

# Job Postings for Working At Home/Remote Work: Before, During, and After COVID-19

- VERY PRELIMINARY -

---

Jörn Boehnke, Boao Zhan, Qian Wang, Yang You, & Richard B. Freeman

NBER Work-From-Home Conference, Nov 3

# Job Postings for Working At Home/Remote Work: Before, During, and After COVID-19

- Job postings provide insight into firm's future employment.
- Postings arise from vacancies and close after hiring.
- LinkUp:
  - Indexes job listings from 60,000+ US employer websites.
  - Data from 2007 with increasing coverage.
  - Includes 6-digit SOC classifications.
- GPT analyzed postings to determine work mode: WFH, hybrid, or in-person.
- Examined variations in firms with COVID-related PPP support.



# Content

1. Job Postings Data and GPT Methodology for Determining WFH/Hybrid
2. Analyzing Job Postings and WFH/Hybrid Posting as Measure of Economic Performance
3. Using Job Postings Data to Examine PPP Support of Firms

# 1. Job Postings Data and GPT Methodology for Determining WFH/Hybrid

- Raw Data: 226 million job postings from LinkUp (2007-2022) from company websites.
- Postings are unique (no duplicates) since sourced directly from firms.
- Coverage: 60,000 companies in 195 countries; 71.3% US-based.
- Data ~10TB in CSV and XML format.
- Excluded postings without firm names; sampled 50%, resulting in 112M observations for analysis.
- Main variable: STOCK of JOB POSTINGS (SJP) - measures postings at a specific time.
- Data indicates if posting was DELETED (job filled!?) or newly CREATED (new job or vacancy from turnover).
- SJP calculation:  $SJP(-1) - DELETED + CREATED$ .

# Why GPT?

- Prior studies used bag-of-words to gauge a company's focus by counting keyword occurrences.
- This method can overestimate due to high false-positive rates.
- Example: “... AppleCare seeks a leader for global contact centers, **home-based** teams ... This position will be based **onsite in Austin, TX ...**”

(Source: LinkUp)

- For accuracy, a sophisticated rule set is needed.
- Human input is possible, but LLM offers a better solution.
- Lack of Work-From-Home labels makes supervised training impractical.
- Solution: Pre-trained LLM.

# Using GPT to determine WFH/Hybrid status

- Analysis aimed to classify postings as WFH, hybrid, or likely in-person.
- 3-stage process:
  1. Extracted sentences with at least one keyword from 50 WFH/hybrid terms generated by GPT-4.
  2. GPT-3.5 API helped create an algorithm for classification into WFH, hybrid, or default.
  3. Probing default group for in-person requirement; most “default” are “in-person” but depends on industry.

# Stage 1: Preprocessing

- Job postings are often lengthy with over 3,000 characters; many are irrelevant for Work-From-Home classification.
- Used bag-of-words to preprocess postings for GPT; keywords generated by GPT-4.
- Postings without keywords are classified default and not processed by GPT.
- Postings with keywords suggest potential work-from-home/hybrid roles; relevant sentences extracted for further analysis.

# Stage 2: Job excerpts classified using GPT-3.5 as: work-from-home, hybrid, or default

- Use excerpt-based GPT-3.5 classification because GPT-4 very costly and slow; cuts computational and monetary cost by about 95% and improves accuracy by 0.3%.
- Default includes postings without WFH information and where WFH/Hybrid is not allowed.
- Chosen GPT-3.5 parameters: n=10, temperature=0.7.

- Prompt:

Read the following job posting excerpt: [INPUT THE JOB POSTING EXCERPT]

Based on the job posting excerpt, please select the category that best describes the modality of this job:

Work from home: This job is done from home, without any on-site/in-person requirements.

In-person/unclear: Inferred from this job posting, this job involves on-site/in-person activities that cannot be done from home, such as traveling.

Hybrid: This job allows working from home but involves (optional) on-site/in-person activities.

If this job posting does not specify whether the job is working from home or in-person so either one may be possible, classify it as "In-person/unclear".

If this job involves remote communication with teammates/clients/third parties but the job posting does not explicitly mention the option to work from home, classify it as "In-person/unclear".

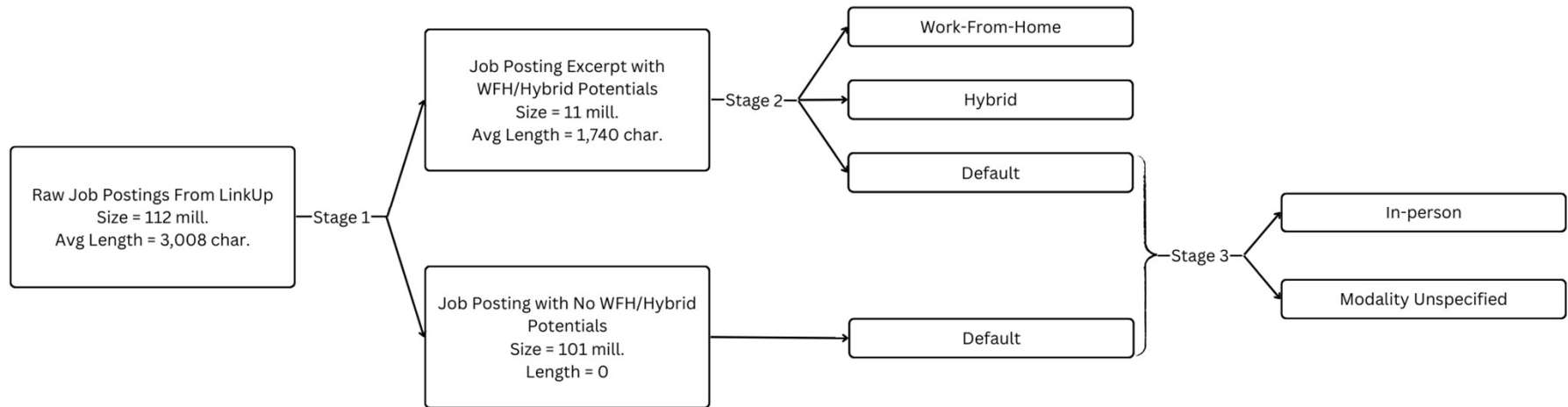
If the "Work from home" option is provided as an optional benefit, classify it as "Hybrid".



