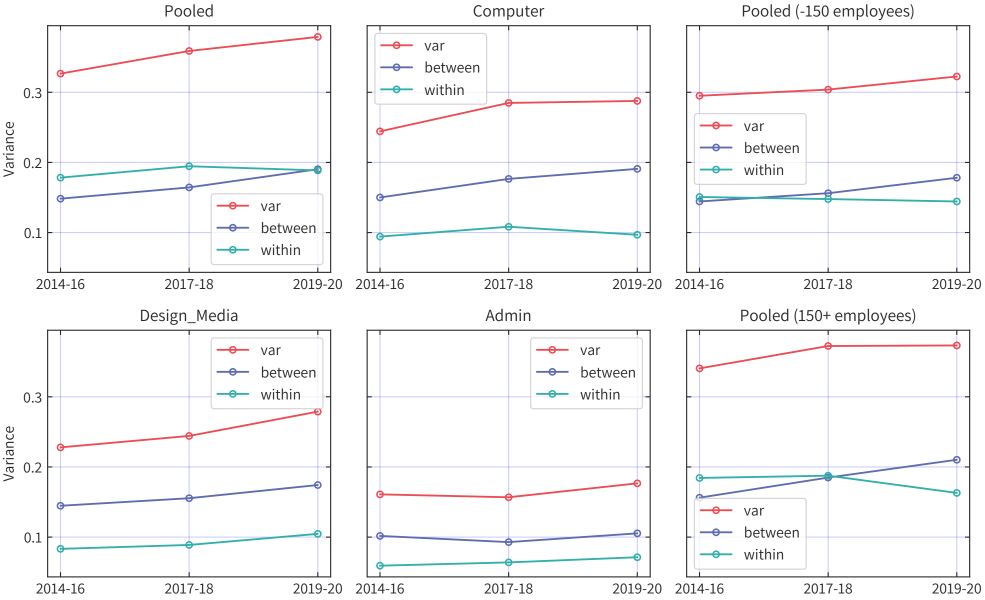
Back



Shares Across Occupations

Back

Posted Wage Variance Trend



Posted Wage Variance Trend Drivers

▸ $\psi_j = \hat{\psi}_j^o$ ▸ new skills

	2014-2016		2017-2018		2019-2020	
	Comp.	Share	Comp.	Share	Comp.	Share
$\text{Var}(\ln w)$	.326	-	.357	-	.377	-
Panel A: $X = \{$**EDU**, **EXP**$, \Xi_2, \ldots, \Xi_8\}$						
$\text{Var}(\theta_i)$	.149	.455	.163	.457	.157	.417
$\text{Var}(\epsilon_i)$	.096	.294	.092	.258	.094	.249
$\text{Var}(\psi_j)$	.048	.148	.050	.141	.059	.157
$2\,\text{Cov}(\theta_i, \psi_j)$	.033	.103	.051	.144	.067	.177
Panel B: Decompose θ Terms						
$\text{Var}(X_{int})$	.039	.121	.043	.120	.041	.109
$\text{Var}(X_{ext})$	.069	.212	.071	.198	.068	.180
$2\,\text{Cov}(X_{int}, X_{ext})$	.040	.123	.049	.139	.048	.128
$2\,\text{Cov}(X_{int}, \psi_j)$	.011	.035	.018	.051	.022	.059
$2\,\text{Cov}(X_{ext}, \psi_j)$	.022	.067	.033	.093	.044	.118
Panel C: Further Decompose X_{ext} Terms						
$\text{Var}(\Xi_g)$	.001	.003	.001	.002	.001	.002
$\text{Var}(\Xi_m)$	.005	.016	.006	.017	.006	.015
$\text{Var}(\Xi_s)$	.039	.120	.039	.109	.037	.098
$2\,\text{Cov}(\Xi_g, \Xi_m)$	.002	.006	.002	.005	.002	.004
$2\,\text{Cov}(\Xi_g, \Xi_s)$	.007	.021	.006	.016	.006	.015
$2\,\text{Cov}(\Xi_m, \Xi_s)$	.015	.046	.018	.049	.017	.045
$2\,\text{Cov}(\Xi_g, X_{int})$	.004	.011	.004	.010	.004	.010
$2\,\text{Cov}(\Xi_m, X_{int})$	.009	.027	.011	.032	.011	.028
$2\,\text{Cov}(\Xi_s, X_{int})$	.028	.085	.034	.096	.034	.090
$2\,\text{Cov}(\Xi_g, \psi_j)$	.002	.005	.002	.006	.003	.008
$2\,\text{Cov}(\Xi_m, \psi_j)$	.007	.020	.010	.027	.011	.030
$2\,\text{Cov}(\Xi_s, \psi_j)$	.014	.043	.022	.060	.030	.080
Obs	930149		1494468		1565866	
Firm	41750		62907		53662	

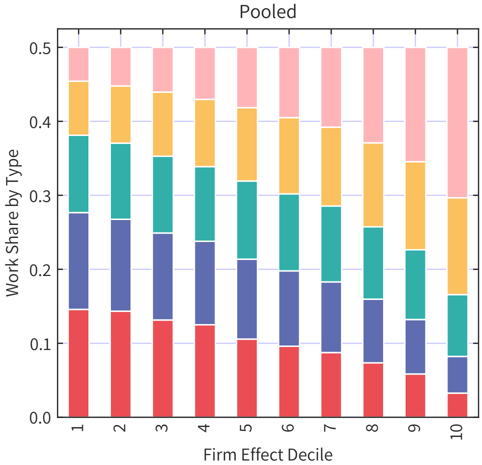
Take-Away Message

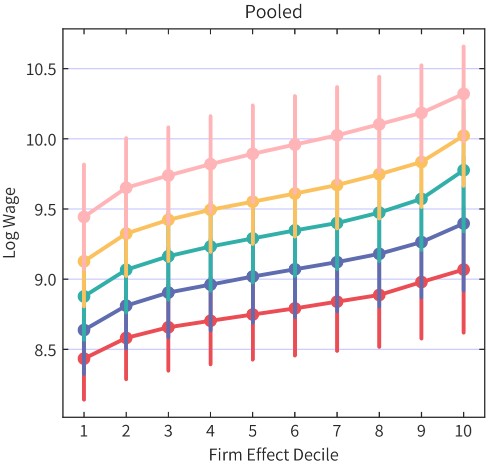
1. Vacancy data + ML $\sim$ EE data + AKM
2. Specificity is (still) an important dimension to think about multidimensional skill/task space
3. Occ-specific & Exp-related skill/task variations are the most important for wage inequality & firm-worker sorting
4. Firms do pay differently for similar-looking jobs, but also varying across occupations
5. Increased posted wage variances in our data is largely due to increased firm-job sorting

