Transforming naturally occurring text data into economic statistics: the case of online job vacancy postings

Arthur Turrell, Bradley Speigner, David Copple, James Thurgood
Bank of England

Jyldyz Djumalieva
Nesta

arthur.e.turrell@frb.gov

The views expressed do not reflect those of the Bank of England or the Federal Reserve Board.
1. Transform naturally occurring ‘big data’ into economic statistics:
   job adverts $\rightarrow$ vacancy measure
Creating new vacancy statistics from online job adverts

1. Transform naturally occurring ‘big data’ into economic statistics: 
   job adverts → vacancy measure

2. Obtaining disaggregate measure of vacancies by occupation: 
   job description → occupational labels
Creating new vacancy statistics from online job adverts

1. Transform naturally occurring ‘big data’ into economic statistics:
   job adverts $\rightarrow$ vacancy measure

2. Obtaining disaggregate measure of vacancies by occupation:
   job description $\rightarrow$ occupational labels

3. New vacancy stocks and their uses
Transforming naturally occurring ‘big data’ into economic statistics
Vacancy statistics in the UK
The ONS Vacancy Survey (2001–): a measure of the UK stock of vacancies

- Very similar to JOLTS (Job Openings and Labor Turnover Survey)
The ONS Vacancy Survey (2001–): a measure of the UK stock of vacancies

- Very similar to JOLTS (Job Openings and Labor Turnover Survey)

- Each month surveys around 6,000 firms on the total number of vacancies that they have open – measures stock of vacancies
The ONS Vacancy Survey (2001–): a measure of the UK stock of vacancies

- Very similar to JOLTS (Job Openings and Labor Turnover Survey)

- Each month surveys around 6,000 firms on the total number of vacancies that they have open – measures stock of vacancies

- Available at monthly frequency with a 40 day lag
The ONS Vacancy Survey (2001–): a measure of the UK stock of vacancies

- Very similar to JOLTS (Job Openings and Labor Turnover Survey)

- Each month surveys around 6,000 firms on the total number of vacancies that they have open – measures stock of vacancies

- Available at monthly frequency with a 40 day lag

- Collected via a business register so new firms (with many vacancies) underrepresented; however aggregate error likely inconsequential
The ONS Vacancy Survey (2001–): a measure of the UK stock of vacancies

- Very similar to JOLTS (Job Openings and Labor Turnover Survey)

- Each month surveys around 6,000 firms on the total number of vacancies that they have open – measures **stock** of vacancies

- Available at monthly frequency with a 40 day lag

- Collected via a business register so new firms (with many vacancies) underrepresented; however aggregate error likely inconsequential

- Firm-level data collection (via form filled in by head offices) allows for cross-section by firm size or sector
The ONS Vacancy Survey (2001–): a measure of the UK stock of vacancies

- Very similar to JOLTS (Job Openings and Labor Turnover Survey)

- Each month surveys around 6,000 firms on the total number of vacancies that they have open – measures stock of vacancies

- Available at monthly frequency with a 40 day lag

- Collected via a business register so new firms (with many vacancies) underrepresented; however aggregate error likely inconsequential

- Firm-level data collection (via form filled in by head offices) allows for cross-section by firm size or sector

- No breakdown of vacancies by region or occupation is available – understandable
JobCentre Plus data (discontinued 2012): UK vacancies using administrative data

- Job vacancies notified to government employment offices by employers
JobCentre Plus data (discontinued 2012): UK vacancies using administrative data

- Job vacancies notified to government employment offices by employers

- Widely used in economic research; Coles and Smith (1996), Burgess and Profit (2001), Smith (2012), Patterson et al. (2016), and Manning and Petrongolo (2017)
- Job vacancies notified to government employment offices by employers

- Widely used in economic research; Coles and Smith (1996), Burgess and Profit (2001), Smith (2012), Patterson et al. (2016), and Manning and Petrongolo (2017)

- Ignoring bias issues, around a third of all UK vacancies were notified to JCP
- Job vacancies notified to government employment offices by employers

- Widely used in economic research; Coles and Smith (1996), Burgess and Profit (2001), Smith (2012), Patterson et al. (2016), and Manning and Petrongolo (2017)

- Ignoring bias issues, around a third of all UK vacancies were notified to JCP

- Those bias issues are big! ▶ More on bias in JCP
Our data: job adverts posted on Reed.co.uk between 2008 and 2017

- 15,242,000 individual job adverts
Our data: job adverts posted on Reed.co.uk between 2008 and 2017