## Stage 3: Probing the defaults with GPT-4

- The default category has two sub-categories: no information provided, and WFH/Hybrid not allowed.
- To distinguish these, we use GPT-4 for a small random sample since GPT-3.5 struggles with the nuanced differences.
- Excerpts in this stage retain details about allowing/not allowing WFH/Hybrid.
- Use 157 keywords to identify jobs not allowing WFH/Hybrid.

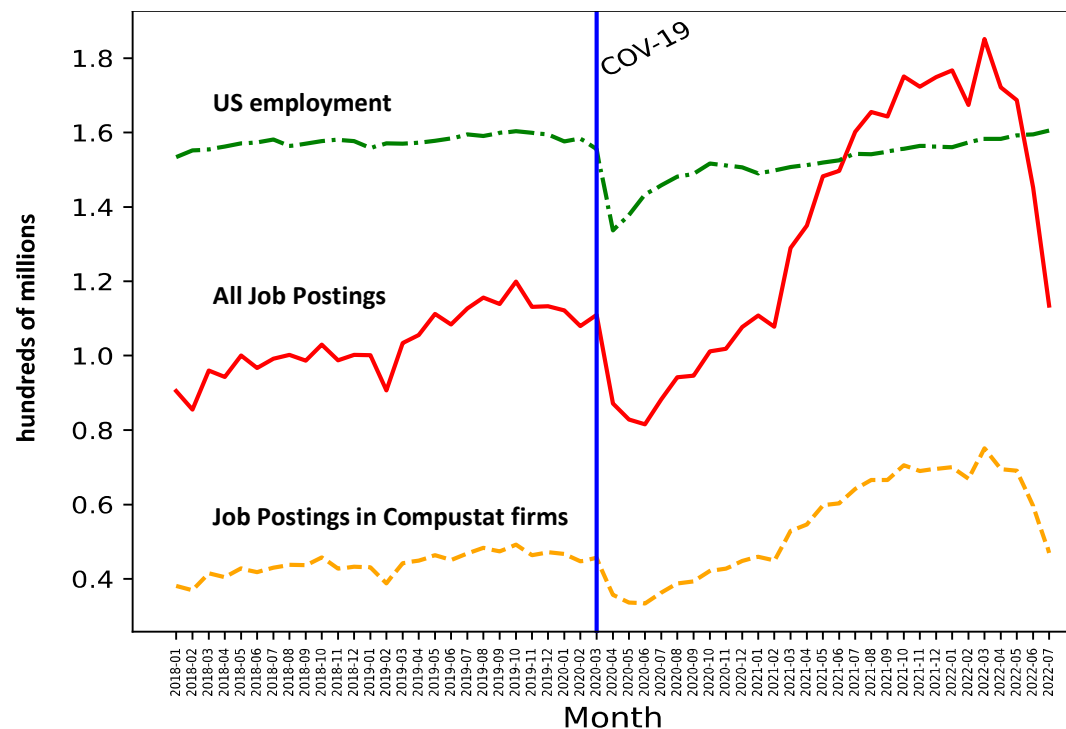
# Summary



# Evaluating the Algorithm's Performance

- Tested GPT's accuracy against 1,000 “difficult” job postings categorized by human experts.
- Lower bound of 3-way classification is 93%.
- Can utilize n=10 setting to achieve higher accuracy levels.
- **Methodology can be applied to any job characteristic determined by a body of data**

## 2. Analyzing Job Postings and WFH/Hybrid Posting as Measure of Economic Performance



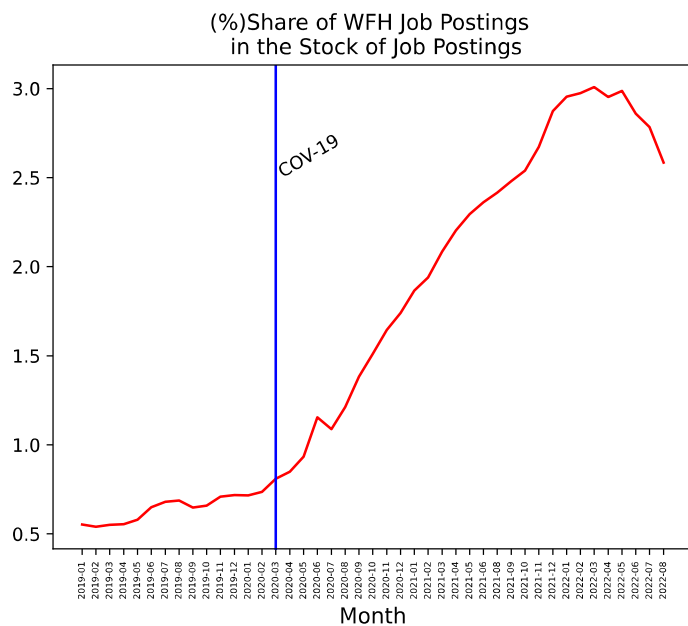
Source: US employment, CPS; Monthly Job Postings, from LinkUp Database; Postings in Compustat firms are for 4671 firms in the LinkUp Data base and are in Compustat

Monthly level of all job postings and of postings for Compustat firms compared to total US employment from Jan 2018 to COVID-19 (March 2020) to July 2022

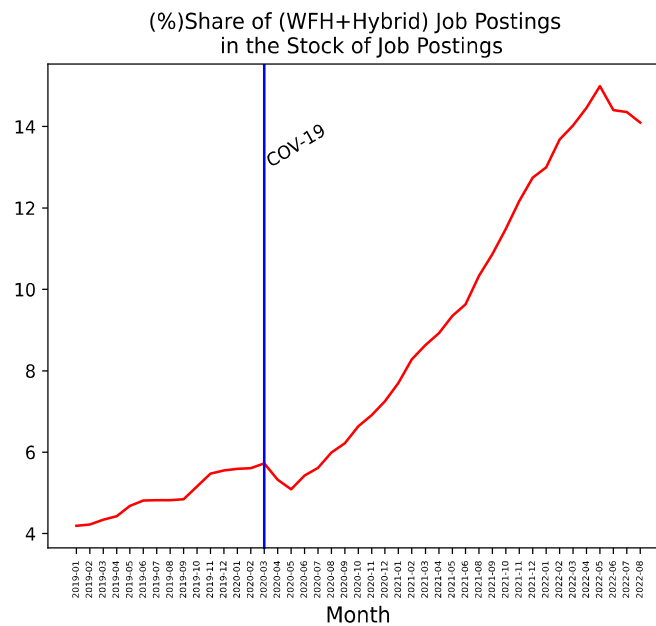
### Comments:

1. Compustat firms account for ~40% of all posts.
2. Postings fall in COVID-19 shock almost simultaneous with employment.
3. Postings show greater increase in recovery
4. Note: drop in summer 2022 is artifact of data collection.

