Pretty Vacant: using job vacancies to understand the effects of mismatch on UK productivity growth

Arthur Turrell†, Bradley Speigner†, James Thurgood†, Jyldyz Djumalieva*, and David Copple†

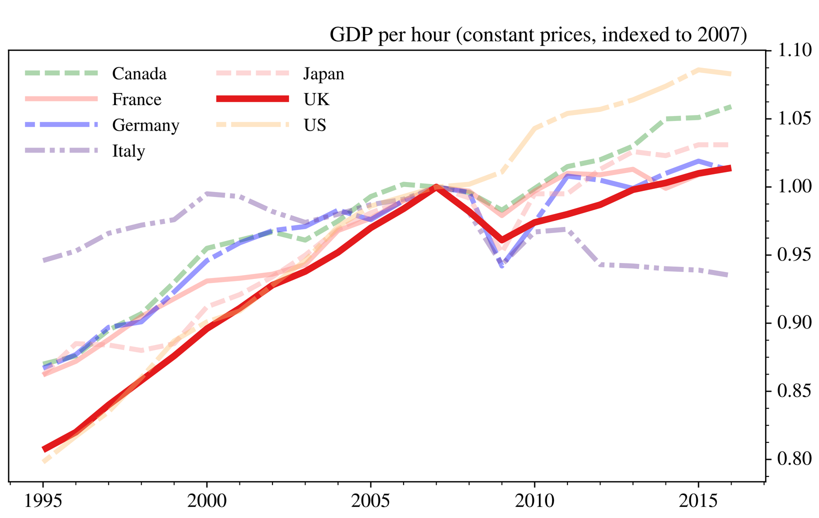
†Bank of England, *NESTA

BANK OF ENGLAND

Objectives

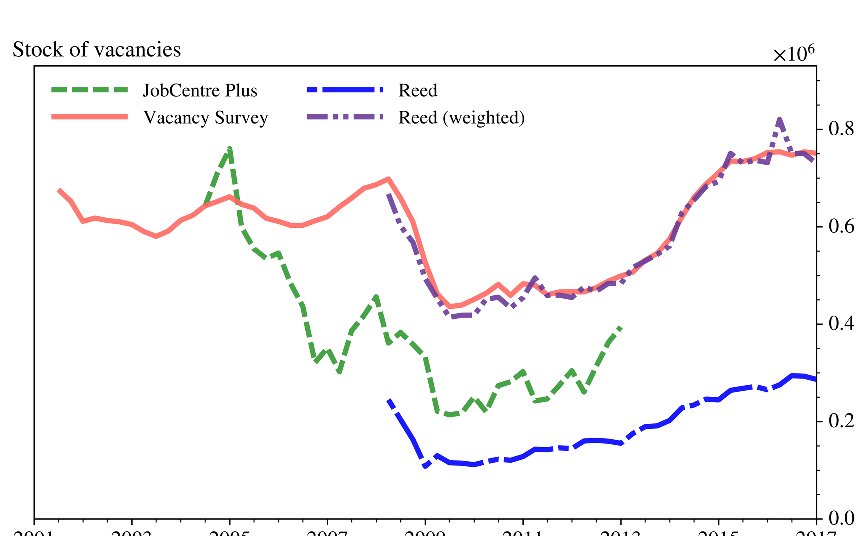
Using ‘big’ data – 15 million vacancies posted online at daily frequency from 2008 to 2016 – we ask:

- How has occupational mismatch affected UK productivity growth?
- How has regional mismatch affected UK productivity growth?
- Can we create a labour market segmentation ‘bottom-up’ which is data-driven in both the **type** and **level** of aggregation?