- 15,242,000 individual job adverts

- Privately run website: firms and recruitment agencies have direct relationship with Reed (not an aggregator)
Our data: job adverts posted on Reed.co.uk between 2008 and 2017

- 15,242,000 individual job adverts

- Privately run website: firms and recruitment agencies have direct relationship with Reed (not an aggregator)

- Cost-to-post: in February 2019, an advert that remains live for 6 weeks is £150 + tax ($197 + tax)
- 15,242,000 individual job adverts

- Privately run website: firms and recruitment agencies have direct relationship with Reed (not an aggregator)

- Cost-to-post: in February 2019, an advert that remains live for 6 weeks is £150 + tax ($197 + tax)

- Fields: job posted date, offered nominal wage, sectoral classification, latitude and longitude of job, job title, and job description
Turning the Reed data into a measure of the stock of vacancies
Creating a stock of job vacancies

- Different definitions (JOLTS, Vacancy Survey, Abraham (1983)). Broadly: vacancies are current, unfilled job openings which are immediately available for occupancy by workers outside a firm and for which a firm is actively seeking such workers (for full-time, part-time, permanent, temporary, seasonal and short-term work)

\[ \text{V}_m = \text{V}_{m-1} + \sum_{d \in m} (\dot{\text{V}}_d - \dot{\text{V}}_{d-6 \times 7}) \]
Creating a stock of job vacancies

- Different definitions (JOLTS, Vacancy Survey, Abraham (1983)). Broadly: vacancies are current, unfilled job openings which are immediately available for occupancy by workers outside a firm and for which a firm is actively seeking such workers (for full-time, part-time, permanent, temporary, seasonal and short-term work)

- last two parts very likely to be satisfied – posting on Reed ensures workers outside of firm see ad, and cost of posting ensures firms are serious about seeking workers

\[ V_m = V_{m-1} + \sum_{d \in m} (\dot{V}_d - \dot{V}_{d-6} \times 7) \]
Creating a stock of job vacancies

- Different definitions (JOLTS, Vacancy Survey, Abraham (1983)). Broadly: vacancies are current, unfilled job openings which are immediately available for occupancy by workers outside a firm and for which a firm is actively seeking such workers (for full-time, part-time, permanent, temporary, seasonal and short-term work).

- Last two parts very likely to be satisfied – posting on Reed ensures workers outside of firm see ad, and cost of posting ensures firms are serious about seeking workers.

- First part needs further work: Reed job adverts are a flow per day, not a stock.

\[ V_m = V_{m-1} + \sum_{d \in m} (\dot{V}_d - \dot{V}_{d-6 \times 7}) \]
Creating a stock of job vacancies

- Different definitions (JOLTS, Vacancy Survey, Abraham (1983)). Broadly: vacancies are current, unfilled job openings which are immediately available for occupancy by workers outside a firm and for which a firm is actively seeking such workers (for full-time, part-time, permanent, temporary, seasonal and short-term work)

- last two parts very likely to be satisfied – posting on Reed ensures workers outside of firm see ad, and cost of posting ensures firms are serious about seeking workers

- First part needs further work: Reed job adverts are a flow per day, not a stock.

- Do not have perfect outflow information but do know that majority of job adverts remain live for 6 weeks\(^1\) after posting so need to transform

\[
V_m = V_{m-1} + \sum_{d \in m} \left( \dot{V}_d - \dot{V}_{d-6 \times 7} \right)
\]  

\(^1\) We are following up with Reed to get more data on how and when they do not.
Vacancy durations vary across the business cycle and 6 weeks probably too long for average (Abraham, 1983; Abraham and Wachter, 1987)
Biases that could affect the stock of vacancies

- Vacancy durations vary across the business cycle and 6 weeks probably too long for average (Abraham, 1983; Abraham and Wachter, 1987)

- Vacancy durations may vary by occupation (though only very weakly for the UK based on JCP)
Biases that could affect the stock of vacancies

- Vacancy durations vary across the business cycle and 6 weeks probably too long for average (Abraham, 1983; Abraham and Wachter, 1987)

- Vacancy durations may vary by occupation (though only very weakly for the UK based on JCP)

- Only some jobs are posted online, only some jobs are posted-at-cost
How biases manifest in stock of vacancies