# Share of WFH and WFH/Hybrid in all postings, based on our classification



Source: The job posting information is from LinkUp.



Source: The job posting information is from LinkUp.

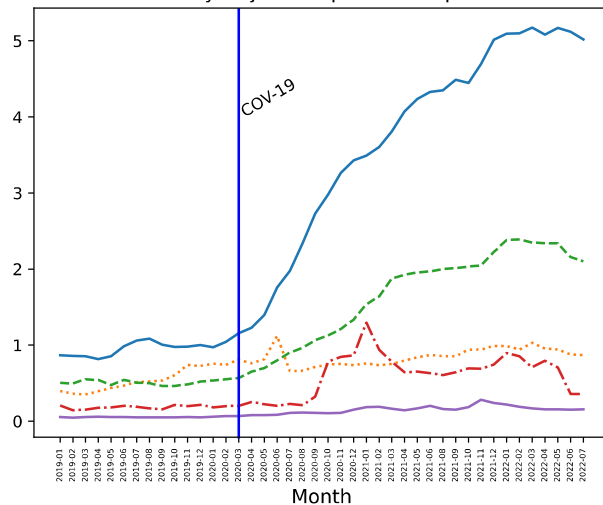
## Comments:

1. WFH Share of Postings rises from <0.7% before COV-19 to peak at ~ 3%
2. WFH/Hybrid Postings are ~5-fold WFH Postings and peaks ~15%
3. Drop in WFH and WFH/Hybrid Postings reflects firm pull-back from WFH as recovery proceeded.

# WFH and WFH/Hybrid share of all postings, by 1-digit occupation

Shows big increase for management and professional related jobs and for sales and office jobs compared to other occupations

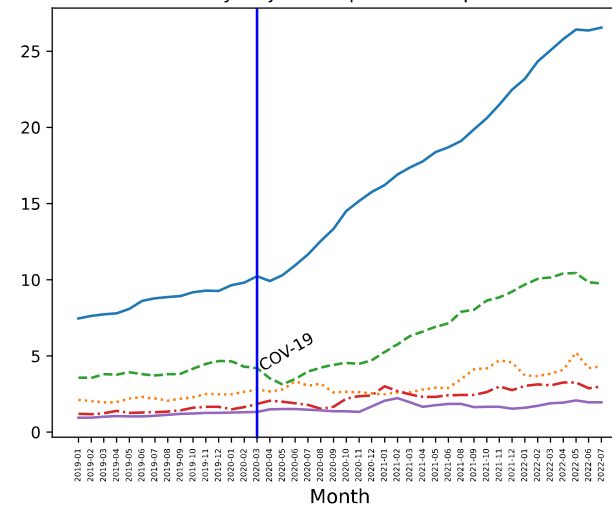
Share of WFH Job Postings(%) in Stock by Major Occupation Group



- Group 1-Management/Finance/Computer/Math/Architecture/Engineer/Science/Law/Education related
- - - Group 2-HealthCare/Protective/Food/Cleaning/PersonalCare/Meida/Entertainment
- - - Group 3-Sales/Office and Administrative Support
- - - Group 4-Farming/Construction/Maintenance
- Group 5-Production/Transportation/Moving

Source: The job posting information is from LinkUp. Major Occupation Group is defined by 2-digit O'net code.

Share of (WFH+Hybrid) Job Postings(%) in Stock by Major Occupation Group

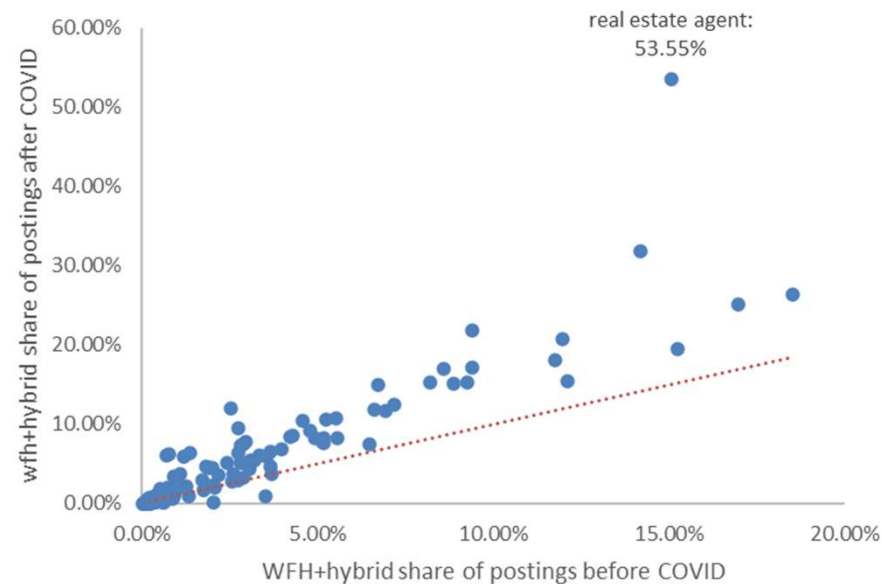


- Group 1-Management/Finance/Computer/Math/Architecture/Engineer/Science/Law/Education related
- - - Group 2-HealthCare/Protective/Food/Cleaning/PersonalCare/Meida/Entertainment
- - - Group 3-Sales/Office and Administrative Support
- - - Group 4-Farming/Construction/Maintenance
- Group 5-Production/Transportation/Moving

Source: The job posting information is from LinkUp. Major Occupation Group is defined by 2-digit O'net code.