Occupational Specific Skill Prices

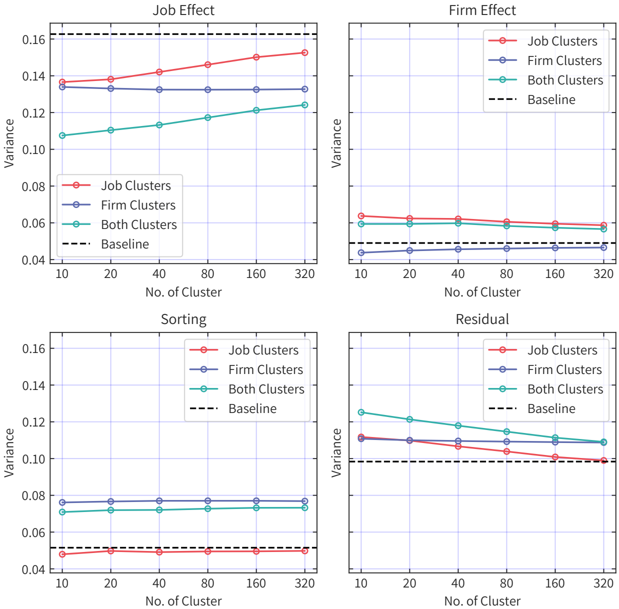
	Benchmark		$X_e\beta_o$		$\tilde{\Xi}\beta_o$		$X\beta_o$		$X\beta_o, \psi_j^o$	
	Comp.	Share	Comp.	Share	Comp.	Share	Comp.	Share	Comp.	Share
$\text{Var}(\ln w)$	.362	-	.362	-	.361	-	.361	-	.359	-
Panel A: $X = \{$**EDU**, **EXP**, $\Xi_2, \ldots, \Xi_8\}$										
$\text{Var}(\theta_i)$	.163	.450	.166	.459	.169	.469	.170	.470	.141	.393
$\text{Var}(\epsilon_i)$	.098	.272	.095	.262	.092	.256	.092	.255	.085	.237
$\text{Var}(\psi_j)$	.049	.136	.050	.137	.049	.136	.049	.136	.063	.175
$2\,\text{Cov}(\theta_i, \psi_j)$	.051	.142	.051	.142	.050	.139	.050	.139	.072	.201
Panel B: Decompose θ Terms										
$\text{Var}(X_{int})$	.042	.115	.053	.146	.040	.111	.048	.134	.039	.108
$\text{Var}(X_{ext})$	.072	.199	.055	.152	.080	.221	.063	.175	.058	.162
$2\,\text{Cov}(X_{int}, X_{ext})$	.049	.136	.058	.161	.049	.136	.058	.161	.044	.123
$2\,\text{Cov}(X_{int}, \psi_j)$	.017	.048	.019	.053	.017	.048	.017	.048	.022	.061
$2\,\text{Cov}(X_{ext}, \psi_j)$	.034	.094	.032	.089	.033	.092	.033	.091	.050	.141
Obs	3998840		3998840		3998840		3998840		3926231	
Firm	86165		86165		86165		86165		300079	

Work Types and Posted Wage by Firm Types



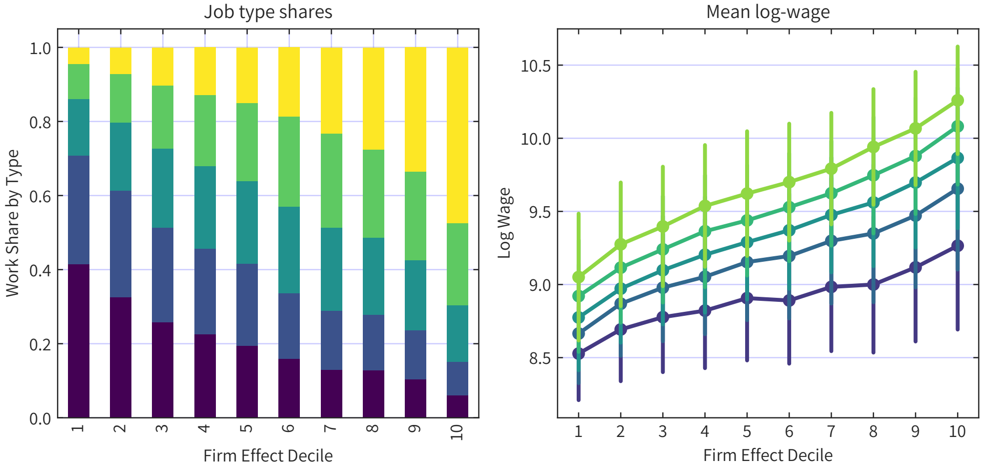


A Shortcut

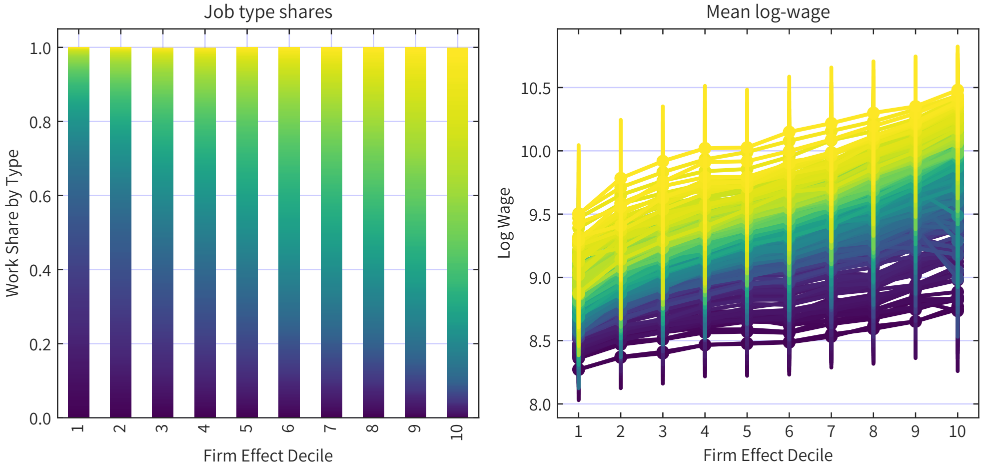


Job Effect
Firm Effect
Sorting
Residual
Variance
No. of Cluster
0.16
0.14
0.12
0.10
0.08
0.06
0.04
10
20
40
80
160
320
Job Clusters
Firm Clusters
Both Clusters
Baseline