- Job ads filled or withdrawn before 6 weeks at aggregate level

  More on this

- with reweighting

- Job ads filled or withdrawn before 6 weeks differentially by occupation

  More on this

- with reweighting

- Aggregate coverage < 100%

  More on this

- with reweighting

- Dissagregate coverage < 100%

  More on this

- Composition different compared to all job ads

  More on this

  reduce with reweighting
How biases manifest in stock of vacancies

- Job ads filled or withdrawn before 6 weeks at aggregate level  
  ▶ More on this

- Job ads filled or withdrawn before 6 weeks differentially by occupation  
  ▶ More on this
How biases manifest in stock of vacancies

- Job ads filled or withdrawn before 6 weeks at aggregate level  
  → More on this

- Job ads filled or withdrawn before 6 weeks differentially by occupation  
  → More on this

- Aggregate coverage < 100%  
  → More on this
How biases manifest in stock of vacancies

- Job ads filled or withdrawn before 6 weeks at aggregate level
  
- Job ads filled or withdrawn before 6 weeks differentially by occupation
  
- Aggregate coverage < 100%
  
- Dissagregate coverage < 100%
How biases manifest in stock of vacancies

- Job ads filled or withdrawn before 6 weeks at aggregate level
  - More on this

- Job ads filled or withdrawn before 6 weeks differentially by occupation
  - More on this

- Aggregate coverage < 100%
  - More on this

- Dissaggregate coverage < 100%

- Composition different compared to all job ads
  - More on this
How biases manifest in stock of vacancies

- Job ads filled or withdrawn before 6 weeks at aggregate level
  - fix with reweighting

- Job ads filled or withdrawn before 6 weeks differentially by occupation
  - reduce with reweighting

- Aggregate coverage $< 100\%$
  - fix with reweighting

- Dissaggregate coverage $< 100\%$
  - reduce with reweighting

- Composition different compared to all job ads
  - reduce with reweighting
Coverage by sector: Mean annual ratios of Reed to Vacancy Survey vacancies

- Accommodation & food service activities
- Administrative & support service activities
- Arts, entertainment & recreation
- Construction
- Education
- Electricity, gas, steam & air conditioning supply
- Financial & insurance activities
- Human health & social work activities
- Information & communication
- Manufacturing
- Other service activities
- Professional scientific & technical activities
- Public admin & defence; compulsory social security
- Real estate activities
- Transport & storage
- Wholesale & retail trade; repair of motor vehicles and motor cycles

Ratio of unweighted Reed stock of vacancies to ONS measure
Reweighting to reduce bias

- Use Vacancy Survey by sector and Reed by sector ratios to create weights for Reed data to reduce bias and match aggregate Vacancy Survey more closely.

The stock weight of an individual vacancy $v$ in sector $i$ and at time $t$ is given by $\omega_{i,t} = \frac{V_{vs,i,t}}{V_{i,t}}$, with $V_{vs}$ the Vacancy Survey, and $V_{Reed}$ vacancies effectively eliminates aggregate stock bias.

- Reduces skill-level bias only to extent that vacancy durations are correlated with sectors.
Reweighting to reduce bias

- Use Vacancy Survey by sector and Reed by sector ratios to create weights for Reed data to reduce bias and match aggregate Vacancy Survey more closely

- Stock weight of an individual vacancy \( v \) in sector \( i \) and at time \( t \) is given by

\[
\omega_{i,t} = \frac{V_{i,t}^{\text{vs}}}{V_{i,t}}
\]

with \( V^{\text{vs}} \) the Vacancy Survey, and \( V \) Reed vacancies
Reweighting to reduce bias

- Use Vacancy Survey by sector and Reed by sector ratios to create weights for Reed data to reduce bias and match aggregate Vacancy Survey more closely

- Stock weight of an individual vacancy $v$ in sector $i$ and at time $t$ is given by

$$\omega_{i,t} = \frac{V_{i,t}^{vs}}{V_{i,t}}$$

with $V_{i,t}^{vs}$ the Vacancy Survey, and $V_{i,t}$ Reed vacancies

- Effectively eliminates aggregate stock bias
Reweighting to reduce bias

- Use Vacancy Survey by sector and Reed by sector ratios to create weights for Reed data to reduce bias and match aggregate Vacancy Survey more closely.

- Stock weight of an individual vacancy $v$ in sector $i$ and at time $t$ is given by

\[ \omega_{i,t} = \frac{V_{i,t}^{vs}}{V_{i,t}} \]

with $V_{i,t}^{vs}$ the Vacancy Survey, and $V_i$ Reed vacancies.

- Effectively eliminates aggregate stock bias.