## Analysis of 106 detailed 6-digit Occupations:

80% had increases in WFH or Hybrid share of job postings



Calculated from LinkUp Data on Job postings, with WFH/Hybrid posting determined by our GPT Algorithm from 6-digit detailed occupations with 500 or more job postings in each month. Left out Telecommunications Equipment Installers and Repairers, Except Line Installers for unstable data over time.

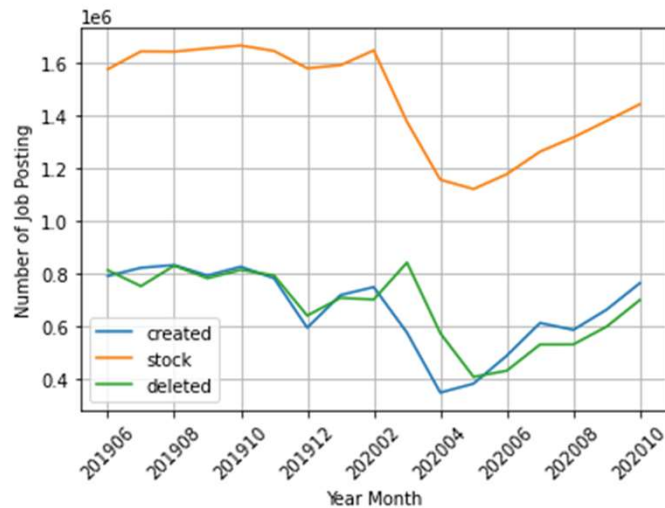
Top occupations with biggest growth are primarily specialized white-collar jobs.  
 While bottom occupations are primarily blue-collar jobs → increased inequality

Occupation	WFH/Hybrid share of job posting, Jan 2015 to Dec 2019	WFH/Hybrid share of job posting, Mar 2020 to Aug 2020	Change
<b>Top 5 occupations Ranked by before pandemic WFH+Hybrid share of job posting</b>			
<i>Mean</i>	<b>13.46%</b>	<b>25.00%</b>	<b>11.55%.</b>
Training and Development Specialists	18.53%	26.38%	7.85%.
Sales Engineers	16.97%	25.19%	8.22%.
Sales Representatives, Wholesale and Manufacturing, Technical and Scientific Products	15.23%	19.55%	4.32%.
Real Estate Sales Agents	15.06%	53.55%	38.49%.
Clinical Research Coordinators	14.18%	31.91%	17.72%.
<b>Bottom 5 occupations Ranked by before pandemic WFH+Hybrid share of job posting</b>			
<i>Mean</i>	<b>0.09%</b>	<b>0.14%</b>	<b>0.05%.</b>
Maids and Housekeeping Cleaners	0.08%	0.04%	-0.04%.
Janitors and Cleaners, Except Maids and Housekeeping Cleaners	0.08%	0.35%	0.27%.
Dishwashers	0.05%	0.00%	-0.05%.
Light Truck Drivers	0.03%	0.04%	0.01%.
Opticians, Dispensing	0.00%	0.00%	0.00%.

Calculated from LinkUp Data on Job postings, with WFH/Hybrid posting based on our classification of 106 6-digit detailed occupations with 500 or more job postings each month



### 3. Using Job Postings Data to Examine the Paycheck Protection Program (PPP) Support of Firms



- Between 2020-2023, government allocated 4-5 trillion for COVID-19, including PPP.
- USAspending.gov data shows 9,614 COVID-related awards over \$10 million, totaling \$1.6 trillion – 1/3 of COVID spending.
- Examined overlap between Compustat and LinkUp; our sample represents 63%-72% of Compustat's total revenue.
- Analysis of 3,505 companies' job postings from June 2019-October 2020:
- Stable postings (1.5-1.6 million) until February 2020.
- March-April 2020 saw a drop by 200,000 postings.
- Recovery from May-October 2020, influenced by government aid and easing restrictions.

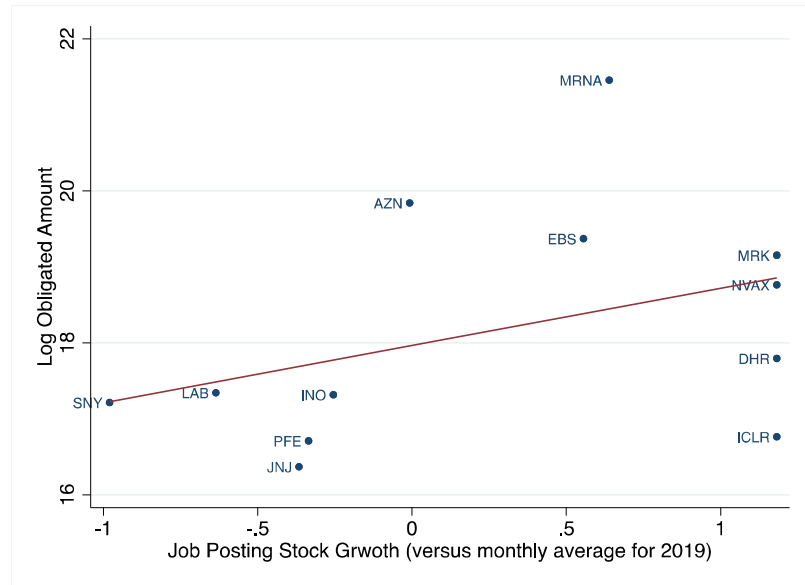
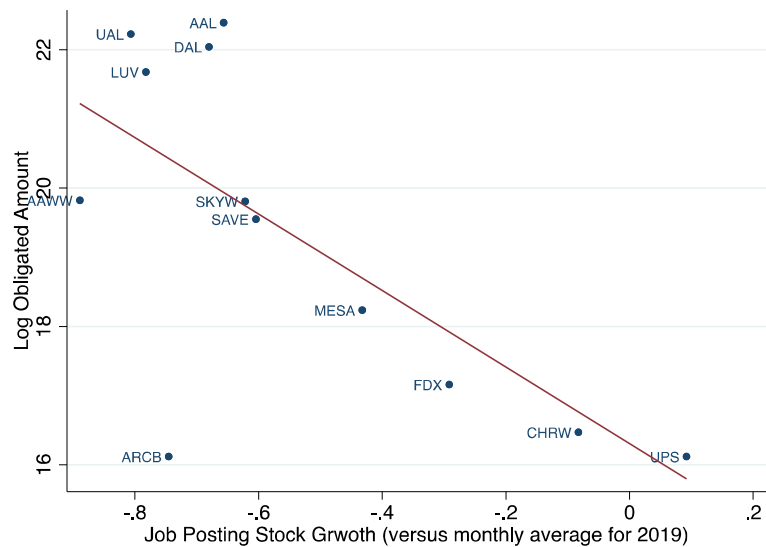
# The Paycheck Protection Program (PPP)