Work Types and Posted Wage by Firm Types

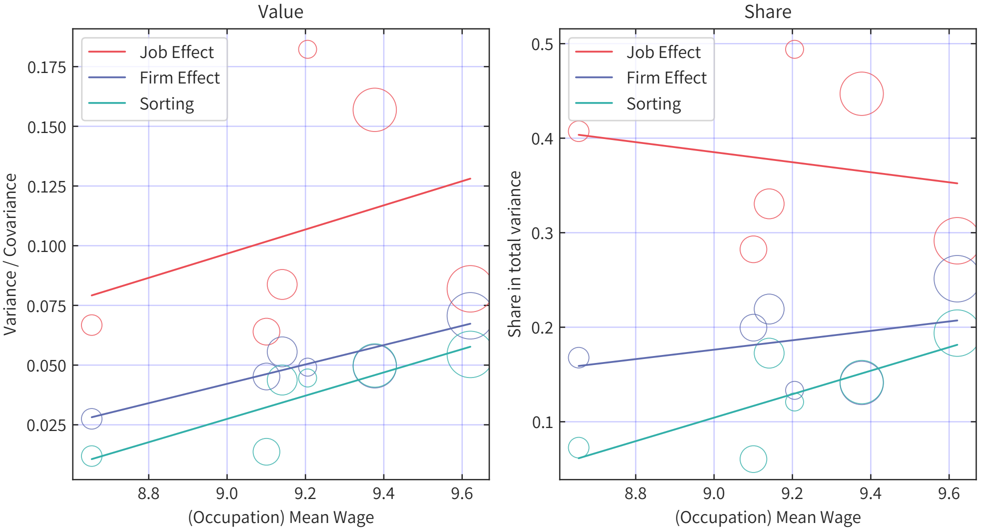


Work Types and Posted Wage by Firm Types



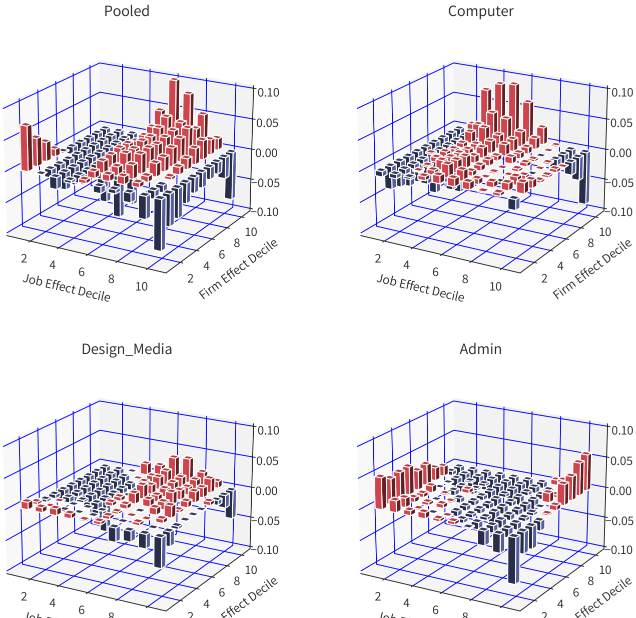
Shares Across Occupations

Back



Mean Residual for Work-Firm cells

Back



Posted Wage Variance Trend Drivers (ψ_j^o) Back

	2014-2016		2017-2018		2019-2020	
	Comp.	Share	Comp.	Share	Comp.	Share
Var($\ln w$)	.322	-	.354	-	.373	-
Panel A: $X = \{$**EDU**, **EXP**, $\Xi_2, \ldots, \Xi_8\}$						
Var(θ_i)	.119	.370	.139	.392	.132	.354
Var(ϵ_i)	.086	.266	.082	.231	.083	.223
Var(ψ_j)	.064	.199	.066	.186	.076	.203
2 Cov(θ_i, ψ_j)	.053	.165	.068	.191	.082	.220
Panel B: Decompose θ Terms						
Var(X_{int})	.038	.117	.041	.115	.039	.104
Var(X_{ext})	.048	.148	.054	.153	.052	.138
2 Cov(X_{int}, X_{ext})	.034	.105	.044	.124	.041	.111
2 Cov(X_{int}, ψ_j)	.017	.053	.024	.067	.028	.075
2 Cov(X_{ext}, ψ_j)	.036	.112	.044	.124	.054	.144
Panel C: Further Decompose X_{ext} Terms						
Var(Ξ_g)	.001	.003	.001	.002	.001	.002
Var(Ξ_m)	.005	.014	.006	.016	.005	.013
Var(Ξ_s)	.025	.079	.028	.078	.026	.071
2 Cov(Ξ_g, Ξ_m)	.001	.004	.002	.005	.001	.004
2 Cov(Ξ_g, Ξ_s)	.005	.015	.005	.014	.005	.013
2 Cov(Ξ_m, Ξ_s)	.011	.034	.014	.039	.013	.036
2 Cov(Ξ_g, X_{int})	.003	.009	.003	.009	.003	.009
2 Cov(Ξ_m, X_{int})	.008	.024	.011	.030	.010	.026
2 Cov(Ξ_s, X_{int})	.023	.072	.030	.084	.029	.077
2 Cov(Ξ_g, ψ_j)	.003	.009	.003	.008	.004	.010
2 Cov(Ξ_m, ψ_j)	.009	.028	.012	.034	.013	.036
2 Cov(Ξ_s, ψ_j)	.024	.075	.029	.083	.037	.099
Obs	888345		1431781		1516033	
Firm	112096		167523		134233	

Posted Wage Variance Trend Drivers ($X\beta_o, \psi_j^o$) ◂ Back

	2014-2016		2017-2018		2019-2020	
	Comp.	Share	Comp.	Share	Comp.	Share
$\text{Var}(\ln w)$	.322	-	.354	-	.373	-
Panel A: $X = \{$**EDU**, **EXP**, $\Xi_2, \ldots, \Xi_8\}$						
$\text{Var}(\theta_i)$	.124	.384	.143	.405	.140	.376
$\text{Var}(\epsilon_i)$	.083	.258	.079	.223	.081	.216
$\text{Var}(\psi_j)$	.062	.192	.063	.179	.073	.195
$2\,\text{Cov}(\theta_i, \psi_j)$	.059	.183	.068	.193	.077	.208
Panel B: Decompose θ Terms						
$\text{Var}(X_{int})$	.036	.113	.039	.111	.037	.100
$\text{Var}(X_{ext})$	.051	.158	.060	.168	.060	.160
$2\,\text{Cov}(X_{int}, X_{ext})$	.036	.113	.044	.125	.043	.116
$2\,\text{Cov}(X_{int}, \psi_j)$	.015	.046	.023	.065	.026	.070
$2\,\text{Cov}(X_{ext}, \psi_j)$	.044	.137	.045	.127	.051	.137
Panel C: Further Decompose X_{ext} Terms						
$\text{Var}(\Xi_g)$	.001	.002	.001	.002	.001	.002
$\text{Var}(\Xi_m)$	.004	.013	.005	.015	.005	.013
$\text{Var}(\Xi_s)$	.031	.095	.033	.092	.033	.089
$2\,\text{Cov}(\Xi_g, \Xi_m)$	.001	.003	.001	.003	.001	.004
$2\,\text{Cov}(\Xi_g, \Xi_s)$	.002	.006	.005	.013	.007	.018
$2\,\text{Cov}(\Xi_m, \Xi_s)$	.010	.033	.016	.044	.014	.037
$2\,\text{Cov}(\Xi_g, X_{int})$	.002	.007	.003	.008	.003	.008
$2\,\text{Cov}(\Xi_m, X_{int})$	.007	.023	.010	.028	.009	.023
$2\,\text{Cov}(\Xi_s, X_{int})$	.026	.082	.032	.089	.032	.085
$2\,\text{Cov}(\Xi_g, \psi_j)$	.005	.015	.003	.008	.001	.003
$2\,\text{Cov}(\Xi_m, \psi_j)$	.010	.031	.011	.032	.013	.036
$2\,\text{Cov}(\Xi_s, \psi_j)$	.029	.091	.031	.088	.037	.099
Obs	888345		1431781		1516033	
Firm	112096		167523		134233	

New Skills/Tasks

Back

	2014-2016		2017-2018		2019-2020	
	Comp.	Share	Comp.	Share	Comp.	Share
$\mathrm{Var}(\ln w)$	.326	-	.357	-	.376	-
Panel A: $X = \{\mathbf{EDU}, \mathbf{EXP}, \Xi_2, \ldots, \Xi_8\}$						
$\mathrm{Var}(\theta_i)$	.148	.455	.163	.456	.156	.415
$\mathrm{Var}(\epsilon_i)$	.096	.294	.092	.257	.093	.248
$\mathrm{Var}(\psi_j)$	.048	.148	.051	.142	.060	.159
$2\,\mathrm{Cov}(\theta_i, \psi_j)$	.034	.103	.052	.145	.067	.178
Panel B: Decompose θ Terms						
$\mathrm{Var}(X_{int})$	.040	.121	.043	.120	.041	.108
$\mathrm{Var}(X_{ext})$	.069	.211	.071	.198	.068	.180
$2\,\mathrm{Cov}(X_{int}, X_{ext})$	.040	.122	.049	.138	.048	.127
$2\,\mathrm{Cov}(X_{int}, \psi_j)$	.012	.035	.018	.052	.023	.060
$2\,\mathrm{Cov}(X_{ext}, \psi_j)$	.022	.067	.033	.093	.044	.118
Panel C: Further Decompose X_{ext} Terms						
$\mathrm{Var}(\Xi_{new})$	.000	.000	.001	.002	.001	.002
$\mathrm{Var}(\Xi_{gm})$	.008	.024	.008	.023	.008	.021
$\mathrm{Var}(\Xi_s)$	.038	.117	.035	.099	.033	.087
$2\,\mathrm{Cov}(\Xi_{new}, \Xi_{gm})$	.001	.002	.001	.004	.002	.004
$2\,\mathrm{Cov}(\Xi_{new}, \Xi_s)$	.001	.004	.003	.009	.003	.009
$2\,\mathrm{Cov}(\Xi_{gm}, \Xi_s)$	.021	.063	.022	.060	.021	.056
$2\,\mathrm{Cov}(\Xi_{new}, X_{int})$	.001	.002	.002	.005	.002	.005
$2\,\mathrm{Cov}(\Xi_{gm}, X_{int})$	.012	.038	.015	.042	.014	.038
$2\,\mathrm{Cov}(\Xi_s, X_{int})$	.027	.083	.033	.092	.032	.084
$2\,\mathrm{Cov}(\Xi_{new}, \psi_j)$	.001	.002	.002	.005	.002	.006
$2\,\mathrm{Cov}(\Xi_{gm}, \psi_j)$	.008	.026	.012	.034	.015	.039
$2\,\mathrm{Cov}(\Xi_s, \psi_j)$	.013	.040	.019	.054	.027	.073
Obs	930149		1494468		1565866	
Firm	41750		62907		53662	