- Reduces skill-level bias only to extent that vacancy durations are correlated with sectors.
Aggregate vacancy stocks from three sources

Stock of vacancies

- JobCentre Plus
- Reed
- Vacancy Survey
- Reed (weighted)

Correlation table
Adding an occupational breakdown
Firms do not post job ads with occupational info. But job descriptions can be used to infer this.

- Simple example

- Use ONS’ Standard Occupational Classification (SOC) codes, e.g. 2425 – ACTUARIES, ECONOMISTS AND STATISTICIANS
Firms do not post job ads with occupational info. But job descriptions can be used to infer this

Simple example

- Use ONS’ Standard Occupational Classification (SOC) codes, e.g. 2425 – ACTUARIES, ECONOMISTS AND STATISTICIANS

- Use the master description of each SOC code, \( d \), from ONS
Firms do not post job ads with occupational info. But job descriptions can be used to infer this.

- Use ONS’ Standard Occupational Classification (SOC) codes, e.g. 2425 – ACTUARIES, ECONOMISTS AND STATISTICIANS

- Use the master description of each SOC code, $d$, from ONS

- Term frequency-inverse document frequency applied to ONS SOC descriptions represents all possible SOC codes with a matrix in which:
  - $t =$ term from within SOC descriptions (all 1- to 3-grams)
  - $d =$ SOC code
  - $\text{tf-idf}(t, d) = \text{tf}(t) \times \left[ \ln \left( \frac{1+D}{1+\text{df}(t,d)} \right) + 1 \right]$
Firms do not post job ads with occupational info. But job descriptions can be used to infer this

- Use ONS’ Standard Occupational Classification (SOC) codes, e.g. 2425 – ACTUARIES, ECONOMISTS AND STATISTICIANS

- Use the master description of each SOC code, $d$, from ONS

- Term frequency-inverse document frequency applied to ONS SOC descriptions represents all possible SOC codes with a matrix in which:
  - $t =$ term from within SOC descriptions (all 1- to 3-grams)
  - $d =$ SOC code
  - $\text{tf-idf}(t, d) = \text{tf}(t) \times \left[ \ln \left( \frac{1+D}{1+\text{df}(t,d)} \right) + 1 \right]$ 

- Can express real job descriptions in the vector space of master SOC descriptions using tf-idf with same terms
Use vector space to find SOC code closest to real job

- \( \hat{\mathbf{v}}' \) = real job vacancy expressed in vector space

Dimension 1
Dimension 2
Dimension T

\( \hat{\mathbf{v}}_e \)
\( \hat{\mathbf{v}}_d \)
\( \hat{\mathbf{v}}'' \)

Tie-break for top 5 matches
Use vector space to find SOC code closest to real job

- \( \hat{\mathbf{v}}' = \text{real job vacancy expressed in vector space} \)
- \( \hat{\mathbf{v}}_d = 3\text{-digit SOC code vector} \)

Tie-break for top 5 matches
Use vector space to find SOC code closest to real job

- \( \hat{V}' = \) real job vacancy expressed in vector space
- \( \hat{V}_d = 3\)-digit SOC code vector
- solve

\[
\arg \max_d \left\{ \hat{V}' \cdot \hat{V}_d \right\}
\]

Tie-break for top 5 matches
**Putting SOCs on – Example**

<table>
<thead>
<tr>
<th>job_title</th>
<th>Physicist</th>
</tr>
</thead>
<tbody>
<tr>
<td>job_description</td>
<td>Make calculations about the universe, do research, perform experiments and understand the physical environment.</td>
</tr>
<tr>
<td>job_sector</td>
<td>Professional, scientific &amp; technical activities</td>
</tr>
</tbody>
</table>
### Example of SOC code assignment

<table>
<thead>
<tr>
<th>job_title</th>
<th>Physicist</th>
</tr>
</thead>
<tbody>
<tr>
<td>job_description</td>
<td>Make calculations about the universe, do research, perform experiments and understand the physical environment.</td>
</tr>
<tr>
<td>job_sector</td>
<td>Professional, scientific &amp; technical activities</td>
</tr>
<tr>
<td>SOC_code</td>
<td>211 – Natural and Social Science Professionals</td>
</tr>
</tbody>
</table>
### Example of SOC code assignment

<table>
<thead>
<tr>
<th>job_title</th>
<th>Physicist</th>
</tr>
</thead>
<tbody>
<tr>
<td>job_description</td>
<td>Make calculations about the universe, do research, perform experiments and understand the physical environment.</td>
</tr>
<tr>
<td>job_sector</td>
<td>Professional, scientific &amp; technical activities</td>
</tr>
<tr>
<td>SOC_code</td>
<td>211 – Natural and Social Science Professionals</td>
</tr>
</tbody>
</table>