- We focus on the first big impact of government spending on public companies, after matching to CUSIP and keeping only companies with LinkUp job posting data
- Ended up with 212 events (awards) covering the period from March 2020 to April 2022 for 9 GICS Sectors, 84 companies.
- **Q1:** Did moneys go to firms with employment/hiring problems evident in job postings?
- **Q2:** Did fiscal spending help these companies recover demand for labor?
- **Q3:** What kind of jobs were firms trying to fill?

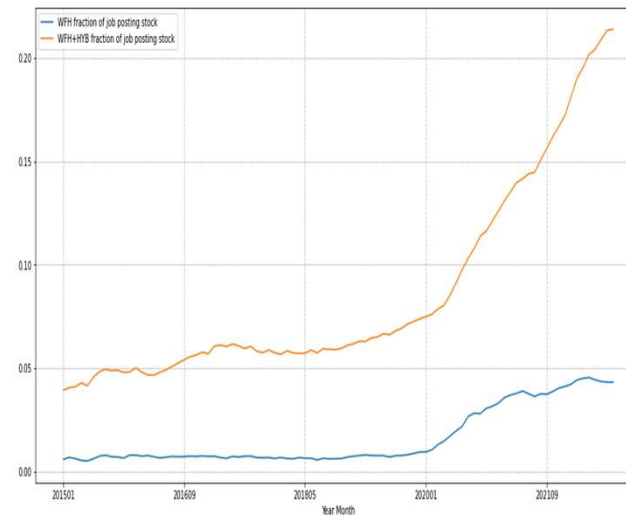
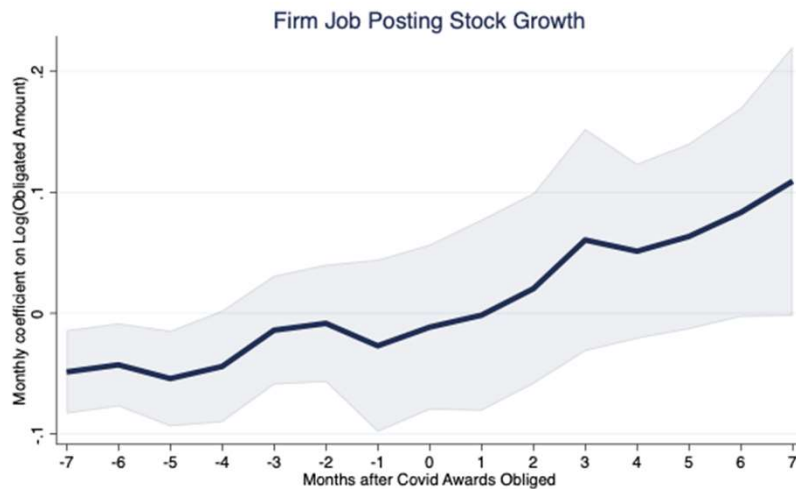
Top 20 companies by total obligated amount

Company Name	Number of Awards	Total Obligated Amount
UNITEDHEALTH GROUP INC	6	268.98 Billion
AMERICAN AIRLINES GROUP INC	3	15.93 Billion
DELTA AIR LINES INC	4	14.98 Billion
UNITED AIRLINES HOLDINGS INC	3	13.51 Billion
SOUTHWEST AIRLINES	3	8.92 Billion
MODERNA INC	2	4.16 Billion
LEIDOS HOLDINGS INC	8	1.25 Billion
SKYWEST INC	3	1.2 Billion
SPIRIT AIRLINES INC	3	929.18 Million
MCKESSON CORP	2	825.0 Million
EMERGENT BIOSOLUTIONS INC	3	774.94 Million
RITE AID CORP	2	761.16 Million
GENERAL DYNAMICS CORP	14	669.2 Million
BOEING CO	6	582.2 Million
GENERAL MOTORS CO	1	476.14 Million
ABBOTT LABORATORIES	4	439.02 Million
GENERAL ELECTRIC CO	4	421.09 Million
ASTRAZENECA PLC	1	413.2 Million
ATLAS AIR WORLDWIDE HLDG INC	1	406.84 Million
HANESBRANDS INC	2	346.5 Million

Q1: Biggest recipients of PPP were transportation (GICS Group 2030) and Pharmaceuticals, Biotechnology & Life Sciences (GICS Group=3520) with very different results...