Appendix for Chapter 3.

Related Literature

1. Literature on Compensating Differential:
   - Classic: Rosen (1974); Brown (1980); Rosen (1986); Hwang et al. (1992)
   - Recent: Mas and Pallais (2017); Maestas et al. (2018); Wissmann (2022) / Sorkin (2018); Taber and Vejlin (2020); Lamadon et al. (2022)

   $\rightarrow$ New insights & New theory that reconciles existed empirical failures
2. Literature on Compensation Provision:
   - Theory: Rosen (1974, 1986); Hwang et al. (1998); Hamermesh (1999); Mortensen (2005); Dey and Flinn (2005); Bonhomme and Jolivet (2009)
   - Empirical: Sockin (2022); Lachowska et al. (2022); Bana et al. (2022); Lamadon et al. (2022)

   $\rightarrow$ New evidences & New theory that explains those new evidences
3. Literature on Efficiency Wage:
   - Salop and Salop (1976); Shapiro and Stiglitz (1984); Katz (1986); Krueger and Summers (1988); Bloesch et al. (2021)

   $\rightarrow$ Apply the insights to a more suitable place: ”Efficiency Compensation”

Unstructured Text Data

- V: full vocabulary set with 110,000+ tokens/features (i.e. words or terms)
- $V_{\text{comp}} \subset V$: compensation vocabulary set with 13,000+ features
  - Not all uniques: synonyms, different versions, typos
  - Common words or stop words
  - Irrelevant texts
- $\mathbf{C}_{\text{comp}} \in \mathbb{R}^{N \times |V_{\text{comp}}|}$: an indicator matrix to run regression
- So, high-dimensional data $\rightarrow$ (basic) Machine Learning methods

Posted Compensation/Amenity Information

- Pros:
  1. Hard to observe in census or survey data
  2. Compensations or amenities that firms regard as important to attract workers
  3. Also observe detailed job information

- Cons:
  1. Not a full list of the compensations that a firm offer
  2. Mainly amenities, rare disamenities (strategic hiding?)
  3. Maybe cheap talk?

- Our empirical results will be mainly descriptive & exploratory
  - No priori, let the data speak
  - Find stylized facts of patterns & correlations in the data
  - Shed new insights in thinking theories

Lasso Regressions ◂ Back

- Lasso regression (L1 penalization):

$$\hat{\zeta} = \arg\min_{\zeta} \sum_{i=1}^{N} \left(\ln w_i - \sum_{k=1}^{K} c_{ik}\zeta_k \right)^2 + \lambda \sum_{k=1}^{K} |\zeta_k|$$

- BIC as the criterion to gauge the hyperparameter λ:

$$\min \mathrm{BIC}(\lambda) = \frac{\| \ln \mathbf{w} - \mathbf{C}\hat{\zeta}_\lambda \|^2}{\sigma^2} + \widehat{df}_\lambda \log N$$

- Inference via subsampling (10x10)

Lasso Regression using V_{comp}: Top Features (Frequency $> 1\%$)

◂ lasso details

	Top Positive			Top Negative		
	token	coef	freq	token	coeff	freq
1	14th month pay	.331	.013	five insurance	-.301	.020
2	large platform	.310	.016	commission	-.195	.022
3	three meals	.263	.013	young	-.186	.012
4	technology	.247	.025	easy	-.181	.014
5	guru	.223	.024	training	-.174	.018
6	flexibility	.149	.091	two-day weekend	-.154	.140
7	options	.146	.043	promotion	-.138	.068
8	shuttle	.144	.015	events	-.104	.010
9	remuneration	.124	.015	holiday	-.093	.017
10	six insurance & one fund	.121	.050	holidays	-.092	.046
11	platform	.114	.046	provide	-.084	.012
12	13th month pay	.114	.021	jobs	-.080	.097
13	supplementary	.107	.011	achievements	-.077	.010
14	stock	.099	.017	work system	-.076	.012
15	salary	.099	.025	travel	-.073	.058
16	good platform	.093	.010	entrepreneurship	-.069	.013
17	listed company	.091	.023	five insurance & one fund	-.068	.261
18	high salary	.074	.018	employees	-.066	.029
19	products	.073	.012	time	-.063	.012
20	lucrative	.069	.018	environment	-.062	.038
21	shareholding	.069	.012	double pay	-.055	.032
22	benefits	.068	.035	office	-.047	.018
23	motivation	.063	.016	company	-.043	.050
24	projects	.058	.030	wide	-.041	.012
25	year-end bonus	.057	.042	snacks	-.041	.013
26	team	.050	.108	growing	-.039	.025
27	treatment	.040	.027	transportation	-.029	.021

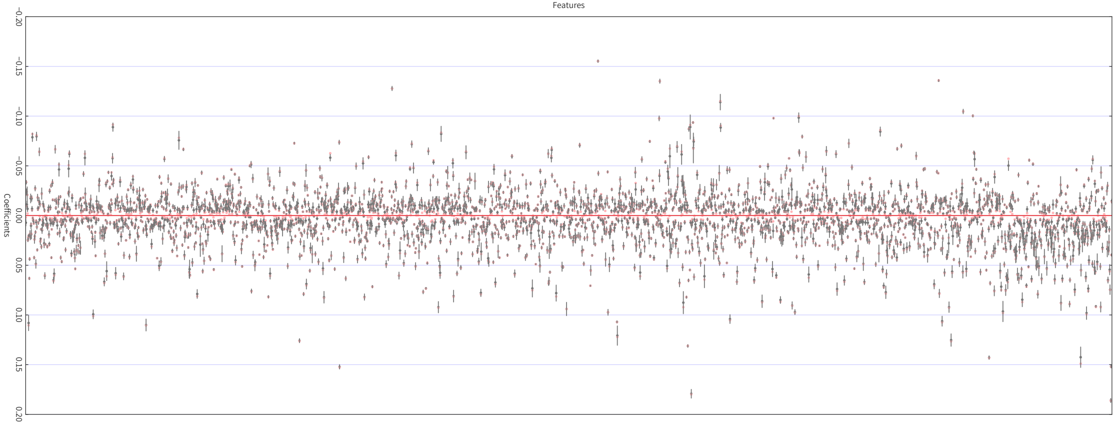
Lasso Regression using V: Top Features (Frequency $> 1\%$)