Code available at

https://github.com/bank-of-england/occupationcoder → Performance
Vacancy stocks
Region
Regional vacancy stock estimates

- latitude & longitude \(\rightarrow\) region

- Can combine with unemployment data from ONS Labour Force Survey for measure of tightness

- Figure shows regional labour market tightness, \(\theta = \frac{\text{vacancies}}{\text{unemployment}}\)
Occupation
Tight occupations agree with UK Government’s ‘Shortage Occupation List’

THE 3 OCCUPATIONS WITH THE HIGHEST MEAN TIGHTNESS

- Welfare Professionals (244)
- Nursing And Midwifery Professionals (223)
- Customer Service Managers And Supervisors (722)

Tightness (ratio of vacancies to unemployment)
Example uses: occupational matching function
- Constant returns to scale matching function indexed by $i$

$$h_{i,t} = \phi_i U_{i,t-1}^{1-\alpha} V_{i,t-1}^\alpha$$

(2)

where $\alpha$ is the vacancy elasticity
Matching function

- Constant returns to scale matching function indexed by $i$

\[
h_{i,t} = \phi_i U_{i,t-1}^{1-\alpha} V_{i,t-1}^\alpha
\]

where $\alpha$ is the vacancy elasticity

- At disaggregate level, even with reweighting, low-skill occupations may be subject to both upward bias (due to vacancy duration) and downward bias (due to under-representation)
Matching function

- Constant returns to scale matching function indexed by $i$

$$h_{i,t} = \phi_i U_{i,t-1}^{1-\alpha} V_{i,t-1}^\alpha$$ (2)

where $\alpha$ is the vacancy elasticity

- At disaggregate level, even with reweighting, low-skill occupations may be subject to both upward bias (due to vacancy duration) and downward bias (due to under-representation)

- Disaggregated matching efficiencies also biased but hard to be quantitative about how much...
Matching function

- Constant returns to scale matching function indexed by $i$

$$h_{i,t} = \phi_i U_{i,t-1}^{1-\alpha} V_{i,t-1}^\alpha$$  \hspace{1cm} (2)

where $\alpha$ is the vacancy elasticity

- At disaggregate level, even with reweighting, low-skill occupations may be subject to both upward bias (due to vacancy duration) and downward bias (due to under-representation)

- Disaggregated matching efficiencies also biased but hard to be quantitative about how much...

- ...problem likely less bad than other, unweighted data (e.g. JCP)
Aggregate ‘Beveridge’ curve: co-movement of \( V \) and \( U \) over 2008–2017
Mismatch unemployment

- Defined in Şahin et al. (2014) as, for heterogeneous labour markets, the extent of unemployment which arises due to mismatch between jobseekers and job vacancies.
Mismatch unemployment

- Defined in Şahin et al. (2014) as, for heterogeneous labour markets, the extent of unemployment which arises due to mismatch between jobseekers and job vacancies.

- Model provides counter-factuals due to a social planner who allocates the unemployed to search in sub-markets so as to optimise output.
Mismatch unemployment

- Defined in Şahin et al. (2014) as, for heterogeneous labour markets, the extent of unemployment which arises due to mismatch between jobseekers and job vacancies

- Model provides counter-factuals due to a social planner who allocates the unemployed to search in sub-markets so as to optimise output

- Mismatch unemployment formally given by gap between actual unemployment, $u$, and counter-factual unemployment, $u^*$
Mismatch unemployment, $u - u^*$ (seasonally adjusted)
Conclusions

- Presented new statistics on vacancies by region and occupation using naturally occurring big data – hard for surveys to get at these dimensions

- Used novel application of text analysis to create disaggregation by occupation

- Biases likely no worse than for other widely used data, and steps taken to reduce bias with weighting

- Take home message – new, big data sources most useful when
  1. they can be combined with existing classifications; and
  2. they are complements, rather than substitutes, to existing data
Conclusions

- Presented new statistics on vacancies by region and occupation using naturally occurring big data – hard for surveys to get at these dimensions

- Used novel application of text analysis to create disaggregation by occupation
Conclusions

- Presented new statistics on vacancies by region and occupation using naturally occurring big data – hard for surveys to get at these dimensions