# Q2: Did fiscal spending help these companies recover hiring?



- Job postings say yes, but WFH share exploding
- Most are likely white collar

# Conclusions

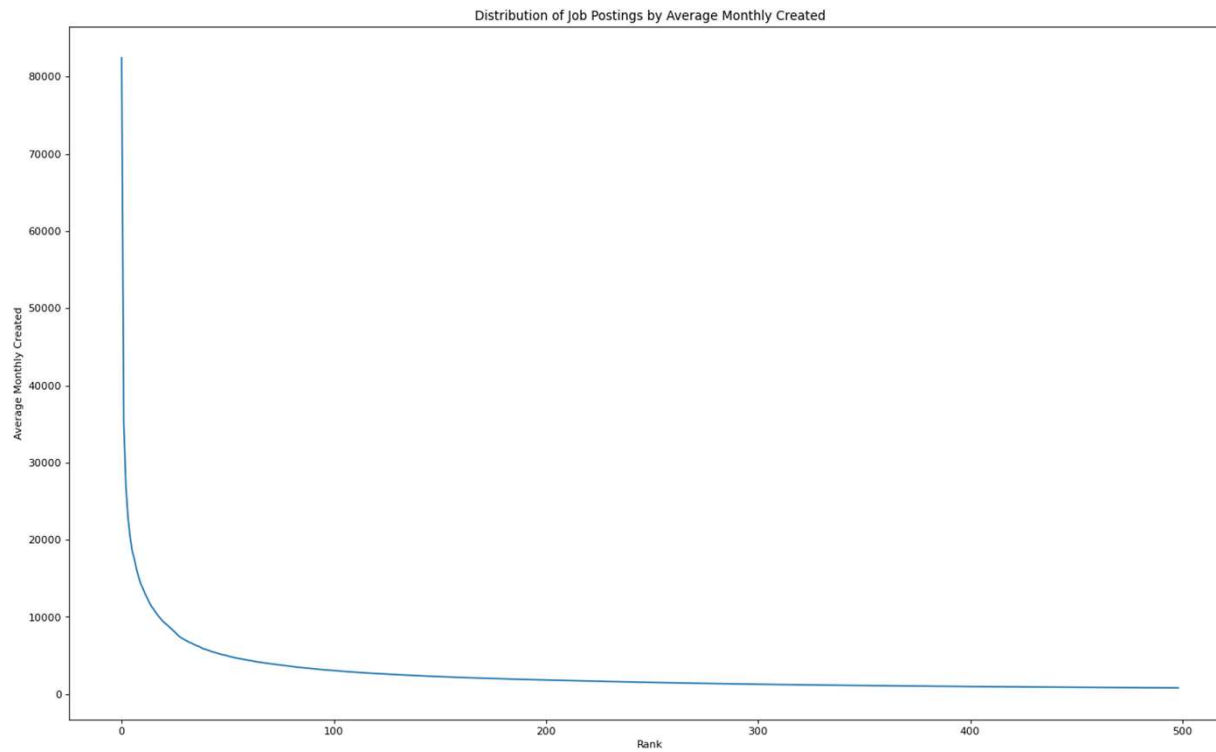
- Analyzed 226 million job postings from 2007-2022.
- Utilized GPT models for classification of job modalities (WFH, Hybrid , and Default/In-person).
- Classification fits time pattern of COVID-19 employment shock and locates rise in remote and hybrid jobs, primarily white-collar roles.
- Increase in the WFH/Hybrid ratio for companies that received substantial government Payroll Protection Program spending.
- Job postings valuable addition to labor market data.



Questions

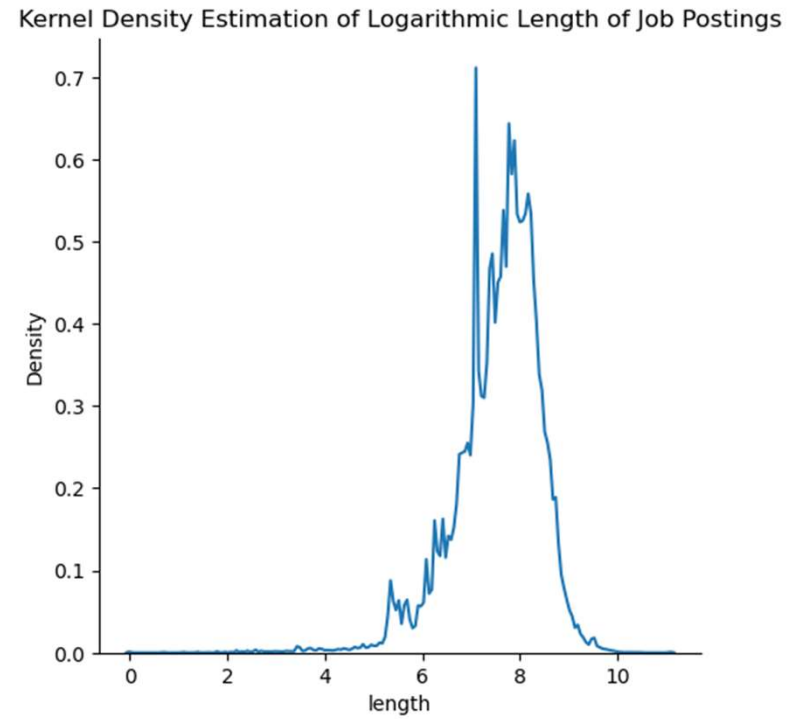
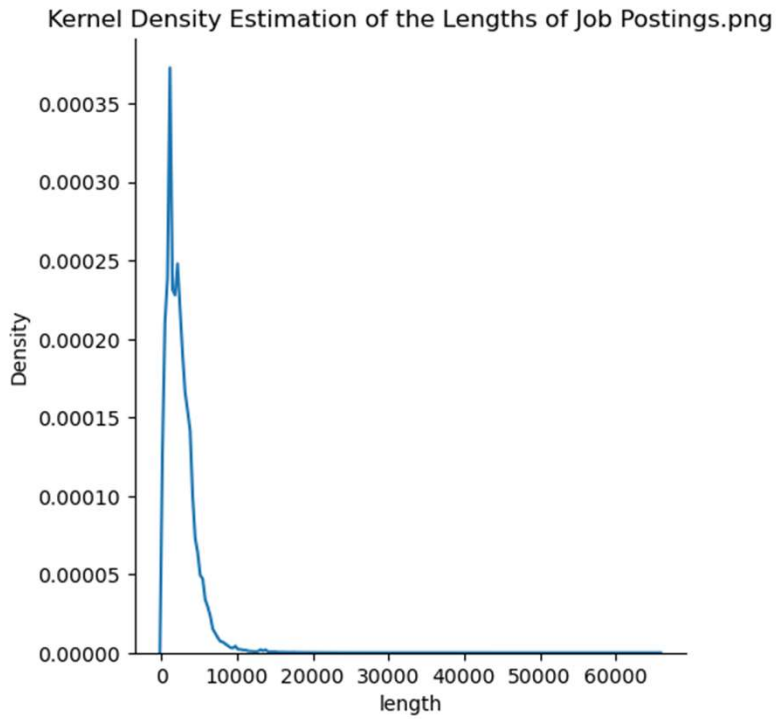
# Appendix