back

	Top Positive			Top Negative		
	token	coef	freq	token	coeff	freq
1	14th month pay	.152	.014	freshmen	-.155	.018
2	three meals	.143	.014	five insurance	-.136	.030
3	large platform	.131	.019	graduates	-.128	.033
4	master degree	.126	.015	vocational major	-.100	.036
5	lead	.107	.041	two-day weekend	-.098	.166
6	c++	.092	.051	vocational college	-.094	.148
7	algorithm	.082	.061	assistant	-.079	.011
8	guru	.082	.028	customer service	-.075	.030
9	famous	.079	.019	social insurance	-.073	.028
10	machine learning	.077	.016	accounting	-.071	.019
11	formation	.076	.013	accommodation	-.067	.016
12	undergraduate	.074	.319	administration	-.067	.027
13	overseas	.072	.026	commissioner	-.063	.011
14	react	.072	.020	taobao	-.059	.015
15	development	.071	.374	assistance	-.058	.164
16	undergraduate	.066	.029	ps	-.056	.029
17	high salary	.063	.028	ltd.	-.056	.012
18	landing	.060	.067	installation	-.055	.020
19	strategy	.057	.047	photoshop	-.052	.039
20	live streaming	.056	.014	careful	-.050	.032
21	listed company	.055	.027	hardworking	-.050	.032
22	large scale	.055	.072	verification	-.048	.011
23	responsibilities	.055	.048	human resources	-.047	.032
24	shuttle	.054	.018	website	-.047	.090
25	finance	.054	.070	any major	-.047	.020
26	six insurance & one fund	.053	.055	humanization	-.046	.012
27	python	.052	.066	excel	-.046	.047

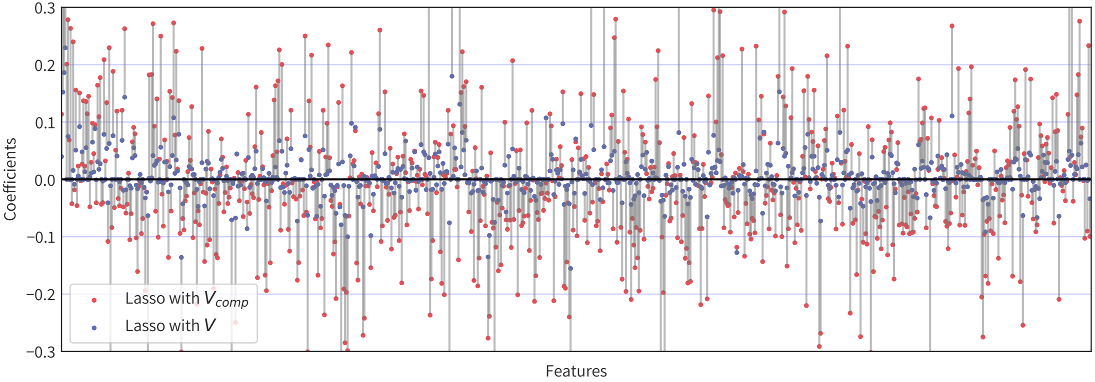
Confidence Intervals on Lasso Coefficients via Subsampling

◂ Back



Compare Lasso Coefficients

Back



Posted-Wage Regression ◂ Back

- So the predictive power of non-wage compensations in part comes from their correlation with job skills/tasks; What about firms?

- Posted wage regression: $\ln w_{i,j,t} = \theta_i + \psi_j + \delta_i + \iota_t + \epsilon_i$
  - $\theta_i \equiv X_i \beta$ (job/worker effect), $X_i = \{\text{EDU}_i, \text{EXP}_i, \mathbf{c}'_{i,\setminus\text{comp}}\}$
  - ψ_j (firm fixed effect)
  - $\delta_i \equiv \mathbf{c}'_{i,\text{comp}} \gamma$ (compensation effect)
  - ι_t (year fixed effect)
  - In practice, further dimensional reduction on $\mathbf{c}'_{i,\setminus\text{comp}}$ & $\mathbf{c}'_{i,\text{comp}}$ using PLS
  - This posted wage regression does a similar job to the AKM framework (Zhu, 2022)

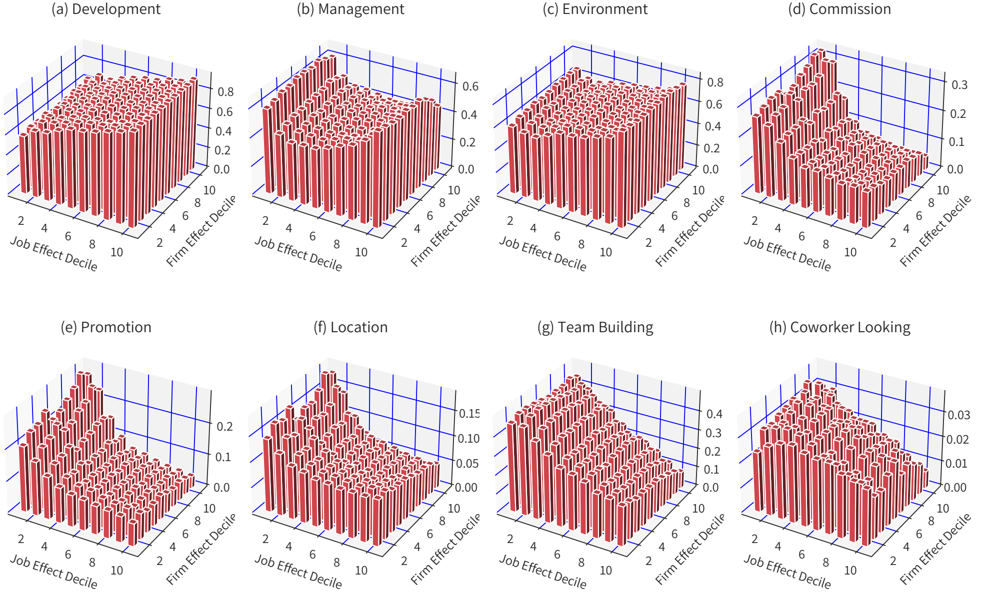
- Variance decomposition: $\text{var}(\ln w_i) =$
  $\text{var}(\theta_i) + \text{var}(\psi_j) + \text{var}(\delta_i) + 2\,\text{cov}(\theta_i, \psi_j) + 2\,\text{cov}(\theta_i, \delta_i) + 2\,\text{cov}(\psi_j, \delta_i) + \text{var}(\epsilon_i)$

Gather Important Types and Check Occurrence

- We can take a direct look on if high/low wage firms or jobs are accompanied with low/high valued amenities
- We do this by selecting a set of major, well-defined, and economic important compensations from V_{comp} based on the frequency & Lasso coefficient
- We gather all relevant terms by checking proximate terms in the embedding space of a work-embedding model trained on the whole job texts
- We then examine how the occurrence ratio for each type differ across different firms & jobs

Compensation Occurrence (More)

Back



Hedonic Regression

	Pooled	Computer	Design_ Media	Admin
	(1)	(2)	(3)	(4)
Advanced Insurance	.014**	.016**	.009**	.002
	(.001)	(.001)	(.002)	(.003)
Backloading Wage	.010**	.013**	.022**	.011**
	(.001)	(.001)	(.002)	(.002)
Stock Option	.087**	.068**	.060**	.040**
	(.001)	(.001)	(.002)	(.003)
Coworker Quality	.024**	.016**	.005*	.008+
	(.001)	(.001)	(.002)	(.004)
Work-Flexibility	.010**	.007**	.009**	.005**
	(.001)	(.001)	(.001)	(.002)
Basic Insurance	-.025**	-.024**	-.017**	-.013**
	(.000)	(.001)	(.001)	(.001)
Training	-.003**	-.019**	-.003	.013**
	(.001)	(.001)	(.002)	(.002)
Work-Time	-.021**	-.018**	-.020**	-.022**
	(.000)	(.001)	(.001)	(.001)
Education FE	✓	✓	✓	✓
Experience FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
$\Xi_2, \ldots, \Xi_8$	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓
Adj. R^2	.738	.748	.730	.657
No. Obs	3998840	1325260	548808	260364