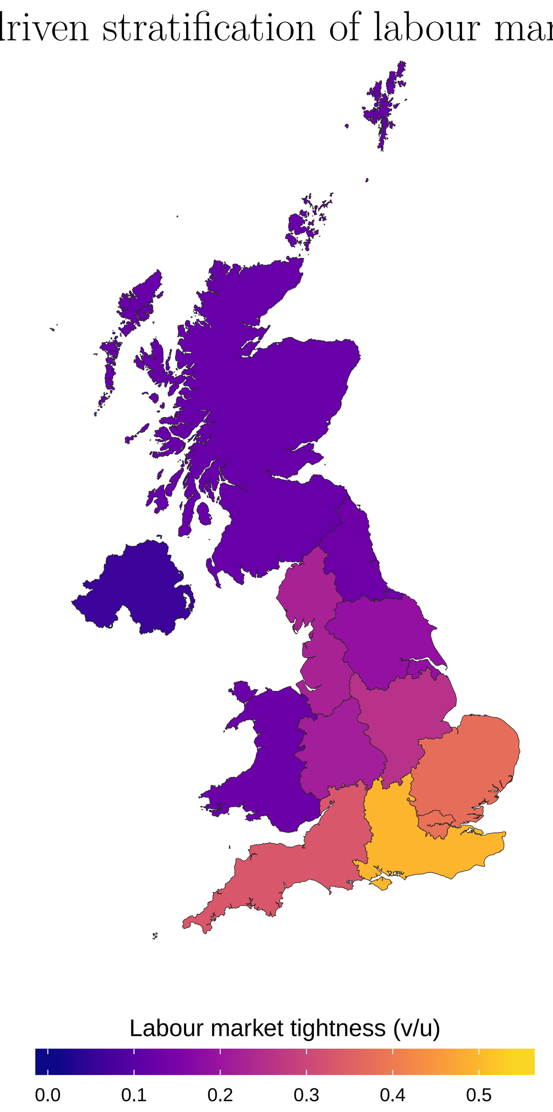
Introduction

Weak UK productivity growth is an enduring puzzle of the post-crisis period. Meanwhile, in 2017, UK job vacancies hit their highest level since official vacancy statistics began (see figure below). The ‘mismatched’ distribution of workers and jobs could have contributed to the productivity puzzle, as suggested for occupational mismatch in Patterson et al. (2016). Mismatch arises when there are barriers to mobility across parts of the labour market: the aggregate market operates as a series of distinct, smaller markets. We bring new data from a recruiter, ‘Reed’, to this problem.



Marginal contribution

- Mapping messy online data into official classifications
  - an algorithm to map job descriptions into standard occupational codes using term frequency-inverse document frequency
- Revisit the contribution of occupational mismatch to weak productivity growth
  - using new data which are better correlated with aggregate official statistics than widely used JobCentre Plus data
- New estimates of matching function parameters
- Examine mismatch by region for the UK
- Data-driven stratification of labour market



Theory

We use the search and match theory with segment specific matching efficiencies, ϕ_i; hires, h_i;, vacancies, v_i, and unemployment, u_i,

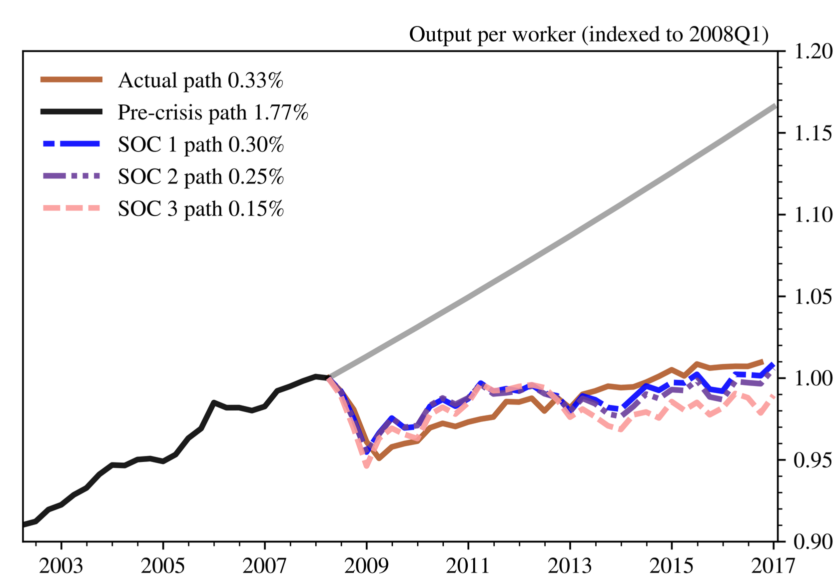
$$h_i = \phi_i M(u_i, v_i) = \phi_i u_i^{1-\alpha} v_i^{\alpha} \tag{1}$$