# Distribution of 500 Top-Job-Posting Companies



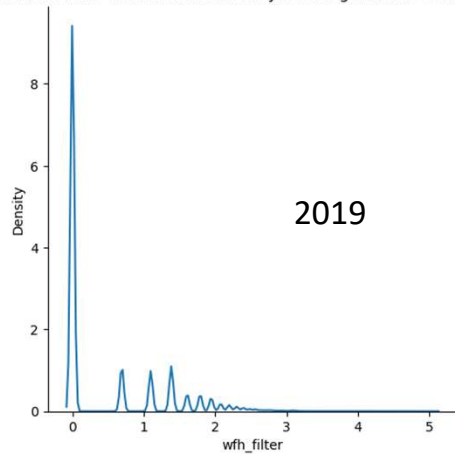


# Distribution of Job Posting Lengths

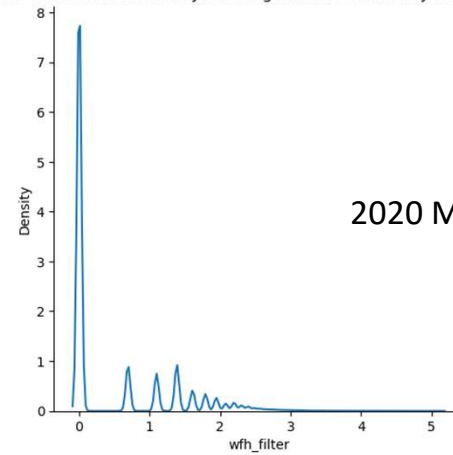


# Distribution of WFH keywords over time (log x+1)

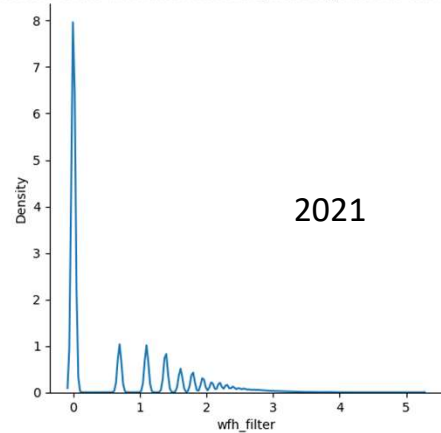
Kernel Density Estimation of the Work-From-Home Keyword Logarithmic Counts in Job Postings in 2019



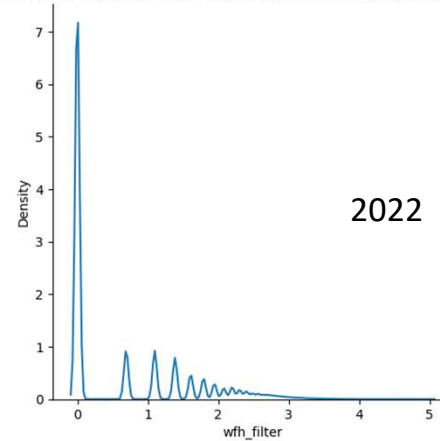
Kernel Density Estimation of the Work-From-Home Keyword Logarithmic Counts in Job Postings in March to December, 2020



Kernel Density Estimation of the Work-From-Home Keyword Logarithmic Counts in Job Postings in 2021

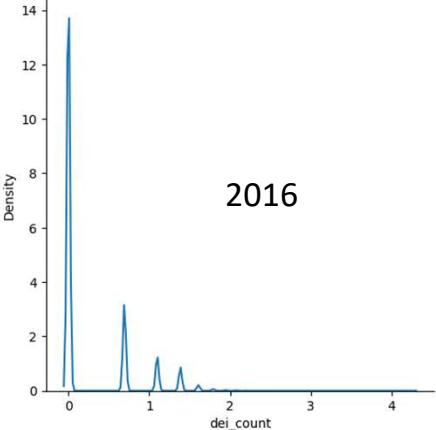


Kernel Density Estimation of the Work-From-Home Keyword Logarithmic Counts in Job Postings in 2022

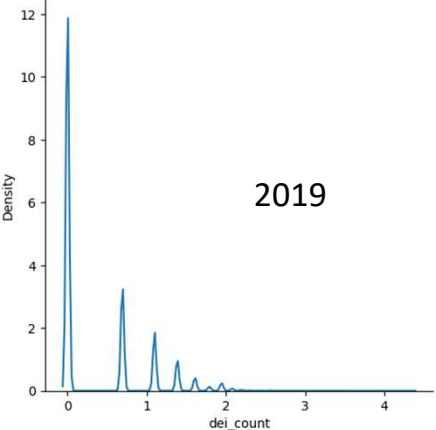


# Just FYI: Distribution of DEI keywords over time (log x+1)

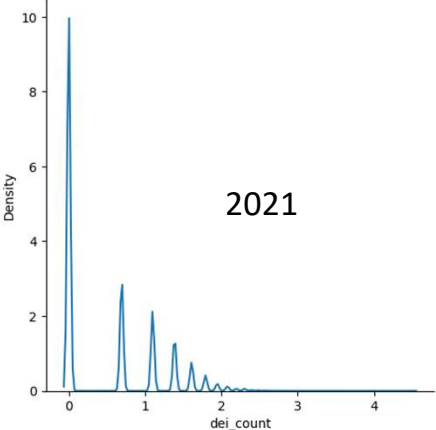
Kernel Density Estimation of the Diversity Equity and Inclusion Keyword Logarithmic Counts in Job Postings in 2016



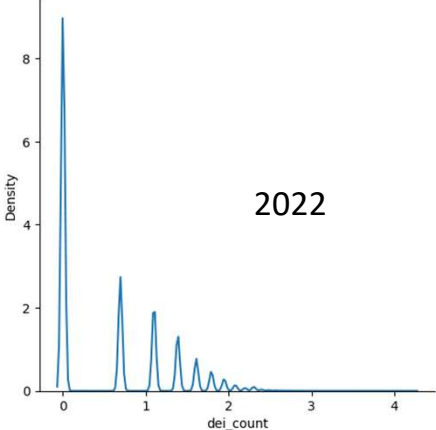
Kernel Density Estimation of the Diversity Equity and Inclusion Keyword Logarithmic Counts in Job Postings in 2019