The Phantom of Unobserved Worker Ability ◂ Back

- Yes, there still could be unobserved worker ability not-captured which cause bias in the estimation above (Rosen, 1986; Hwang et al., 1992)
- But would unobserved skill heterogeneity matter so much?
  - In our job vacancy data, the usually-unobserved job heterogeneity accounts for additional 5 percent of the posted wage variances
  - Unobserved job heterogeneity is typtically positively correlated with observed job heterogeneity
- Perhaps compensation differential is not the sole or the major force?
  - The toughness of the omitted-variable problem indicates other dominant mechanism of compensating dispersion

Firms' Problem ◂ Back

- Firm problem: $$\max_{\{q_i\}_{i=1}^N, a, h, w(q)} AN^{1+\alpha} \prod_{i=1}^N q_i e(a,h) - \sum_{i=1}^N w(q_i) - a\kappa N$$
$$\text{s.t.} \quad w(q) + \phi_a a - \frac{h^{1+\phi_h}}{1+\phi_h} \geq u(q) \quad \forall q \in \{q_i\}_{i=1}^N$$
  - Complementary production function & additively separable utility function ensure positive assortative matching (PAM) even under imperfect transferable utility
    $\rightarrow$ a firm will employ workers with same q

- Rewrite the firm problem given equilibrium allocation:
  $\max_{q,a,h} \; AN^{1+\alpha} q^N (1 + \gamma_a a + \frac{h^{\gamma_h}}{\gamma_h}) - N \left(u(q) - \phi_a a + \frac{h^{1+\phi_h}}{1+\phi_h} \right) - a\kappa N$

- FOCs: $$AN^{1+\alpha} q^{N-1} e(a,h) = u'(q)$$
$$AN^{\alpha} q^N h^{\gamma_h - 1} = h^{\phi_h}$$

Market Utility Profile ◂ Back

- $u(q) = \begin{cases} \frac{\left(\bar{A}q^N\right)^{1+\omega}}{(1+\omega)(1+\gamma_h)} + (1+\gamma_a)\bar{A}q^N + u_a, & \text{if } q \geq q_a \\ \frac{\left(\bar{A}q^N\right)^{1+\omega}}{(1+\omega)(1+\gamma_h)} + \bar{A}q^N + u_0, & \text{if } q < q_a \end{cases}$
  - where $\bar{A} \equiv AN^{\alpha}$, $\omega = \frac{1+\gamma_h}{1+\phi_h-\gamma_h}$, $u_0 = 0$, and $u_a = \phi_a - \kappa$.

If Firm Size Is Endogenous (Typical O-Ring Results) ◂ Back

- N is also a choice of the firm
- Additional FOC: $AN^{\alpha} q^{N} e(a, h) \left(1 + \alpha + N \ln(q)\right) = w + ac$
- Optimal choice on firm size: $N(q) = \frac{1+\alpha}{-\ln(q)}$
- Firm size increases in productivity q and is irrelevant to the choices of amenities
- All the relationships between productivity and amenity provision can be now directly translate to the firm size

Model Implications 1. Compensating Differential

1.1 Compensating effects can be confounded with productivity effects
   - Esp. for the up-end labor market where (in)efficiency forces are strong

1.2 The result of an empirical test on compensating differential will depend on the targeted labor market
   - If focusing on low-end labor market (close to q_a or $q < q_a$ with imperfectly mandated policies) $\rightarrow$ easy to find clear evidence
   - If focusing on board or high-end labor market (& with heterogeneous usage in efficiency compensation or imperfect matching) $\rightarrow$ tests likely to fail

1.3 Available variations for wage-amenity packages can be limited conditional on worker
   - Depends on exogenous heterogeneity v.s. endogenous heterogeneity
   - Constrains on both low-end and high-end markets

$\rightarrow$ Field/choice experiments (WtP) or RCT-like experiments (exogenous variations) not necessarily capture the whole picture of how labor market works

Model Implications 2. Labor Market Inequality

2 Efficiency compensations can enlarge both utility dispersion & wage dispersion
  - Ignoring non-wage compensations can underestimate labor market inequality
  - Moreover those compensations per se can actually be the drivers of wage inequality

$\rightarrow$ Increased sorting or better use of efficiency compensations increases wage inequality

Model Implications 3. Job Mobility & Choice

3.1 The set of non-wage compensations that can justify job moves to low wage-premium firms is likely limited to inefficient amenities

- Work-time/effort is the most likely culprit for moving downgrade

3.2 Greater compensating than just ”compensating differential”

- A worker with a ϕ_h shock would suffer not only traditional compensation differential but also a worse matching & an inferior package of other compensations
- Again, available choices for wage-amenities packages are limited

$\rightarrow$ Potential implications for gender wage gap and etc.

Take-Away Message

1. Think explicitly about non-wage compensations: insurance/fund, work-time, pay schemes, work environment, fringe benefits, ...
   $\rightarrow$ empirical focus & policy targets & intuition when back-out revealed preference
2. Different Firms in different jobs have distinct provision patterns
   $\rightarrow$ compensating differential $\neq$ provision inequality
3. (In)Efficiency compensations & productivity sorting reconciles empirical findings and generates important implications
   $\rightarrow$ high-wage firms can also offer better compensations without wage discounts

Appendix for Chapter 4.

Related Literature

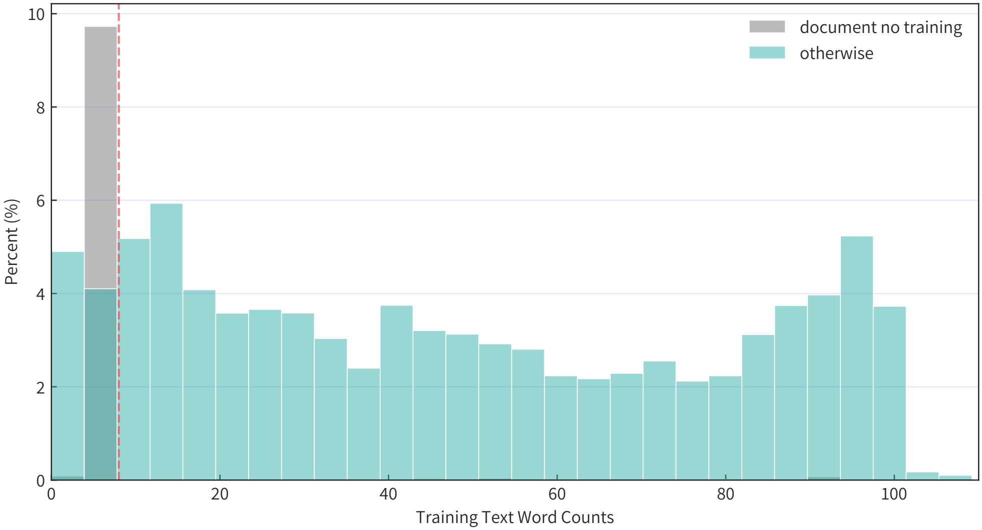
- Literature on the impact of human capital or labor market on technology adoption: Nelson and Phelps (1966), Greenwood and Yorukoglu (1997), Chari and Hopenhayn (1991), Adão et al. (2021), Galor and Moav (2000), Krueger and Kumar (2004a,b), Acemoglu and Zilibotti (2001); esp. due to the holdup problem: Acemoglu (2003), Acemoglu et al. (2006); esp. empirical evidences on IT technology: Bloom et al. (2012), Arora et al. (2013), Michaels et al. (2014)
- Literature on training and human capital investment under non-Walrasian market: Acemoglu (1996), Acemoglu (1997), Acemoglu and Pischke (1998), Acemoglu and Pischke (1999b), Acemoglu and Shimer (1999), Moen and Rosén (2004), Wasmer (2006), Doepke and Gaetani (2020), Engbom (2022)
- Literature on cross-country relationship between labor market turnover and training, development, or lifecycle wage growth: Blinder and Krueger (1996), Donovan et al. (2022), Ma et al. (2021), Engbom (2022)
- Literature on endogenous labor market institutions: Acemoglu et al. (2006), Acemoglu et al. (2017); esp. on the Japanese labor market institutions: Hashimoto (1979), Hashimoto and Raisian (1985), Morita (2001), Owan (2004)