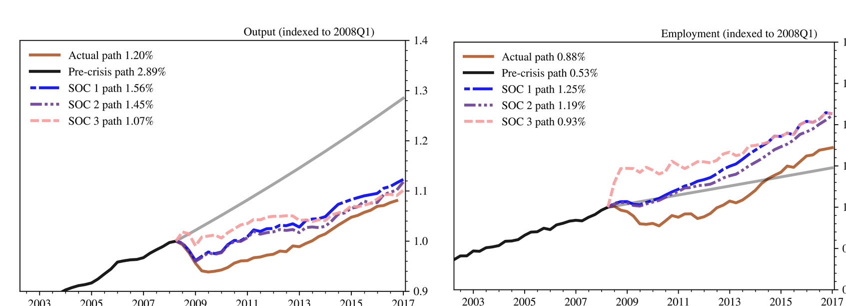
We use the mismatch framework developed by Şahin et al. (2014) and used by Patterson et al. (2016). A social planner's problem is given by

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

where u_{it} is the control variable, ξ the job destruction rate, γ a re-training penalty, and z_{it} productivity. Counter-factuals maximise output based on the social planner's allocation of the unemployed, u^*_{it}, with total output given by $Y^*_t = \sum_i^I z_{it} e^*_{it}$ where $e^*_{i,t+1} = (1-\xi_t)e^*_{i,t} + h^*(v_{it}, u^*_{it})$.

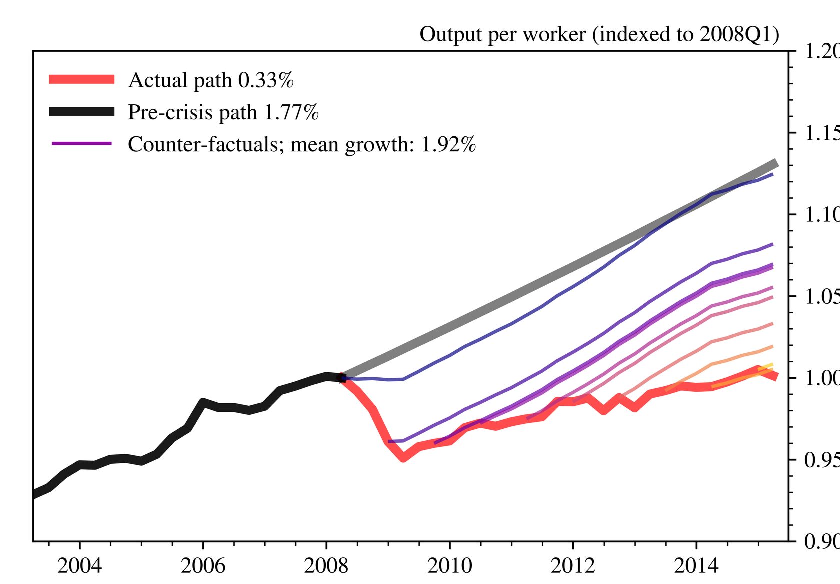
Results – occupation

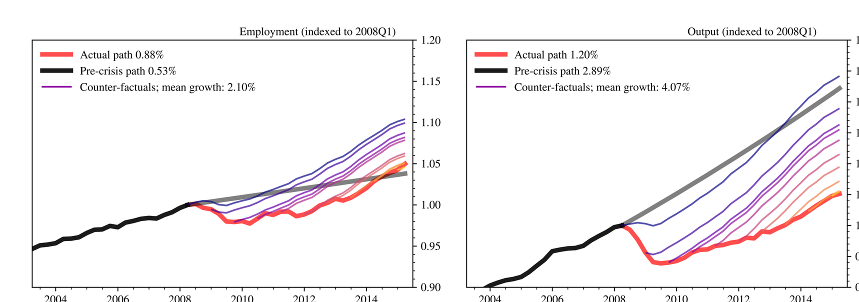




The social planner's path results in higher output but *lower* productivity. This is driven by the heterogeneity in matching efficiency (ϕ), in productivity (z), and in tightness (θ) across occupations. Broadly, matching efficiency and productivity are anti-correlated. Occupational mismatch cannot explain the UK's current productivity puzzle.

Results – region





Mismatch by region predates the financial crisis, and so cannot explain the productivity puzzle. But unwinding it (shown with simulations beginning from every third quarter) shows that it has played a significant role as an inhibitor of output and productivity growth. The results are driven by regions where $\{\phi, z, \theta\}$ are *positively* correlated.

Main results

Neither occupational nor regional mismatch can explain the enduring productivity puzzle, but resolving regional mismatch could provide a significant boost to productivity growth.