- Used novel application of text analysis to create disaggregation by occupation

- Biases likely no worse than for other widely used data, and steps taken to reduce bias with weighting
Conclusions

- Presented new statistics on vacancies by region and occupation using naturally occurring big data – hard for surveys to get at these dimensions

- Used novel application of text analysis to create disaggregation by occupation

- Biases likely no worse than for other widely used data, and steps taken to reduce bias with weighting

- Take home message – new, big data sources most useful when
  1. they can be combined with existing classifications; and
  2. they are complements, rather than substitutes, to existing data
Thank you
Appendix
Bias and coverage in JobCentre Plus data

- Large variation between regions, sectors, and over time depending on business cycle and policies of JCP offices (Machin, 2003)

- Burgess and Pro/uniFB01t (2001) show a disproportionate share of low-skilled, manual jobs + more likely to be matched to the long-term unemployed; Patterson et al. (2016) /uniFB01nd some sectors over-represented

- Not included in labour market statistics releases from 2005 (Bentley, 2005) because it was up to /uniFB01rms to notify when vacancies /uniFB01lled or withdrawn → biased stock upwards by as much as multiple tens of thousands (of total numbers of ∼ 600,000)

- this lead to large amount of 'vacancy deadwood' building-up

Go back
Bias and coverage in JobCentre Plus data

- Large variation between regions, sectors, and over time depending on business cycle and policies of JCP offices (Machin, 2003)

- Burgess and Profit (2001) show a disproportionate share of low-skilled, manual jobs + more likely to be matched to the long-term unemployed; Patterson et al. (2016) find some sectors over-represented
Bias and coverage in JobCentre Plus data

- Large variation between regions, sectors, and over time depending on business cycle and policies of JCP offices (Machin, 2003)

- Burgess and Profit (2001) show a disproportionate share of low-skilled, manual jobs more likely to be matched to the long-term unemployed; Patterson et al. (2016) find some sectors over-represented

- Not included in labour market statistics releases from 2005 (Bentley, 2005) because
  - it was up to firms to notify when vacancies filled or withdrawn → biased stock upwards by as much as multiple tens of thousands (of total numbers of ∼600,000)
  - this lead to large amount of ‘vacancy deadwood’ building-up
Stock-flow bias when job ads are actually filled or withdrawn before 6 weeks – aggregate

- Biases aggregate vacancy stock *upwards*
  - Bias depends on average vacancy duration, known to vary across the business cycle (Abraham, 1983; Abraham and Wachter, 1987) and likely to be less than 6 weeks
  - Vacancy aggregators, e.g. Burning Glass, typically provide no outflow data at all
- Biases aggregate vacancy stock **upwards**
  - Bias depends on average vacancy duration, known to vary across the business cycle (Abraham, 1983; Abraham and Wachter, 1987) and likely to be less than 6 weeks
  - Vacancy aggregators, e.g. Burning Glass, typically provide no outflow data at all

- At the aggregate level, bias is not fixed in time, but is no worse than in the JCP

› Aggregate deviation comparison to JCP
Biases aggregate vacancy stock upwards
- Bias depends on average vacancy duration, known to vary across the business cycle (Abraham, 1983; Abraham and Wachter, 1987) and likely to be less than 6 weeks
- Vacancy aggregators, e.g. Burning Glass, typically provide no outflow data at all

- At the aggregate level, bias is not fixed in time, but is no worse than in the JCP

  Aggregate deviation comparison to JCP

- Can be fixed on aggregate by reweighting with Vacancy Survey stock
Stock-flow bias when job ads are actually filled or withdrawn before 6 weeks – disaggregate

- Vacancy durations may vary by occupation, introducing differential occupation duration bias
Stock-flow bias when job ads are actually filled or withdrawn before 6 weeks – disaggregate

- Vacancy durations may vary by occupation, introducing differential occupation duration bias

- US data (2001-2018) → construction, leisure & hospitality, and trade have shortest durations
  - If broad relationship true for UK too, low skill vacancies biased upwards relative to the average
Stock-flow bias when job ads are actually filled or withdrawn before 6 weeks – disaggregate

- Vacancy durations may vary by occupation, introducing differential occupation duration bias

- US data (2001-2018) → construction, leisure & hospitality, and trade have shortest durations
  - If broad relationship true for UK too, low skill vacancies biased upwards relative to the average