Data Source

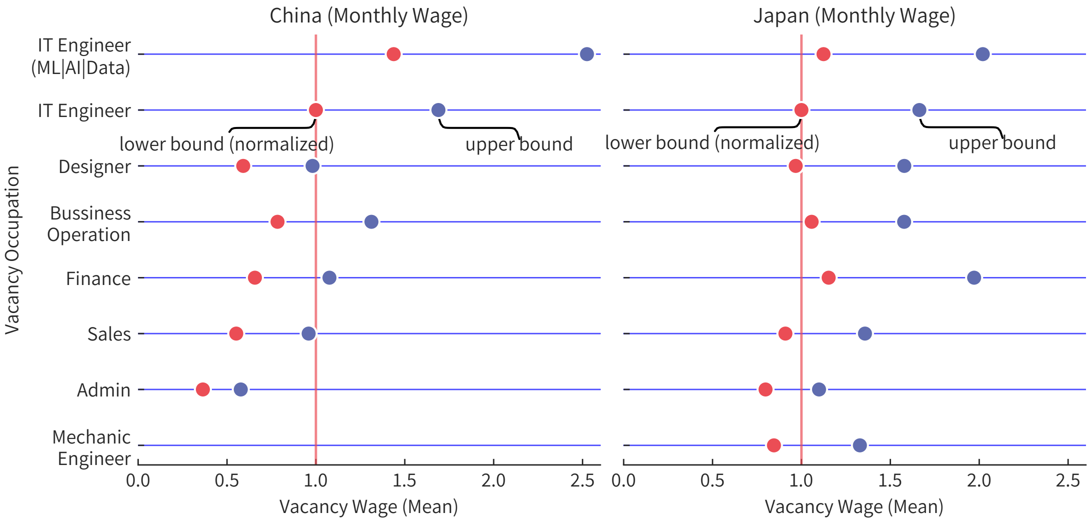
- Japan: vacancy data from `Doda.com`
  - Largest general Job boards in Japan
  - IT vacancy: 34,000 / All vacancy: 216,000
  - Time period: 2019/06-2020/03
- China: vacancy data from `Lagou.com`
  - Largest IT-centered job board in China
  - IT vacancy: 278,000 / All vacancy: 909,000
  - Time period: 2019/01-2019/12
- Only regular jobs, but same results applying to new graduates
- Confirm by using Labor Census data in Japan

Training Text in Japanese IT Vacancies

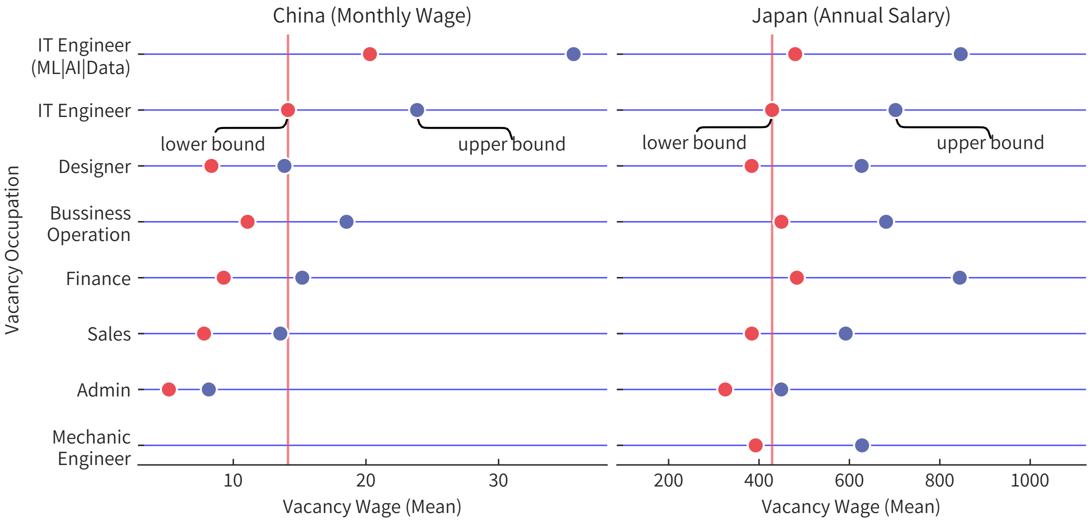
◂ Back



Posted Wage (Monthly)

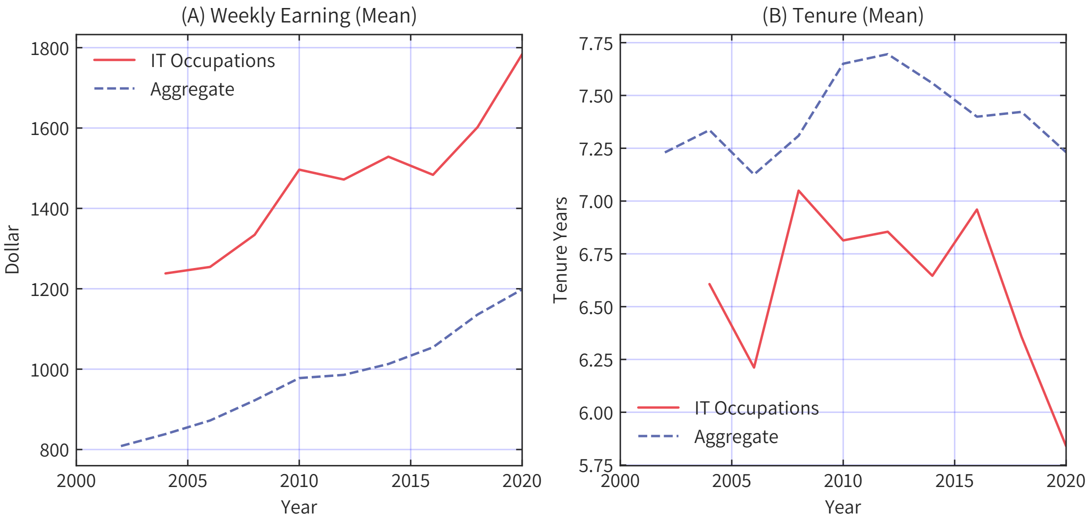


Posted Wage (Real)



US Census Data (CPS)

Back



Other Evidences from Literature

- IT capital productivity: Bloom et al. (2012) finds European affiliates of US multinationals have higher productivity in using IT capital than non-US multinationals and domestic firms, which can be accounted by different "people management" practices (promotions, rewards, hiring, and firing)
- Patent data: Arora et al. (2013) shows that Japanese firms were increasingly lagging behind US firms in IT-related invention during a software-biased shift in the innovation process in IT sectors
- Demand on skill: Michaels et al. (2014) shows that a positive correlation between high-skill(education) workers' demand/wages and ICT adoption(investment) across countries and industries