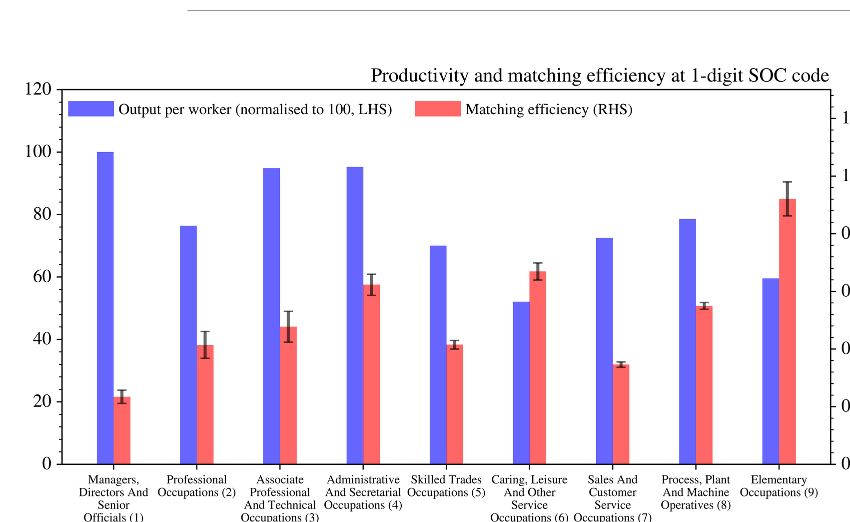
Matching function estimates

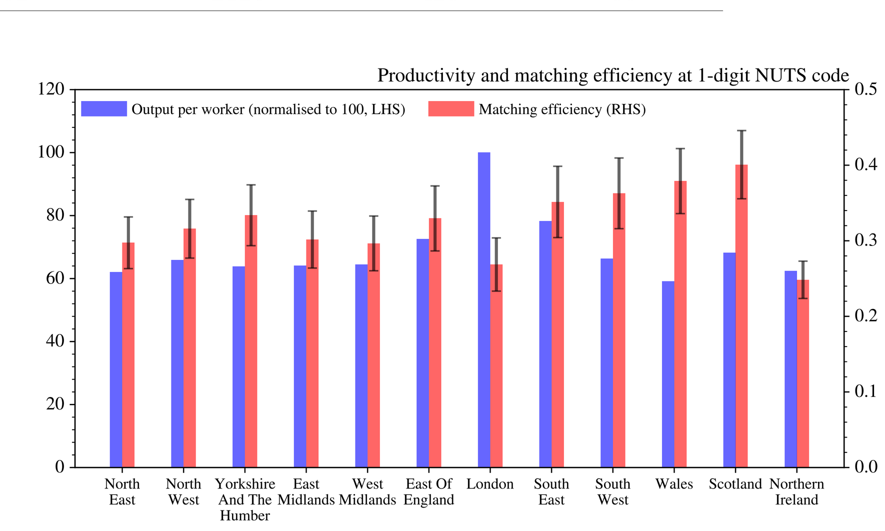
The baseline empirical matching regression is

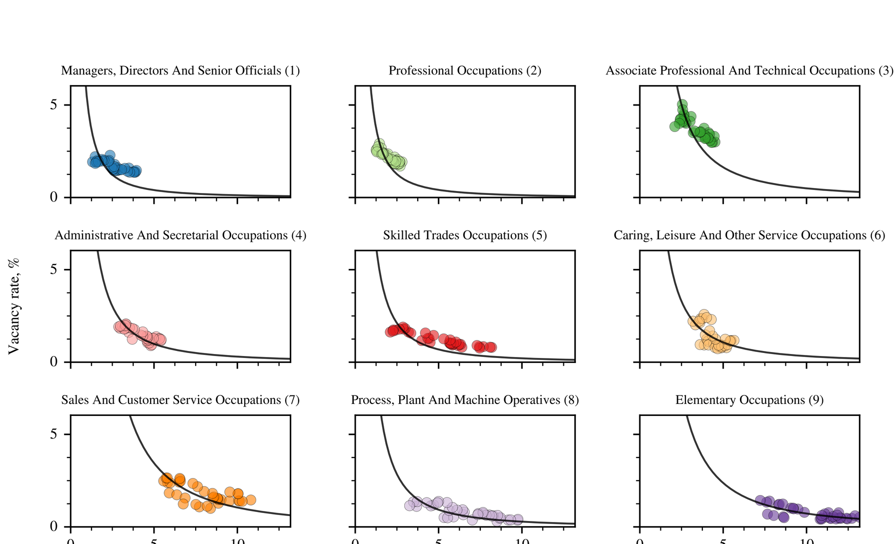
$$\ln\left(\frac{h_{i,t}}{U_{i,t-1}}\right) = \ln \phi_i + \alpha \ln\left(\frac{V_{i,t-1}}{U_{i,t-1}}\right) + \epsilon_{i,t} + d \tag{2}$$

where ϕ_i capture cross-section fixed effects and d is a quarterly dummy variable. In the table below, ordinary least squares with cross-section clustered standard errors is applied to the pooled data, while the instrumental variable is a single lag of tightness. All results are significant at the 1% level.