- UK JCP (2004–2012) → little link between occupation & duration
  - Median durations (1-digit SOC) have mean & standard deviation of $4.5 \pm 0.6$ weeks
  - Mean durations (1-digit SOC) have mean & standard deviation of $9.8 \pm 1.3$ weeks
  - If true for UK now, means differential occupation duration bias not a big problem
Stock-flow bias when job ads are actually filled or withdrawn before 6 weeks – disaggregate

- Vacancy durations may vary by occupation, introducing differential occupation duration bias

- US data (2001-2018) → construction, leisure & hospitality, and trade have shortest durations
  - If broad relationship true for UK too, low skill vacancies biased upwards relative to the average

- UK JCP (2004–2012) → little link between occupation & duration
  - Median durations (1-digit SOC) have mean & standard deviation of 4.5 ± 0.6 weeks
  - Mean durations (1-digit SOC) have mean & standard deviation of 9.8 ± 1.3 weeks
  - If true for UK now, means differential occupation duration bias not a big problem

- Reweight with Vacancy Survey to reduce differential occupation duration bias
- Online vacancies are not all vacancies: stock biased downwards with compositional differences
- Online vacancies are not all vacancies: stock biased *downwards* with *compositional* differences

- Vacancies are costly – stock biased *downwards* with *compositional* differences if more cost-effective, alternative channels exist for some jobs
- Online vacancies are not all vacancies: stock biased downwards with compositional differences

- Vacancies are costly – stock biased downwards with compositional differences if more cost-effective, alternative channels exist for some jobs

- Because of these two factors, Reed data likely to over represent middle and high-skilled jobs
- Online vacancies are not all vacancies: stock biased downwards with compositional differences.

- Vacancies are costly – stock biased downwards with compositional differences if more cost-effective, alternative channels exist for some jobs.

- Because of these two factors, Reed data likely to over represent middle and high-skilled jobs.

- Reweight with Vacancy Survey (by sector) to reduce the extent of this bias, and to fix aggregate coverage.
Vacancy durations not correlated with occupation classification: median JCP vacancy durations, 2004–2012 (mean of medians is 4.5 weeks)
Percentage deviations from mean ratio relative to the Vacancy Survey

% deviation of vacancy stock from mean ratio relative to Vacancy Survey

- Reed ratio
- JobCentre Plus ratio
### Correlation between aggregate vacancy time series

<table>
<thead>
<tr>
<th></th>
<th>JobCentre Plus</th>
<th>Vacancy Survey</th>
<th>Reed</th>
<th>Reed (weighted)</th>
</tr>
</thead>
<tbody>
<tr>
<td>JobCentre Plus</td>
<td>1</td>
<td>0.71</td>
<td>0.68</td>
<td>0.69</td>
</tr>
<tr>
<td>Vacancy Survey</td>
<td>-</td>
<td>1</td>
<td>0.93</td>
<td>0.98</td>
</tr>
<tr>
<td>Reed</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>0.90</td>
</tr>
<tr>
<td>Reed (weighted)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
</tbody>
</table>
No occupational labels – firms don’t care about SOCs. How can we use the text of the job descriptions?
Tie-break for top 5 matches

- Choose between the top five matching SOC codes using fuzzy matching on job titles and SOC code job titles
  - use Python package fuzzywuzzy, based on Levenshtein distance (Levenshtein, 1966)
  - this counts number of changes needed to make one string become another
Evaluation of SOC coding algorithm against ONS coding at 3-digit level (200,000 submitted).

<table>
<thead>
<tr>
<th></th>
<th>Manually assigned</th>
<th>Proprietary algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td>330</td>
<td>67,900</td>
</tr>
<tr>
<td>Accuracy</td>
<td>76%</td>
<td>91%</td>
</tr>
</tbody>
</table>

Go back
Stylised fact in Vacancy Survey disaggregation also exists in Reed disaggregations

- Vacancy Survey by sector follows a Taylor power law such that the monthly mean and monthly variance are related by

\[ \sigma_t^2 = aV_t^b \]

with \( R^2 = 0.86 \) and \( b = 2.04 \pm 0.06 \)
Stylised fact in Vacancy Survey disaggregation also exists in Reed disaggregations

- Vacancy Survey by sector follows a Taylor power law such that the monthly mean and monthly variance are related by

\[ \sigma_t^2 = a \bar{V}_t^b \]

with \( R^2 = 0.86 \) and \( b = 2.04 \pm 0.06 \)

- Do our data also follow Taylor power law when disaggregated?
Stylised fact in Vacancy Survey disaggregation also exists in Reed disaggregations

- Vacancy Survey by sector follows a Taylor power law such that the monthly mean and monthly variance are related by

\[
\sigma_t^2 = a \bar{V}_t^b
\]

with \( R^2 = 0.86 \) and \( b = 2.04 \pm 0.06 \)

- Do our data also follow Taylor power law when disaggregated?