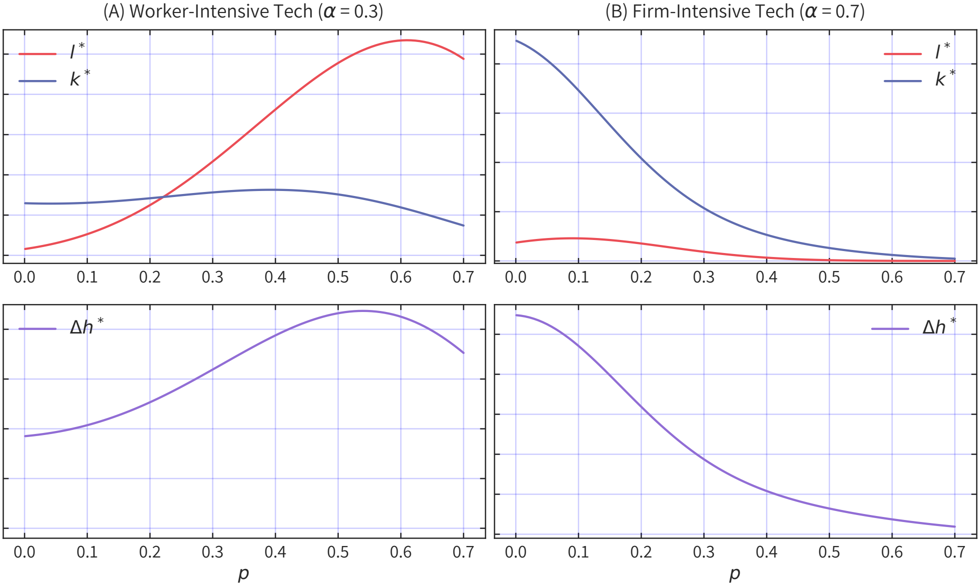
Fact Implications

- In the case of recent IT sectors, firm training (Japan) seems to be less efficient than worker learning (China) in technology adoption and innovation
- More generally, it implies that firm training and worker learning are not equally efficient or not perfect substitutes in human capital investment and their importances may vary across sectors and technologies, due to
  - Technological reasons: e.g. if reply on equipment or work environment
  - Contractual reasons: e.g. moral hazard problems, credit constraints
  - Often both
- It further implies that Japan might be trapped in its labor market institutions
  - The well-known Japanese labor market institutions featured by more-training, less-turnover, and less skill-premium seem to extend to the newly emerged IT sectors, despite of its inefficiency

Investment Choices

- FOCs: $\Gamma_l(1-\alpha)Ak^{*\alpha}l^{-\alpha} = \kappa l^{\gamma}$; $\Gamma_k A\alpha k^{(\alpha-1)}l^{*(1-\alpha)} = r$
  - $\Gamma_l = (1-p)\,\beta z_1 + p\,((1-\beta)z_1 + \beta z_2)$ increases in p
  - $\Gamma_k = (1-p)(1-\beta)z_1$ decreases in p
- $l^* = \left(\Gamma_l^{1-\alpha}\Gamma_k^{\alpha}A\alpha^{\alpha}(1-\alpha)^{1-\alpha}r^{-\alpha}\kappa^{\alpha-1}\right)^{\frac{1}{\gamma(1-\alpha)}}$
- $k^* = \left(\Gamma_l^{1-\alpha}\Gamma_k^{\alpha+\gamma}A^{1+\gamma}\alpha^{\alpha+\gamma}(1-\alpha)^{1-\alpha}r^{-(\alpha+\gamma)}\kappa^{\alpha-1}\right)^{\frac{1}{\gamma(1-\alpha)}}$
- $\Delta h^* = \left(\Gamma_l^{1-\alpha}\Gamma_k^{\alpha(1+\gamma)}A^{1+\gamma}\alpha^{\alpha(1+\gamma)}(1-\alpha)^{1-\alpha}r^{-\alpha(1+\gamma)}\kappa^{\alpha-1}\right)^{\frac{1}{\gamma(1-\alpha)}}$

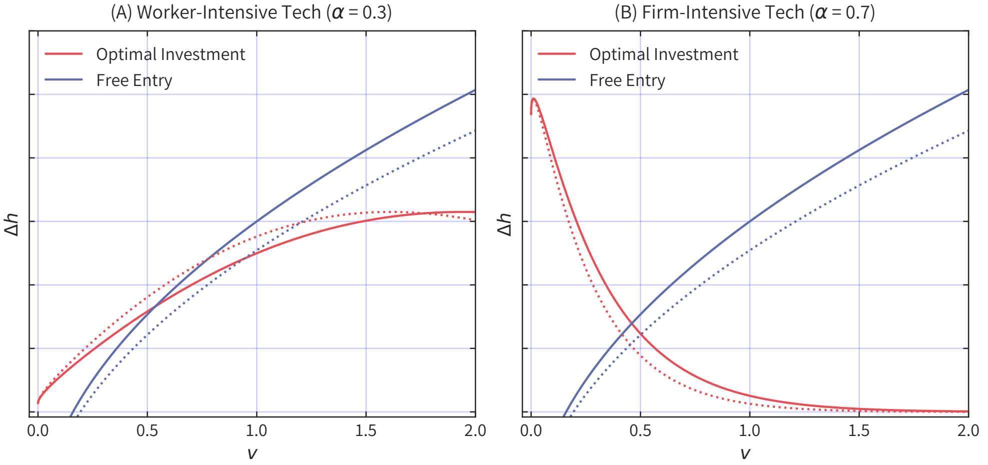
$$\Delta h = A k^{\alpha} (l + l_0)^{(1-\alpha)}$$



Efficient Allocation

- An allocation is (constrained) efficient if it maximizes the net output of the economy subject to search frictions
- A social planner chooses training investment and vacancy opening to maximize the output in the second period:
  $\max_{l,k,v}[(1-p(v))z_1 + p(v)z_2](1+\Delta h(k,l)) - rk - \kappa\frac{l^{1+\gamma}}{(1+\gamma)} - cv$
  s.t. $\Delta h(k,l) = Ak^{\alpha}l^{(1-\alpha)}$ and $p = \xi(v/s)^{\phi}$

Increase in Search Efficiency ξ



The Hypothesis

- Japan-China differences lay in the different labor market institutions developed to solve the firm hold-up problem in the early stage
- Japan: achieved by social norms or customs (under historical contingencies) which applies to the whole economy
- China: achieved by using a large state-owned sector with low turnover (Feng and Guo, 2021) and high investment (Song et al., 2011) but left an intact and fluid labor market in the private sector.
    - Along with structural transform, the government sectors dampened or reformed gradually (Hsieh and Song, 2015) and the new industries like IT grow entirely from a very fluid labor market in the private sector
    - The state-owned sector might be less efficient in solving the firm hold-up problem than the Japanese system because worker can still flow out to the private sector and there may have other moral hazard problems in SOEs

Labor Market Mobility and Economics Development

- Donovan et al. (2022): labor market liquidities are negatively correlated with development, with workers moving on and off the bottom rungs of the job ladder
- Ma et al. (2021): the levels of firm-provided training are positively correlated with development, suggesting a hold-up problem in firm training in LDCs with large portion of self-employment
- Engbom (2022): a positive relationship between labor market fluidity and lifecycle wage increase among European countries
- Our model suggests a U-shape relationship:
  - Low income countries need rigidity to solve the firm hold-up problem to catch-up
  - High income countries need liquidity to solve the worker incentive problem under new TC