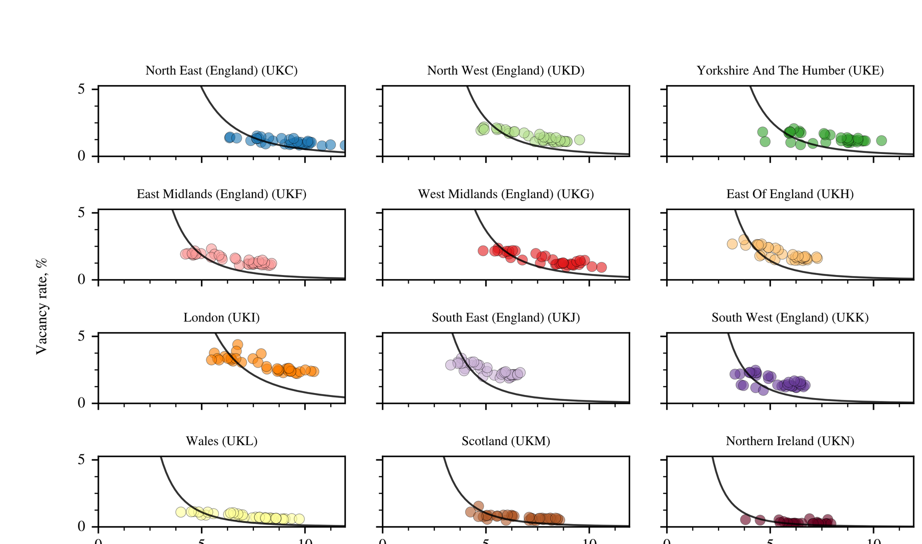
	1-digit occupation	2-digit occupation	3-digit occupation	1-digit region	Aggregate data
Elasticity parameter (α)					
Point estimate (least squares)	.396	.427	.431	.254	.367
Standard error	.075	.050	.037	.020	.030
Point estimate (IV)	.392	.442	.371	.275	.350
Standard error	.073	.061	.048	.026	.031
Cross-sections	9	25	90	12	-
Observations	324	852	2120	423	35







Beveridge curves by 1-digit occupation



Beveridge curves by 1-digit region

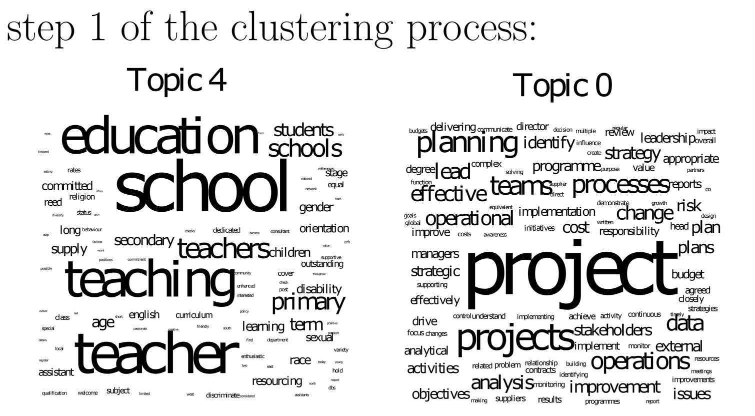
Bottom-up market segmentation

- In mismatch, determining both the appropriate **type** and **level** of disaggregation is an open problem (Barnichon and Figura, 2015; Petrongolo and Pissarides, 2001)
- Our analysis supports this:
  - We find that standard mismatch indices are increasing in the level of disaggregation, and differ by type of disaggregation
  - We find that estimates of structural parameters also differ by level and type
  - Data show that many job-to-job moves are across sectors or occupations
- We try to develop a ‘bottom-up’, data-driven segmentation of the labour market