- Yes – shown for 3-digit occupations (but also true for regional data).
Data: Beveridge curve by region at the 1-digit UK NUTS level
Econometric results on matching function estimation

The baseline empirical matching regression is

\[
\ln \left( \frac{h_{i,t}}{U_{i,t-1}} \right) = \ln \phi_i + \alpha \ln \left( \frac{V_{i,t-1}}{U_{i,t-1}} \right) + \epsilon_{i,t}
\]  

(3)

NB: \( \phi_i \) capture cross-section fixed effects.

<table>
<thead>
<tr>
<th>Elasticity parameter (( \alpha ))</th>
<th>1-digit SOC</th>
<th>2-digit SOC</th>
<th>3-digit SOC</th>
<th>1-digit NUTS</th>
<th>Aggregate data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Point estimate (least squares)</td>
<td>.396</td>
<td>.427</td>
<td>.431</td>
<td>.254</td>
<td>.367</td>
</tr>
<tr>
<td>Standard error</td>
<td>.075</td>
<td>.050</td>
<td>.037</td>
<td>.020</td>
<td>.030</td>
</tr>
<tr>
<td>Point estimate (IV)</td>
<td>.392</td>
<td>.442</td>
<td>.371</td>
<td>.275</td>
<td>.350</td>
</tr>
<tr>
<td>Standard error</td>
<td>.073</td>
<td>.061</td>
<td>.048</td>
<td>.026</td>
<td>.031</td>
</tr>
<tr>
<td>Cross-sections</td>
<td>9</td>
<td>25</td>
<td>90</td>
<td>12</td>
<td>-</td>
</tr>
<tr>
<td>Observations</td>
<td>324</td>
<td>852</td>
<td>2120</td>
<td>423</td>
<td>35</td>
</tr>
</tbody>
</table>

Matching function parameter estimates. All results are significant at the 1% level.
Model

- Şahin et al. (2014) model – optimal path for output due to social planner assigning unemployed to sub-markets  
  More details on model

- The planner chooses $\tilde{u}_t$ to maximise output:

$$V(u_t, \tilde{e}_t; \Xi_t) = \max_{\{u_{i,t}\}} \left\{ \sum_i z_{i,t}(e_{i,t} + \gamma h_{i,t}) - \kappa u_t + \beta \mathbb{E} [V(u_{t+1}, \tilde{e}_{t+1}; \Xi_{t+1})] \right\}$$

- Counter-factual employment path

$$e_{it}^* = (1 - \xi_{t-1}) e_{i,t-1}^* + h_{it}(v_{it}, u_{it}^*)$$

- Counter-factual output path

$$Y_t^* = \sum_i z_{it} e_{it}^* + y_t^*$$
- Follow methodology of Şahin et al. (2014) – optimal path for output due to social planner assigning unemployed to sub-markets

- The planner chooses $\tilde{u}_t$ to maximise output:

$$ V(u_t, \tilde{e}_t; \Xi_t) = \max_{\{u_{i,t}\}} \left\{ \sum_i z_{i,t}(e_{i,t} + \gamma h_{i,t}) - \kappa u_t + \beta \mathbb{E}[V(u_{t+1}, \tilde{e}_{t+1}; \Xi_{t+1})] \right\} $$

such that $\sum_i u_{i,t} \leq u_t$ where $u_{t+1} = L_{t+1} - \sum_i e_{i,t+1}$.

- $\gamma$ is ‘hit’ of 2/3 to productivity after a hire

- $\Xi_t = (\tilde{z}_t, \tilde{V}_t, \tilde{\phi}_t, \tilde{\zeta}_t)$ with $\tilde{\zeta}$ the job destruction rate
- Social planner’s optimal allocation is $\tilde{u}_t^*$

- Gives rise to counter-factual employment path

\[ e_{it}^* = (1 - \zeta_{t-1}) e_{i,t-1}^* + h_{it}(v_{it}, u_{it}^*) \]

- Counter-factual output is

\[ Y_t^* = \sum_i z_{it} e_{it}^* + y_t^* \]

- Output per worker in the realised and counter-factual cases given by $Y_t / e_t$ and $Y_t^* / e_t^*$ respectively
References


Smith, Jennifer C. 2012. “Unemployment and Mismatch in the UK.”