Kernel Density Estimation of the Diversity Equity and Inclusion Keyword Logarithmic Counts in Job Postings in 2021



Kernel Density Estimation of the Diversity Equity and Inclusion Keyword Logarithmic Counts in Job Postings in 2022



# Results: Probing the defaults with GPT-4

	inp%
<b>Real estate and rental and leasing</b>	71.081081
<b>Educational services</b>	45.283019
<b>Mining, quarrying, and oil and gas extraction</b>	58.236659
<b>Other services (except public administration)</b>	63.779528
<b>Transportation and warehousing</b>	70.193638
<b>Information and cultural industries</b>	39.553703
<b>Professional, scientific and technical services</b>	49.398281
<b>Administrative and support, waste management and remediation services</b>	81.307339
<b>Accommodation and food services</b>	76.421725
<b>Wholesale trade</b>	78.052550
<b>Manufacturing</b>	58.454183
<b>Utilities</b>	57.692308
<b>Retail trade</b>	78.887827
<b>Arts, entertainment and recreation</b>	88.888889
<b>Finance and insurance</b>	48.034223
<b>Construction</b>	69.019608
<b>Health care and social assistance</b>	80.489741

# Stage 1: Preprocessing

- This prompt is fed to GPT-4 and run 5 times with temperature set to 0.7 to generate a comprehensive list of keywords. 5 additional informative keywords are added. After dropping the duplicated keywords (e.g., “remote work” and “remotely” are considered duplicated after we have the keyword “remote”), a list of 50 keywords are finalized:

['nationwide', 'offsite', 'cloud-based', 'work from home', 'online', 'hybrid', 'cloudworking', 'asynchronous', 'e-commute', 'telepresence', 'cloud based', 'location', 'off-site', 'off site', 'nomadic', 'web work', 'work at home', 'relocate', 'web-based', 'decentralized', 'work-at-home', 'home-based', 'home based', 'anywhere', 'ecommuting', 'geographically dispersed', 'remote', 'home work', 'e-work', 'virtual', 'web based', 'e-commuting', 'zoom work', 'wfh', 'telecommute', 'travel', 'home-office', 'zoom job', 'work-from-home', 'distance', 'global', 'digital nomad', 'worldwide', 'mobile', 'web job', 'distributed', 'telework', 'internet-based work', 'home office', 'flex']

# Stage 3: 4-way Classification with GPT-4

- The corresponding list of keywords is as follows:

['direct service', 'desk job', 'offsite', 'zoom work', 'telepresence', 'client site', 'web-based', 'on duty', 'need to be', 'off site', 'customer interaction', 'work at home', 'physically present', 'cloud based', 'on-ground', 'physical job', 'face-to-face', 'required to be', 'relocate', 'hands-on', 'be on site', 'direct contact', 'worldwide', 'physical tasks', 'flex', 'shift', 'e-work', 'personal interaction', 'hospital', 'face to face', 'co-located', 'nine to five', 'telework', 'resident', 'work on site', 'digital nomad', 'cloudworking', 'direct supervision', 'manufacturing', 'zoom job', 'internet-based work', 'attend', 'asynchronous', 'cloud-based', 'customer site', 'colocated', 'branch', 'traditional hours', 'client meetings', 'lab work', 'web work', 'web job', 'on-site', 'personal contact', 'customer service', 'headquarters', 'client visits', 'home work', 'plant', 'travel', 'physical work', 'machinery operation', 'facility', 'work place', 'under supervision', 'regional', 'part time', 'campus', 'job site', 'distance', 'should be', 'anywhere', 'manual work', 'physical presence', 'off-site', 'workplace', 'full time', 'fixed hours', 'home-based', 'manual tasks', 'clinic', 'e-commuting', 'be present', 'distributed', 'work from home', 'remote', 'shop', 'local', 'has to be', 'office', 'work-at-home', 'field work', 'wfh', 'ecommuting', 'on-duty', 'in-situ', 'site visits', 'in person', 'directly supervised', 'work-from-home', 'geographic', 'residency', '9-to-5', 'direct interaction', 'must be', 'hybrid', 'in situ', 'domestic', 'daily presence', 'web based', 'online', 'nationwide', 'presence required', 'on-premise', 'on the ground', 'global', 'presence needed', 'manual labor', 'nomadic', 'site inspections', 'collocated', 'on premises', 'warehouse', 'full-time', 'report to', 'station', 'fieldwork', 'factory', 'immediate supervision', 'store', 'onsite', 'co located', 'equipment operation', 'virtual', 'physical labor', 'hq', 'home based', 'laboratory', 'location', 'direct interface', 'on the job', 'commute', 'visiting clients', 'must reside', 'decentralized', 'in-person', 'in house', 'on ground', 'mobile', 'in-house', 'personal interface', 'site-based', 'part-time', 'in the field']

## Optional Stage 4: 3-way Classification with GPT-4

- The 3-way classification result generated with GPT-3.5 includes a level of confidence, which is positively related to the accuracy rate. Classification results with relatively low levels of confidence can be rerun with GPT-4 to further improve the overall accuracy.
- An ensembled approach is expected to further increase the accuracy.

	confidence	valid	hash
0	0.4	True	6
1	0.5	False	33
2	0.5	True	9
3	0.6	False	12
4	0.6	True	36
5	0.7	False	25
6	0.7	True	43
7	0.8	False	16
8	0.8	True	81
9	0.9	False	21
10	0.9	True	234
11	1.0	False	87
12	1.0	True	4397

<u>PROMPT ID</u>	<u>ACCURACY</u>	<u>CUM_ACCURACY</u>	<u>COVERAGE</u>
Overall	96.12%	96.12%	100.00%
0.4	100.00%	96.12%	100.00%
0.5	21.43%	96.12%	99.88%
0.6	75.00%	96.75%	99.04%
0.7	63.24%	96.96%	98.08%
0.8	83.51%	97.44%	96.72%
0.9	91.76%	97.72%	94.78%
1.0	98.06%	98.06%	89.68%