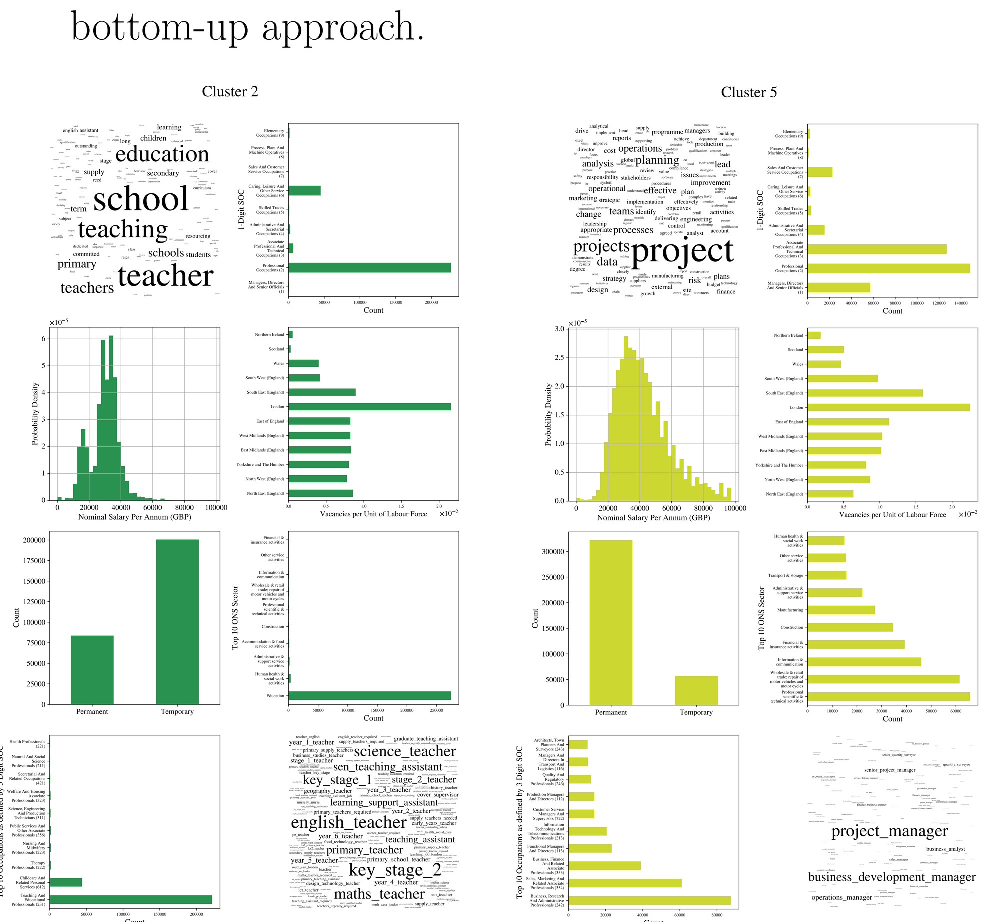
We group vacancies based upon the **demand** expressed in the job description. We use machine learning algorithms to do this, processing the text in three steps:

1. Latent Dirichlet Allocation to express each vacancy in terms of a vector space composed of N topics
   - Fix N using ‘weighted saliency’ (Goldsmith-Pinkham et al., 2016)
2. The K-means clustering algorithm to group vacancies expressed in topic space into market segments (the **type** of disaggregation), for many values of K
3. Silhouette scores to determine K (the **level**)

Below are word clouds for two of the topics selected by step 1 of the clustering process:



$K = 20$ is chosen by the algorithm as the **level** of disaggregation. Two example market segments are shown below; market segment 2 is mostly made up of vacancies strong in topic 4, and clearly corresponds to an official classification. Market segment 5 cuts across official classifications, demonstrating the usefulness of the approach. The ultimate aim is to determine the extent of mismatch using this bottom-up approach.



References

Barnichon, R. and A. Figura
2015. Labor market heterogeneity and the aggregate matching function. *American Economic Journal: Macroeconomics*, 7(4):222–249.

Goldsmith-Pinkham, P., B. Hirtle, and D. Lucca
2016. Parsing the content of bank supervision. Staff Reports 770, Federal Reserve Bank of New York.

Patterson, C., A. Şahin, G. Topa, and G. L. Violante
2016. Working hard in the wrong place: A mismatch-based explanation to the uk productivity puzzle. *European Economic Review*, 84:42–56.

Petrongolo, B. and C. A. Pissarides
2001. Looking into the black box: A survey of the matching function. *Journal of Economic literature*, 39(2):390–431.

Şahin, A., J. Song, G. Topa, and G. L. Violante
2014. Mismatch unemployment. *The American Economic Review*, 104(11):3529–3564.

Contact information

arthur.turrell@bankofengland.co.uk